Topics in nonlinear conic optimization and applications

Leonardo Makoto Mito

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 1: Prof. Dr. Gabriel Haeser

Advisor 2: Prof. Dr. Héctor Ariel Ramírez Cabrera

Financial support by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)
grant 2017/17840-2.

São Paulo, March, 2022

Topics in nonlinear conic optimization and applications

Leonardo Makoto Mito

This is the final version of the doctoral thesis written by Leonardo Makoto Mito and revised by the examination committee, submitted to the Institute of Mathematics and Statistics of the University of São Paulo.

Examination committee:

- Dr. Gabriel Haeser (co-advisor, president) - *Universidade de São Paulo*
- Dr. Héctor Ariel Ramírez Cabrera (co-advisor) - *Universidad de Chile*
- Dr. Defeng Sun - *The Hong Kong Polytechnic University*
- Dr. Gábor Pataki - *University of North Carolina*
- Dr. Jérôme Malick - *Université Grenoble Alpes*
- Dr. Patrick Mehlitz - *Brandenburgische Technische Universität Cottbus-Senftenberg*

Acknowledgements

Written in January 31, 2022.

Doctoral theses tend to be forgotten in time and this one is unlikely to be an exception, yet I wrote it assuming it would be read. I'll read it again eventually, hoping I've grown at least enough to read it with pride and not nostalgia. Further, although I recognize how conceited this may sound, I assume to have at least one *nontrivial reader*. In fact, this note is also meant to thank you, my nontrivial reader, regardless of the fact you probably stumbled upon this thesis for reasons that are unknown even to yourself, and to tell you that I sincerely hope you can benefit from what I've written, nevertheless.

I always tell people that I'm a humanities person disguised as a mathematician, as what catches my attention in mathematics is not the results, but the definitions and proofs. To me, definitions are other people's interpretations of things and proofs are, loosely speaking, extremely detailed accounts of how they convinced themselves of something they believe it's true. Thus, from my point of view knowing mathematics is knowing people, and I was fortunate enough to know many interesting people during my doctoral course. At this point I should mention my first advisor Gabriel Haeser, my informal advisor Nino (Roberto Andreani), my second formal advisor Héctor Ramírez, and my first external collaborator Walter Gómez. They all offered to teach me for free at some point, and I made a huge effort in return so they wouldn't waste their time; to them I am immeasurably grateful. I am also deeply grateful to Ellen Fukuda and Bruno Lourenço, who also accepted me as their pupil and gave me much of their time, even though I was never capable of repaying a fraction of their good will.

Besides the people who mentored me there are also those who taught me involuntarily. They are the people for whom I have solid admiration, which I've been consistently imitating since I knew of their existence in the hope of becoming like them, but they're probably not aware of it. To cite a few names: Ernesto Birgin, Paulo Silva, Alberto Ramos, Leonardo Secchin, Felipe Bueno, Mituhiro Fukuda, Jorge Sotomayor Tello, Sandra Santos, Luiz Rafael dos Santos, Pedro Peixoto, and Daniel Steck, who is the only person in this list that I never met[1].

Defeng Sun, Gabor Pataki, Jérôme Malick, and Patrick Mehlitz. To me, these names are more than special. I admire their work so much, and I've seen their names so many times, that I started to believe that they are not real human beings; perhaps demigods, I don't know. Can you

[1]I'll explain: Steck's Ph.D. thesis and papers were very helpful to me.

remotely guess how happy I was to have them judging my work? I am eternally grateful to them for being part of my thesis' committee. Now, I feel truly accomplished; even though I decided to pursue new challenges outside of academia, I will never forget the unique experience that I lived until this moment. I'm twice as grateful to Patrick Mehlitz for giving such a careful read in my thesis, correcting some very important parts of it along the way.

I am also grateful to all my friends from IME-USP and ICT-UNIFESP for their support and awesome discussions on random (sometimes weird) subjects. I'm lucky for having too many names in mind to list them all here, so I won't list any because I don't feel capable of selecting just a few.

Over the past six years I have received a decent amount of money from the government through Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Regardless of the amount, I strongly believe people who receive money from the government must work harder than those who don't, specifically here in Brazil. I have no idea of whether I have paid my debt with the Brazilian population or not, because the quality of my work at this point is completely determined by the opinion of other researchers on it (meaning this is out of my control, to some extent), but I can assure I did my best to not waste the money of those who never chose to pay me. That said, I'm absolutely grateful to whoever approved my Ph.D. scholarship request; perhaps I didn't even deserve it at the time it was approved, but I know there were people who advocated for me at some point, and I am truly grateful to them.

Last but certainly not least, I must write about my wife, Rayara. At the time I'm writing this text, we are together for almost 10 years; I was her husband before I was a mathematician or whatever it is that I am now. She's much more to me than a friend or a lover, she's the protagonist of my story. My wife is probably the only reason why I do *something* instead of *nothing*, and even if she wasn't my wife she would be the most incredible person in the world to me. Although this may sound like some sort of romantic applesauce, it's not. We are both completely different people from what we were when we started our relationship, life has not always been nice to us, we had our crises, but we always stayed together, we changed together, and we slowly became the best match for one another. Just as many people of my generation, I've never had any true ambition or any life purpose besides avoiding the inherent inconveniences of living. But I do have one single person I believe that deserves to have good reasons to actually enjoy living. I dedicate my thesis to her, just as everything else in my existence.

Sincerely,

Leonardo M. Mito

Abstract

MITO, L. M. **Topics in nonlinear conic optimization and applications**. Ph.D. thesis. Institute of Mathematics and Statistics of the University of São Paulo. Brazil, 2022.

This thesis has three main parts: in part one, we develop new sequential optimality conditions for *Nonlinear Conic Programming* (NCP) problems, which are used to study convergence of algorithms in a simplified and unified way. In part two, we extend the so-called *Constant Rank Constraint Qualification* (CRCQ) and the *Constant Rank of the Subspace Component* (CRSC) conditions to the context of NCP over reducible cones by means of new geometric characterizations of them; we use these conditions to prove strong second-order optimality results that improve the classical one obtained under Robinson's Constraint Qualification, and we show how CRSC is related to a nonlinear extension of the celebrated *facial reduction* preprocessing technique. In part three, we present an alternative approach to extending CRCQ and the *Constant Positive Linear Dependence* (CPLD) conditions to Nonlinear Semidefinite and Second-Order Cone Programming, which has applications in the global convergence theory of a class of numerical methods to first-order stationary points. Then, we combine some of the ideas presented in part two with the CRCQ extension of part three to derive a Weak Constant Rank property that modifies the second-order optimality condition induced by Robinson's CQ to a notion that better suits convergence of algorithms.

Keywords: Conic optimization, First-order optimality conditions, Second-order optimality conditions, Constraint qualifications, Algorithms, Global convergence.

Resumo

MITO, L. M. **Tópicos em otimização cônica não-linear e aplicações**. Tese de doutorado. Instituto de Matemática e Estatística da Universidade de São Paulo. Brasil, 2022.

Esta tese pode ser dividida em três partes: na parte um, nós desenvolvemos novas condições sequenciais de otimalidade para problemas de *Otimização Cônica Não-Linear* (NCP), que são usadas para estudar a convergência global de algoritmos de modo unificado e simplificado. Na parte dois, nós estendemos a chamada *Condição de Qualificação do Posto Constante* (CRCQ) e a condição do *Posto Constante do Subespaço Componente* (CRSC) para o contexto de NCP sobre cones redutíveis, por meio de novas caracterizações geométricas destas condições. Nós as usamos para provar resultados de otimalidade fortes de segunda ordem que melhoram os resultados clássicos obtidos sob a Condição de Qualificação de Robinson, e mostramos como CRSC está relacionada com uma extensão não-linear de uma técnica popular de pré-processamento conhecida como *redução facial*. Na parte três, nós apresentamos uma abordagem alternativa para estender tanto CRCQ quanto sua variante conhecida como a condição da *Dependência Linear Positiva Constante* (CPLD), para problemas não-lineares de otimização sobre os cones semidefinido e de segunda ordem. Essas extensões alternativas têm aplicações na teoria de convergência global de uma classe de métodos numéricos para pontos estacionários de segunda ordem. Então, nós incorporamos algumas das ideias apresentadas na parte dois com a extensão de CRCQ da parte três para derivar uma propriedade fraca do tipo posto constante, que modifica a noção de segunda ordem induzida pela condição de Robinson para algo mais aplicável à convergência de algoritmos.

Keywords: Otimização cônica, Condições de otimalidade de primeira ordem, Condições de otimalidade de segunda ordem, Condições de qualificação, Algoritmos, Convergência global.

Contents

Abbreviature list

(N)LP	(Non)linear Programming
(N)SOCP	(Nonlinear) Second-Order Cone Programming
(N)SDP	(Nonlinear) Semidefinite Programming
(N)CP	(Nonlinear) Conic Programming
(S)CQ	(Strict) Constraint Qualification
LICQ	Linear Independence Constraint Qualification
MFCQ	Mangasarian-Fromovitz Constraint Qualification
MSCQ	Metric Subregularity Constraint Qualification
(R)CPLD	(Relaxed) Constant Positive Linear Dependence (Constraint Qualification)
(R)CRCQ	(Relaxed) Constant Rank Constraint Qualification
CRSC	Constant Rank of the Subspace Component
WCR	Weak Constant Rank (Property)
(A)KKT	(Approximate) Karush-Kuhn-Tucker
CAKKT	Complementary Approximate Karush-Kuhn-Tucker
AGP	Approximate Gradient Projection
(B)SOC	(Basic) Second-Order (Necessary) Condition
WSOC	Weak Second-Order (Necessary) Condition
SSOC	Strong Second-Order (Necessary) Condition
(S)SOSC	(Strong) Second-Order Sufficient Condition
ALM	Augmented Lagrangian Method
SQP	Sequential Quadratic Programming (Method)
IP(M)	Interior-Point (Method)

Notation

Set-related notation:

$\mathbb{N}$	Natural numbers (without zero)
$\mathbb{R}$	Real numbers
$\mathbb{R}^m$	m-dimensional Euclidean space; n-fold Cartesian product of $\mathbb{R}$
$\mathbb{R}^{m\times n}$	Set of matrices with real entries, m rows, and n columns
$\mathbb{R}^m_+$	Nonnegative orthant of $\mathbb{R}^m$
$\mathbb{R}^m_-$	Nonpositive orthant of $\mathbb{R}^m$
$\mathbb{S}^m$	Linear space of all $m \times m$ symmetric matrices with real entries
$\mathbb{S}^m_+$	Cone of $m \times m$ symmetric positive semidefinite matrices
$\mathbb{S}^m_-$	Cone of $m \times m$ symmetric negative semidefinite matrices
$\mathbb{L}^m$	Lorentz cone, ice-cream cone, second-order cone
$\mathrm{int}(C)$	Interior of a set C
$\mathrm{cl}(C)$	Closure of a set C
$\mathrm{bd}(C)$	Boundary of a set C
$\mathrm{bd}_+(C)$	Boundary of a cone C excluding the origin
$\preceq$	Partial order induced by $\mathbb{S}^m_+$
$\leqslant$	Partial order induced by $\mathbb{R}^m_+$ (abuse of notation)
$C + K$	Minkowski sum of two sets C and K

Linear algebra and convex analysis notation:

$\langle x, y\rangle$	Unspecified inner product between x and y
$\|x\|$	Unspecified norm of x
$\|x\|_2$	Euclidean (2-)norm of $x \in \mathbb{R}^m$
$\|x\|_\infty$	Supremum (∞-)norm of $x \in \mathbb{R}^m$
$\mathrm{dist}(x, C)$	Distance between a point $x \in \mathbb{Y}$ and a set $C \subseteq \mathbb{Y}$
$\Pi_C(x)$	Orthogonal projection of a point $x \in \mathbb{Y}$ onto a closed convex set $C \subseteq \mathbb{Y}$
$B(x, \varepsilon)$	Open ball centered at x with radius ε
$B[x, \varepsilon]$	Closed ball centered at x with radius ε
$\mathbb{Y}^\perp$	Orthogonal complement of a subspace $\mathbb{Y}$
$\dim(\mathbb{Y})$	Dimension of a linear space $\mathbb{Y}$

span(C)	Smallest linear subspace that contains a given set C
$\mathrm{lin}(C)$	Largest subspace contained in a given set C
aff(C)	Smallest affine subspace that contains a given set C (affine hull)
conv(C)	Smallest convex set that contains a given set C (convex hull)
cone(C)	Smallest cone that contains a given set C (conic hull)
C°	Polar cone of a set C
$T_C(x)$	Tangent cone to a set C at $x \in C$
$N_C(x)$	Normal cone to a set C at $x \in C$
$L_C(x)$	Linearized tangent cone to a set C at $x \in C$
$F \trianglelefteq C$	F is a face of C
$F^{\triangle}$	Conjugate face of $F \trianglelefteq C$; that is, $F^{\triangle} \doteq F^{\perp} \cap C^{\circ}$, where $F^{\perp} \doteq \mathrm{span}(F)^{\perp}$

Matrix notation:

$\mathbb{I}_m$	$m \times m$ identity matrix
$M^{\top}$	Transpose (real adjoint) of $M \in \mathbb{R}^{m\times n}$
M^{-1}	Inverse of $M \in \mathbb{R}^{m\times m}$
$M^{\dagger}$	Moore-Penrose pseudoinverse of $M \in \mathbb{R}^{m\times n}$
tr(M)	Trace of $M \in \mathbb{R}^{m\times m}$
$\det(M)$	Determinant of $M \in \mathbb{R}^{m\times m}$
$\mathrm{diag}(M)$	Vector formed by extracting the diagonal elements of $M \in \mathbb{R}^{m\times m}$
$\mathrm{Diag}(v)$	Matrix that has a given vector $v \in \mathbb{R}^{m}$ in its diagonal and zeros elsewhere
$[M_{ij}]_{\substack{i\in\{1,\ldots,n\}\\ j\in\{1,\ldots,m\}}}$	Matrix whose (i,j)-th entry is M_{ij}
$(v_i)_{i\in\{1,\ldots,n\}}$	Vector whose i-th entry is v_i
$\Vert M\Vert_F$	Frobenius norm of $M \in \mathbb{R}^{m\times m}$

Sequences and functions:

$\{x^k\}_{k\in I}$	Sequence labelled by a variable x and indexed by $I \subseteq \mathbb{N}$
	Family of vectors indexed by $I \subseteq \mathbb{N}$
$\{x^k\}_{k\in I} \to \overline{x}$	Convergence of $\{x^k\}_{k\in I}$ to $\overline{x}$
$\{t^k\}_{k\in I} \to 0^+$	Convergence of a sequence $\{t^k\}_{k\in I} \subseteq \mathbb{R}_+$ to 0
$Df(x)$	Jacobian of a function f at x
$\nabla f(x)$	Gradient (transpose of the Jacobian) of a function $f\colon \mathbb{R}^n \to \mathbb{R}$ at x
$D^2 f(x)$	Second derivative of a function f at x
$\nabla^2 f(x)$	Hessian (Jacobian of the gradient) of a function $f\colon \mathbb{R}^n \to \mathbb{R}$ at x
$\partial_B f(x)$	B-subdifferential (or B-generalized Jacobian) of a function f at x
$\partial_C f(x)$	Clarke subdifferential (or generalized Jacobian) of a function f at x
$f'(x,d)$	Directional derivative of f at x along a given direction d

$\partial_i f(x)$	i-th partial derivative of f at x
L^*	Adjoint of a linear operator L
$L[v]$ or Lv	Image of a vector v by a linear operator L
$\mathrm{Im}(L)$	Image of a linear operator L
$\mathrm{Ker}(L)$	Null space of a linear operator L
$\mathrm{rank}(L)$	Rank of a linear operator L, dimension of $\mathrm{Im}(L)$
$\limsup_{k\to\infty} C^k$	Upper (or outer) limit of a sequence of sets $\{C^k\}_{k\in\mathbb{N}}$; that is, the collection of all points y such that there is an infinite subset $I \subseteq_\infty \mathbb{N}$ and a sequence $\{y^k\}_{k\in I}$, with $y^k \in C^k$ for each $k \in I$, such that $\lim_{k\in I} y^k = y$
$\liminf_{k\to\infty} C^k$	Lower (or inner) limit of a sequence of sets $\{C^k\}_{k\in\mathbb{N}}$; the collection of all y such that there exists a terminal (a "tail") subset $I \subseteq_\infty \mathbb{N}$ and a sequence $\{y^k\}_{k\in I}$, with $y^k \in C^k$ for each $k \in I$, such that $\lim_{k\in I} y^k = y$
$\limsup_{x\to\overline{x}} C(x)$	Upper limit of a set-valued mapping $x \mapsto C(x)$; the union for all sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ of $\limsup_{k\to\infty} C(x^k)$
$\liminf_{x\to\overline{x}} C(x)$	Lower limit of a set-valued mapping $x \mapsto C(x)$; the union for all sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ of $\liminf_{k\to\infty} C(x^k)$

Chapter 1

Overview

Optimization is a consolidated and interesting research field on its own, but it became popular in the last century due to its many practical applications in Business Intelligence and Engineering. In summary, whenever there is some sort of classification in a collection of objects, one can ask which one is the best one, and this is the most general form of an optimization problem. In a structured problem, this "classification" criterion is usually determined by a function f which is called an *objective function*, and this "collection" is a set Ω, called the *feasible set*, which is expected to admit a good mathematical representation, for instance, as the preimage of a simpler set $\mathcal{K}$ through a continuous function G, called a *constraint function*.

To solve optimization problems with some degree of confidence, one must first characterize its solutions to find out how to compute them. Any criterion to decide whether a given point is a (potential) solution of an optimization problem is called an *optimality condition* and the methods for computing them are called *optimization algorithms*. The main goal of this thesis is to propose new concepts and results regarding optimality conditions for a very general class of optimization problems and study their applications in studying the "reliability" of practical algorithms. Our advances were reported in a few papers [9, 11, 12, 13, 14, 15, 16, 48], some of which have already been peer reviewed and accepted for publication, and as expected these papers are the main source of inspiration for this thesis. However, it is not limited to them. Many of the concepts therein are presented here in a clearer and simpler way, from a more intuitive point of view, and some results from our papers were improved and extended to more general contexts in this thesis. On the other hand, because such papers are attached to this thesis, we took the liberty of omitting some technical proofs here and even whole parts of some papers that in our opinion could cause some sort of confusion.

1.1 A summary of our contributions

- We present in Chapter 3 an extension of the so-called *Approximate Karush-Kuhn-Tucker* (AKKT) sequential optimality condition and its variants from the context of *Nonlinear Programming* (NLP) [10] to a general *Nonlinear Conic Programming* (NCP) one, based on our paper [9]. We start with a full exposition of the main properties of AKKT, which are essentially the fact it is necessary for optimality without relying on any *Constraint Qualification* (CQ) condition, a relaxed CQ-less convergence theory of an Augmented Lagrangian-type method, and a characterization of the weakest CQ that makes AKKT equivalent to the standard KKT condition. Then, we proceed to improve such results by means of the stronger *Approximate Gradient Projection* (AGP) condition (extended from [67]) and its enhanced version called *Complementary AKKT* (CAKKT) that was originally proposed for NLP problems in [23]; in particular, we proved that the Augmented Lagrangian method produces AGP-type sequences without any additional assumption – a new and somewhat surprising result even in NLP – and CAKKT-type sequences under a mild regularity assumption, which extends an analogous result from NLP, but is new for the other particular cases of NCP. Because AGP/CAKKT and KKT are very closely related in a certain sense, this is a significant improvement of the classical convergence theory of the Augmented Lagrangian method. Moreover, in some particular cases of NCP, the AGP and CAKKT conditions have been used as a unified framework for proving convergence of several other methods (see Section 3.3.2) in a very simple language, and this degree of generality is also expected when it comes to more general NCP algorithms. In the last section of Chapter 3, we proposed an alternative way of using sequential optimality conditions: as mathematical tools for analysing the effect of some minor variants of classical algorithms that usually appear in a practical context, and this was illustrated by and Augmented Lagrangian method. We were also able to prove that being an accumulation point of a sequence generated by this method is itself an optimality condition (that is, the algorithm produces its own certificate of optimality mid-execution) and to characterize the weakest CQ required to prove its convergence to stationary (KKT) points; such results first appeared in our paper [16] for NLP, but in this thesis they are presented for the general NCP problem.

- In Chapter 4, we present the first extension of the so-called *Constant Rank Constraint Qualification* (CRCQ) [57] from NLP to NCP. Such an extension is based in a new geometric characterization of CRCQ – novel even in NLP – that was introduced in our paper [14] concerning two very popular particular cases of NCP: *Nonlinear Semidefinite Programming* (NSDP) and *Nonlinear Second-Order Cone Programming* (NSOCP). The main feature of such extension of CRCQ is the fact it induces a *Strong Second-Order Optimality Condition* (SSOC) at local minimizers of a NSDP or NSOCP problem, which holds for any Lagrange multiplier and any direction in the cone of critical directions. Of course, there are classical second-order results related to the celebrated *Nondegeneracy* CQ and the popular Robinson's CQ, but our result improves the classical ones in some aspects. In fact, perhaps the major motivation to study extensions of CRCQ is the fact they are expected to be independent of Robinson's CQ (thus complementing it) while still being strictly weaker than Nondegeneracy, which is the case of the condition presented in [14]. Noteworthy, in this thesis we improved [14] in several aspects. First, we presented a relaxed version of CRCQ which happens to be an extension of the so-called *Constant Rank of the Subspace Component* (CRSC) condition from NLP [18], and we showed how this is related to a nonlinear variant of the famous *facial reduction* preprocessing technique for linear conic programs. It is worth remarking that CRSC is strictly weaker than both Robinson's CQ and CRCQ. Then, we proposed a completely new CQ that is in-between CRCQ and CRSC, called the *Strong-CRSC* condition, which we used to obtain the same type of result on SSOC as the one of [14], but for a much more general class of problems, with simpler proofs and more general assumptions, since Strong-CRSC is also strictly weaker than CRCQ and independent of Robinson's CQ. However, there are strong reasons to believe that this type of extension is more prone to theoretical uses; to cover practical applications, we followed a completely different approach that is presented in Chapter 5.

- In Chapter 5 we are led to investigate alternative extensions of CRCQ and its variants that can be used for proving convergence of practical algorithms. To do so, we relied on a new characterization of the Nondegeneracy condition that was presented in our manuscript [13] together with an idea of using the eigenvalues of the constraint function with respect to the spectral structure of NSOCP and NSDP problems to study constant rank, from our previous paper [11]. This allowed us to extend not only CRCQ, but also one of its most relevant variants: the so-called *Constant Positive Linear Dependence* (CPLD) condition from NLP [21, 80]. These extensions were called in our papers [12] (for NSDP) and [15] (for NSOCP) as *Sequential-CRCQ* and *Sequential-CPLD*, respectively, to distinguish them from the CRCQ extension of Chapter 4; also, they are "sequential" for their natural relation with AKKT and its variants. In particular, we proved that the feasible accumulation points of any AKKT-type sequence that satisfy Sequential-CPLD (or Sequential-CRCQ) must always satisfy the KKT condition also. In other words, any algorithm that generates AKKT-type sequences admits a convergence theory based on Sequential-CPLD. This is meaningful because Sequential-CPLD is in turn strictly weaker than Robinson's CQ and Sequential-CRCQ, which is in turn independent of it. Moreover, the characterization of Nondegeneracy from [13] allowed us to develop a relaxed version of it called *Weak-Nondegeneracy*, which in turn induced "weak" forms of the Sequential-CPLD and -CRCQ conditions. Still in [13], we studied an intuitive way of dealing with structural sparsity in the constraints of a NSDP problem, similarly to what Forsgren had done in [47]. Although dealing with structural sparsity is somewhat trivial in NLP and NSOCP, this is not necessarily the case of NSDP problems; curiously, such a treatment is also closely related to another analogue of the facial reduction technique we mentioned before, but no relation with CRSC has been found yet. We end Chapter 5 by deriving an amalgam of Sequential-CRCQ and the CRCQ from Chapter 4, which is not a CQ on its own, but when paired with Robinson's CQ is able to induce a Weak Second-Order Optimality Condition that seems to be the only one with applications in the convergence of algorithms (we also provide a careful explanation of why).

1.2 How to navigate through this thesis

This thesis is designed to be as self-contained as possible, but also as short and easy-to-read as possible. Most classical theorems and definitions needed for our own results are presented in Chapter 2, most of them without proofs. The intricate results, however, are presented with expanded proofs for the sake of understanding[1]. All other chapters will refer to this one from time to time.

Chapters 3 and 4 only depend on Chapter 2, but they are independent of each other. Chapter 5, on the other hand, depends on both Chapters 3 and 4. It is worth emphasizing that this thesis is a compendium of some papers that are attached to it. Here is a relational list for quick reference.

Chapter 3 is based on:

1. *On optimality conditions for nonlinear conic programming* (with R. Andreani, W. Gómez, G. Haeser, and A. Ramos). To appear in Mathematics of Operations Research, 2022. DOI: 10.1287/moor.2021.1203. See Appendix A.

2. *On the best achieavable quality of limit points of augmented Lagrangian schemes* (with R. Andreani, G. Haeser, A. Ramos, and L. D. Secchin). To appear in Numerical Algorithms, 2022. DOI: 10.1007/s11075-021-01212-8. See Appendix B.

[1] *Personal note:* My own understanding, I mean, though I'd be very happy should any of these expanded proofs help new students, since I struggled a lot to understand them for the first time.

Chapter 4 is based on:

3. *First- and second-order optimality conditions for second-order cone and semidefinite programming under a constant rank condition.* (with R. Andreani, G. Haeser, H. Ramírez, and T. P. Silveira). Under review. ArXiv: 2107.04693. See Appendix C.

Chapter 5 is based on:

4. *Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming* (with R. Andreani, G. Haeser, H. Ramírez, D. O. Santos, and T. P. Silveira). To appear in Optimization Letters, 2022. DOI: 10.1007/s11590-021-01737-w. See Appendix D.

5. *On the weak second-order optimality condition for nonlinear semidefinite and second-order cone programming* (with E. H. Fukuda and G. Haeser). Under review. Optimization Online: 7951. See Appendix E.

6. *Weak notions of nondegeneracy in nonlinear semidefinite programming* (with R. Andreani, G. Haeser, and H. Ramírez). Under review. ArXiv: 2012.14810. See Appendix F.

7. *Sequential constant rank constraint qualifications for nonlinear semidefinite programming with applications* (with R. Andreani, G. Haeser, and H. Ramírez). Under review. ArXiv: 2106.00775. See Appendix G.

8. *Global convergence of algorithms under constant rank conditions for nonlinear second-order cone programming* (with R. Andreani, G. Haeser, H. Ramírez, and T. P. Silveira). Under review. ArXiv: 2110.12015. See Appendix H.

There is no conclusion chapter in this thesis, but the last section of each chapter contains a quick summary of the content presented in it. Some ideas for future works are spread throughout the main text in footnotes and other proposals can be found in the conclusion sections of our papers.

Chapter 2

Fundamentals of optimization theory

In this chapter, we present several fundamental results of Nonlinear (Conic) Optimization, Linear Algebra, and Convex Analysis, to support our own results. If you are already familiarized with these concepts, we encourage you to skip this chapter. For the sake of simplicity, we do not prove well-known (i.e. classical) results; instead, we point the reader towards a proper reference. Moreover, for the reader's convenience, each section of this chapter is divided in tiny modules to allow quick navigation.

2.1 Basic structures, convex sets, and cones

Let $\mathbb{Y}$ be an arbitrary finite-dimensional linear space over $\mathbb{R}$, equipped with an inner product $\langle\cdot,\cdot\rangle$ and its induced norm $\|\cdot\| \doteq \sqrt{\langle\cdot,\cdot\rangle}$. A subset C of $\mathbb{Y}$ is said to be *affine* when the (unique) *line that passes through* every two points $x_1 \in C$ and $x_2 \in C$, expressed as $\{\alpha_1 x_1 + \alpha_2 x_2 \colon \alpha_1 + \alpha_2 = 1\}$, is entirely contained in C. Moreover, given any set $S \subseteq \mathbb{E}$, the collection of all (finite) *affine combinations* of its elements, denoted by

$$\mathrm{aff}(S) \doteq \left\{ \sum_{i=1}^{s} \alpha_i x_i \colon s \in \mathbb{N},\ \forall i \in \{1,\ldots,s\}, x_i \in S,\ \text{and}\ \sum_{i=1}^{s} \alpha_i = 1 \right\},$$

is called the *affine hull* of S, which coincides with the smallest affine subset of $\mathbb{E}$ (by inclusion) that contains S. In particular, if S is affine, then $\mathrm{aff}(S) = S$. Roughly speaking, affine sets are invariant to the addition of lines crossing their elements. Moreover, C is *convex* when the *line segment that joins* every two points $x_1 \in C$ and $x_2 \in C$,

$$\{\alpha_1 x_1 + \alpha_2 x_2 \colon \alpha_1 \geqslant 0, \alpha_2 \geqslant 0, \alpha_1 + \alpha_2 = 1\},$$

belongs to C. Observe that *convexity* is a similar notion to *affinity* but it describes sets that are invariant to the addition of line segments between their elements, instead of full lines. The collection of all *convex combinations* of the elements of a given set S,

$$\mathrm{conv}(S) \doteq \left\{ \sum_{i=1}^{s} \alpha_i x_i \colon s \in \mathbb{N},\ \forall i \in \{1,\ldots,s\}, \alpha_i \geqslant 0,\ x_i \in S,\ \text{and}\ \sum_{i=1}^{s} \alpha_i = 1 \right\},$$

is known as the *convex hull* of S, and of course, $\mathrm{conv}(S)$ is also the smallest convex subset of $\mathbb{E}$ that contains S. Also, $\mathrm{conv}(S) = S$ when S is convex.

It is worth pointing out that every affine set C is a translation of a (unique) linear subspace, namely $C - x$ for any $x \in C$, and as such the *dimension* of C is usually defined as the dimension of its associate subspace. Furthermore, we will often abuse this terminology to talk about the dimension of an arbitrary set, which will in turn be regarded as the dimension of its affine hull; that is,

$$\dim(C) \doteq \dim(\mathsf{aff}(C)) = \dim(\mathsf{aff}(C) - x),$$

where $x \in \mathsf{aff}(C)$. For instance, points have dimension 0, line segments have dimension 1, and squares have dimension 2. Similarly, as an abuse of notation, we will define the orthogonal complement of any set $C \subseteq \mathbb{Y}$ as the orthogonal complement of its linear span, that is,

$$C^{\perp} \doteq \mathsf{span}(C)^{\perp}.$$

2.1.1 Relative interiors and a separation theorem

Definition 2.1.1 (Relative interior)**.** The *relative interior* of a set $C \subseteq \mathbb{E}$, denoted by $\mathsf{ri}(C)$, is the collection of all points $x \in C$ such that $B(x, \varepsilon) \cap \mathsf{aff}(C) \subseteq C$ for some $\varepsilon > 0$.

The relative interior, which is one of the most useful concepts in Convex Analysis, essentially consists of the interior of a set restricted to its affine hull. It allows us to study the "relative topology" of sets embedded in a larger dimensional topological space, such as a line segment in $\mathbb{R}^2$ or a (flat) circle embedded in $\mathbb{R}^3$. Following Rockafellar's book [81, Section 6], let us recall a couple of useful properties of relative interiors:

Theorem 2.1.1 (Theorem 6.1 of [81])**.** *Let $C \subseteq \mathbb{Y}$ be a convex set, and let $x_1 \in \mathsf{ri}(C)$ and $x_2 \in \mathrm{cl}(C)$. Then, $(1-\alpha)x_1 + \alpha x_2 \in \mathsf{ri}(C)$ for every $\alpha \in [0, 1)$.*

Further, if C is nonempty, then $\mathsf{ri}(C)$ is also nonempty; and if C is convex, then $\mathsf{ri}(C)$ is also convex and it holds that

$$\mathrm{cl}(\mathsf{ri}(C)) = \mathrm{cl}(C) \text{ and } \mathsf{ri}(\mathrm{cl}(C)) = \mathsf{ri}(C).$$

Another important fact to keep in mind is that $C_1 \subseteq C_2$ does *not* necessarily imply $\mathsf{ri}(C_1) \subseteq \mathsf{ri}(C_2)$, although this holds true for usual interiors. A simple counterexample comes by letting C_2 be a regular triangle in $\mathbb{R}^2$ and C_1 be one of its sides; then, it is easy to see that $C_1 \subseteq C_2$, but $\mathsf{ri}(C_1) \cap \mathsf{ri}(C_2) = \emptyset$. What is true is the following:

Lemma 2.1.2 (Theorem 6.5 of [81])**.** *Let $C_1, C_2 \subseteq \mathbb{Y}$ be convex sets such that $\mathsf{ri}(C_1) \cap \mathsf{ri}(C_2) \neq \emptyset$. Then $\mathsf{ri}(C_1 \cap C_2) = \mathsf{ri}(C_1) \cap \mathsf{ri}(C_2)$ and if, in addition, $C_1 \subseteq \mathrm{cl}(C_2)$ then $\mathsf{ri}(C_1) \subseteq \mathsf{ri}(C_2)$.*

Another useful information about relative interiors to our analyses is their behavior under a linear transformation, which can be summarized as follows:

Theorem 2.1.3 (Theorems 6.6 and 6.7 of [81])**.** *Let $\mathbb{X}$ and $\mathbb{Y}$ be finite-dimensional linear spaces, $C \subseteq \mathbb{Y}$ be a convex set, and let $U\colon \mathbb{X} \to \mathbb{Y}$ and $V\colon \mathbb{Y} \to \mathbb{X}$ be linear operators. Then:*

1. *$V[\mathsf{ri}(C)] = \mathsf{ri}(V[C])$;*
2. *If $U^{-1}(\mathsf{ri}(C)) \neq \emptyset$, then $U^{-1}(\mathsf{ri}(C)) = \mathsf{ri}(U^{-1}(C))$.*

In particular, note that $\mathsf{ri}(\alpha C) = \alpha \mathsf{ri}(C)$ for every $\alpha \in \mathbb{R}$. Moreover, given any two convex sets $C_1, C_2 \subseteq \mathbb{Y}$ one has $\mathsf{ri}(C_1 \times C_2) = \mathsf{ri}(C_1) \times \mathsf{ri}(C_2)$ and as a consequence $\mathsf{ri}(C_1 + C_2) = \mathsf{ri}(C_1) + \mathsf{ri}(C_2)$.

There are many relevant results in optimization regarding the existence of special objects, that can be derived from a simple idea such as *separability*. Given two nonempty convex sets

$C_1, C_2 \subseteq \mathbb{Y}$, we say that they can be *separated* when there exists some nonzero $w \in \mathbb{Y}$ and some $\gamma \in \mathbb{R}$ such that

$$\forall x \in C_1,\ \langle w, x\rangle \geqslant \gamma \quad \text{and} \quad \forall y \in C_2,\ \langle w, y\rangle \leqslant \gamma.$$

Observe that every nonzero vector $w \in \mathbb{Y}$ defines a *hyperplane* – that is, a $(\dim(\mathbb{Y}) - 1)$-dimensional affine space – in the form

$$H \doteq \{x \in \mathbb{Y} \colon \langle x, w\rangle = \gamma\}, \tag{2.1}$$

that has w as its defining normal vector. Also, notice that H divides $\mathbb{Y}$ in two half-spaces:

$$H_+ \doteq \{x \in \mathbb{Y} \colon \langle x, w\rangle \geqslant \gamma\} \quad \text{and} \quad H_- \doteq \{x \in \mathbb{Y} \colon \langle x, w\rangle \leqslant \gamma\}, \tag{2.2}$$

and that H separates C_1 and C_2 by letting $C_1 \subseteq H_+$ and $C_2 \subseteq H_-$. However, this weak definition of separability allows, for instance, C_1 and C_2 to be contained in H. For this reason, one may be interested in considering stronger notions of separability:

Definition 2.1.2 (Proper and strong separability)**.** Let $C_1, C_2 \subseteq \mathbb{Y}$ be convex sets.

- We say that C_1 and C_2 can be *properly separated* if there exists a hyperplane H that separates C_1 and C_2, but $C_1 \not\subseteq H$ or $C_2 \not\subseteq H$;
- We say that C_1 and C_2 can be *strongly separated* if there exists a hyperplane H and some $\varepsilon > 0$ such that $C_1 + B[0, \varepsilon] \subseteq H_+$ and $C_2 + B[0, \varepsilon] \subseteq H_-$, where $B[0, \varepsilon] \doteq \{x \in \mathbb{Y} \colon \|x\| \leqslant \varepsilon\}$ is the closed ball[1] with radius ε at the origin and H_+ and H_- are as in (2.2).

We are now ready to present the classical *strong separation theorem*:

Theorem 2.1.4 (Theorem 11.3 and Corollary 11.4.2 of [81])**.** *Let $C_1, C_2 \subseteq \mathbb{Y}$ be nonempty convex sets. Then, C_1 and C_2 can be properly separated if, and only if,* ri$(C_1) \cap$ ri$(C_2) \neq \emptyset$*. Moreover, if $\mathrm{cl}(C_1) \cap \mathrm{cl}(C_2) = \emptyset$ and either C_1 or C_2 is bounded, then these sets can be strongly separated.*

As a corollary, it follows that every nonempty closed convex set $C \subseteq \mathbb{Y}$ is the intersection of all half-spaces containing it [81, Theorem 11.5]. Indeed, if there was any point $x \notin C$ in the intersection of all half-spaces that contain C, then x and C would be strongly separated by a hyperplane H such that $x \in H_+$ and $C \subseteq H_-$ but $x \notin H$, which is absurd since $x \in H_-$ by construction. This is an appropriate moment to recall that hyperplanes H such that $C \subseteq H_+$ or $C \subseteq H_-$ (that is, such that $H \cap$ ri$(C) = \emptyset$) are called *supporting hyperplanes* to C and their associated half-spaces that contain C are called *supporting half-spaces*.

2.1.2 Cones and variants of Carathéodory's Lemma

We say that a nonempty subset C of $\mathbb{Y}$ is a *cone* when $\alpha x \in C$, for every $x \in C$ and every $\alpha \in \mathbb{R}$ such that $\alpha > 0$. Some popular examples of cones, which will be widely studied in this thesis, are:

- The *nonnegative orthant*, defined as

$$\mathbb{R}^m_+ \doteq \{x \in \mathbb{R}^m \colon \forall i \in \{1, \ldots, m\},\ x_i \geqslant 0\};$$

[1] As pointed out by Patrick Mehlitz, strong separation can be equivalently defined with an open ball $B(0, \varepsilon)$.

- The *positive semidefinite cone*, which can be defined as

$$\mathbb{S}^m_+ \doteq \{M \in \mathbb{S}^m \colon \forall d \in \mathbb{R}^m,\ d^\top M d \geqslant 0\};$$

- The *Lorentz cone* (or *ice-cream cone*, or *second-order cone*), given by

$$\mathbb{L}^m \doteq \{(x_0, \widehat{x}) \in \mathbb{R} \times \mathbb{R}^{m-1} \colon x_0 \geqslant \|\widehat{x}\|_2\}.$$

Another important example of cone that shall be recalled several times throughout this thesis is the *polar cone* that can be defined for any given set $C \subseteq \mathbb{Y}$ as follows:

$$C^\circ \doteq \{y \in \mathbb{Y} \colon \forall x \in C,\ \langle y, x\rangle \leqslant 0\}.$$

The set C° is always a closed convex cone, even if C is not a cone or not closed[2]. If C is itself a cone, then it holds that $(C^\circ)^\circ = \mathrm{cl}(C)$, and if C is also closed, then of course $(C^\circ)^\circ = C$. Directly from the definition, one can also verify that if C is a linear subspace, then $C^\circ = C^\perp$, meaning that the polar cone generalizes the orthogonal complement. In addition, a cone C is called *self-dual* when $C^\circ = -C$. For instance, the cones we mentioned before: $\mathbb{R}^m_+$, $\mathbb{S}^m_+$, and $\mathbb{L}^m$, are all self-dual in this sense, whereas the singleton $\{0\}$ is not. Moreover, given two cones $C_1, C_2 \subseteq \mathbb{Y}$, the relations

$$(C_1 + C_2)^\circ = C_1^\circ \cap C_2^\circ \quad \text{and} \quad (C_1 \cap C_2)^\circ = \mathrm{cl}(C_1^\circ + C_2^\circ)$$

hold true.

Note that if C is a cone, then $\mathsf{aff}(C) = \mathsf{span}(C)$, and C is called *full-dimensional* when $\dim(C) = \dim(\mathsf{span}(C)) = \dim(\mathbb{Y})$. If $0 \in C$, then $\mathsf{span}(C) = C - C$. On the other side of the playground, the largest subspace contained in C, called its *lineality space* and denoted by $\mathrm{lin}(C)$, satisfies the relations

$$\mathrm{lin}(C) = C \cap (-C) = \mathsf{span}(C^\circ)^\perp \tag{2.3}$$

provided C is closed and convex. When $\mathrm{lin}(C) = \{0\}$ the cone C is called *pointed*, otherwise it is said to *contain lines*. It is immediate from (2.3) that C is pointed if, and only if, C° is full-dimensional. Also, every pointed convex cone induces a partial order $\preceq_C$, defined by the relation $x \preceq_C y \ \Leftrightarrow\ y - x \in C$.

A useful fact to keep in mind is that if C is a nonempty closed convex cone, then every point $w \in C^\circ$ defines a supporting hyperplane to C in the form (2.1) with $\gamma = 0$ and, moreover, C is the intersection of all supporting half-spaces associated with these hyperplanes – see also [81, Corollary 11.7.1].

Given any set $S \subseteq \mathbb{Y}$, its *conic hull* (i.e. the smallest convex cone that contains S) can be characterized as

$$\mathsf{cone}(S) \doteq \left\{ \sum_{i=1}^{s} \alpha_i x_i \colon s \in \mathbb{N},\ \forall i \in \{1, \ldots, s\}, \alpha_i \geqslant 0,\ x_i \in S \right\},$$

which is precisely the union of all *half-lines* starting at the origin that intersect $\mathsf{conv}(S)$.

Example 2.1.1 (Polyhedral cones). *A* polyhedral cone *is the intersection of finitely many half-spaces in the form* $\{y \in \mathbb{Y} \colon \langle y, w\rangle \geqslant 0\}$ *where* $w \in \mathbb{Y}$ *is nonzero. A famous result known as*

[2] It is important to remark that the *polar cone* of a set C is *not* the same as the *polar set* of C, given by $C^\diamond \doteq \{y \in \mathbb{Y} \colon \forall x \in C,\ \langle y, x\rangle \leqslant 1\}$, unless C is a closed convex cone. Every incidence of the word "polar" (of C) in this thesis, without further specification, refers to C° and not $C^\diamond$.

Weyl-Minkowski Theorem for cones *(see, for instance, [44, Theorem 1.3.12 and Corollary 1.3.13]) tells us that a cone C is polyhedral if, and only if,*

$$C = \mathsf{cone}(\{x_1, \ldots, x_s\}) = \left\{ \sum_{i=1}^{s} \alpha_i x_i \colon \forall i \in \{1, \ldots, s\}, \alpha_i \geqslant 0 \right\}. \tag{2.4}$$

for some $x_1, \ldots, x_s \in \mathbb{Y}$; that is, if C is finitely generated*. For instance, $\mathbb{R}^m_+$ is polyhedral since it coincides with* cone$(\{e_1, \ldots, e_m\})$*, where e_i is the i-th vector of the canonical basis of $\mathbb{R}^m$, while $\mathbb{S}^m_+$ and $\mathbb{L}^m$ are not. As a very important remark, following [54, Example 3.2.2], we see that if C has the form (2.4), then*

$$C^{\circ} = \{y \in \mathbb{Y} \colon \forall i \in \{1, \ldots, s\}, \langle x_i, y \rangle \leqslant 0\},$$

which means that polars of polyhedral cones are also polyhedral.

Now, let us discuss about representation of the elements of a cone in terms of special bases. The first result of such kind that we recall is the celebrated *Carathéodory's Lemma* (for cones), as characterized by Bertsekas [31, Exercise B.1.7]:

Lemma 2.1.5 (Carathéodory's Lemma)**.** *Let $x_1, \ldots, x_s \in \mathbb{Y}$, and let $\alpha_1, \ldots, \alpha_s \in \mathbb{R}$ be arbitrary. Then, there exists some subset $J \subseteq \{1, \ldots, s\}$ and some scalars $\tilde{\alpha}_j$ with $j \in J$, such that $\{x_j\}_{j \in J}$ is linearly independent,*

$$\sum_{j=1}^{s} \alpha_j x_j = \sum_{j \in J} \tilde{\alpha}_j x_j,$$

and $\alpha_j \tilde{\alpha}_j > 0$, for all $j \in J$.

Carathéodory's Lemma 2.1.5 is the conic analogue (and a consequence) of a well-known result with the same name that states that every point in the convex hull of a subset S of $\mathbb{Y}$ can be written as a combination of at most $\dim(\mathbb{Y}) + 1$ points of S. Lemma 2.1.5 states that any point of the conic hull of $\{x_1, \ldots, x_s\}$ can be represented using only a linear independent subset of it while keeping the sign of its correspondent scalars. However, note that some of these scalars may be zero, meaning that they are somehow superfluous for this particular representation. This leads us to a slightly different version of Carathéodory's Lemma that shall be useful later on:

Lemma 2.1.6. *Let C be a closed convex cone such that $s \doteq \dim(C) \geqslant 1$ and let $x \in$ ri(C) be nonzero. Then, there exist some linearly independent vectors $x_1, \ldots, x_s \in C$ and scalars $\alpha_1 > 0, \ldots, \alpha_s > 0$, such that*

$$x = \sum_{i=1}^{s} \alpha_i x_i.$$

Proof. Let $y_1, \ldots, y_{s-1}$ be a basis of span$(C) \cap \{x\}^{\perp}$ and define $y_s \doteq -(y_1 + \ldots + y_{s-1})$. Now, let $\varepsilon > 0$ be such that $x_i \doteq x + \varepsilon y_i \in$ ri(C) for every $i \in \{1, \ldots, s\}$ and note that $x_1, \ldots, x_s$ are linearly independent, since

$$\beta_1 x_1 + \ldots + \beta_s x_s = (\beta_1 + \ldots + \beta_s)x + \varepsilon(\beta_1 - \beta_s)y_1 + \ldots + \varepsilon(\beta_{s-1} - \beta_s)y_{s-1} = 0$$

implies $\beta_1 = \ldots = \beta_{s-1} = \beta_s = 0$. Let $\alpha_i \doteq 1/s > 0$ for every $i \in \{1, \ldots, s\}$ and the proof is complete[3]. □

[3] Thanks to Joe Higgins of Math.StackExchange for this ingenious proof.

Observe that Lemma 2.1.6 is very intuitive since it basically tells us that every element in the relative interior of C can be written as a positive combination of a basis of span(C) that can be found inside the cone itself; in particular, it is a trivial statement when $C = \mathbb{R}^m_+$.

We end this section with a very well known result about images of closed convex cones by linear mappings, in a weak form, as stated in the introduction of Pataki [75].

Proposition 2.1.7. *Let $C \subseteq \mathbb{Y}$ be a closed convex cone and let $U\colon \mathbb{Y} \to \mathbb{X}$ be a linear mapping, where $\mathbb{X}$ is also finite-dimensional, and let U^* denote the adjoint of U. If* $\mathrm{Im}(U) \cap \mathrm{ri}(C) \neq \emptyset$, *then $U^*[C^\circ]$ is closed. Alternatively, if C is a polyhedral cone, then $U^*[C^\circ]$ is closed also.*

This kind of result that describes situations when linear images of closed convex cones or sets are also closed have shown to be very useful in several fields of mathematics [75], including optimization, where it plays a central role.

2.1.3 Faces and minimal faces of convex sets

Let $C \subseteq \mathbb{Y}$ be a convex set and let F be a convex subset of C. Following Rockafellar's book [81, Section 18], we say that F is a *face* of C if for every $x \in F$ and every $y, w \in C$ such that $x = \alpha y + (1-\alpha)w$ for some $\alpha \in (0,1)$, we have that $y, w \in F$. Moreover, we use the standard notation $F \trianglelefteq C$ to say that F is a face of C. Loosely speaking, a face $F \trianglelefteq C$ is essentially an "extreme flat convex subset of C" and, in particular, zero-dimensional faces are called *extreme points*. By definition, the empty set and C itself are both faces of C.

Let us recall some basic properties of faces from [81, Section 18]:

Lemma 2.1.8. *Let $C \subseteq \mathbb{Y}$ be a convex set. Then:*

1. *If $F_1 \trianglelefteq F_2 \trianglelefteq C$, then $F_1 \trianglelefteq C$;*
2. *If $F \trianglelefteq C$ and D is a convex set satisfying $F \subseteq D \subseteq C$, then $F \trianglelefteq D$;*
3. *$F \trianglelefteq C$ if, and only if, $F = C \cap \mathrm{aff}(F)$;*
4. *If $F \trianglelefteq C$ and $D \subseteq C$ is a convex set such that $\mathrm{ri}(D) \cap F \neq \emptyset$, then $D \subseteq F$;*
5. *If $F_1 \trianglelefteq C$ and $F_2 \trianglelefteq C$ are such that $\mathrm{ri}(F_1) \cap \mathrm{ri}(F_2) \neq \emptyset$, then $F_1 = F_2$.*

Proof. Items 1 and 2 follow directly from the definition of face. Item 3 was extracted from [39, Exercise 5.4]. Items 4 and 5 are restatements of [81, Theorem 18.1 and Corollary 18.1.2], respectively. □

In view of item 3 of Lemma 2.1.8, it is important to remark that, in general, not every face F of C can be written in the form $F = C \cap H$ for some hyperplane H, although every set in this form is a face of C. Such faces are called *exposed* and Figure 2.1 illustrates a set $C \subseteq \mathbb{R}^2$, which is the union of a (green) circle and a (blue) square, with non-exposed faces given by the intersection of the boundaries of these sets (red points). All of its remaining faces, however, are exposed.

We now reproduce a couple of useful results regarding intersections and products of faces of closed convex cones from a paper by Lourenço et al. [65]:

Lemma 2.1.9. *Let $C \subseteq \mathbb{Y}$ be a nonempty closed convex cone and let $F \trianglelefteq C$.*

1. *If $C \doteq C_1 \cap C_2$ for some closed convex cones $C_1, C_2 \subseteq \mathbb{Y}$, then there exist faces $F_1 \trianglelefteq C_1$ and $F_2 \trianglelefteq C_2$ such that $F = F_1 \cap F_2$ and $\mathrm{ri}(F) \subseteq \mathrm{ri}(F_1) \cap \mathrm{ri}(F_2)$;*

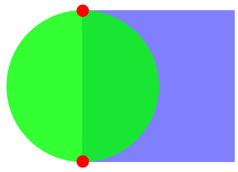


Figure 2.1: *Convex set with non-exposed faces.*

2. *If* $\mathbb{Y} \doteq \mathbb{Y}_1 \times \mathbb{Y}_2$ *and* $C \doteq C_1 \times C_2$ *for some closed convex cones* $C_1 \subseteq \mathbb{Y}_1$ *and* $C_2 \subseteq \mathbb{Y}_2$, *then there exist faces* $F_1 \trianglelefteq C_1$ *and* $F_2 \trianglelefteq C_2$ *such that* $F = F_1 \times F_2$.

Proof. Item 1 is proved in [65, Proposition 2.2]. For item 2, note that

$$C = (C_1 \times \mathbb{Y}_2) \cap (\mathbb{Y}_1 \times C_2),$$

and by item 1 there is a face $S_1 \trianglelefteq C_1 \times \mathbb{Y}_2$ and a face $S_2 \trianglelefteq \mathbb{Y}_1 \times C_2$ such that $F = S_1 \cap S_2$. For each $i \in \{1, 2\}$, let π_i be the projection mapping of $\mathbb{Y}_1 \times \mathbb{Y}_2$ onto $\mathbb{Y}_i$ and it is elementary to see that $F_i \doteq \pi_i(S_i)$ is a face of C_i. Moreover, observe that $S_1 = F_1 \times \mathbb{Y}_2$ and $S_2 = \mathbb{Y}_1 \times F_2$, which yields

$$F = (F_1 \times \mathbb{Y}_2) \cap (\mathbb{Y}_1 \times F_2) = (F_1 \cap \mathbb{Y}_1) \times (\mathbb{Y}_2 \cap F_2) = F_1 \times F_2,$$

thus completing the proof[4]. □

For any given convex subset $S \subseteq C$, the *minimal face* associated with S, denoted by $F_{\min}(S)$, is defined as the smallest (by inclusion) face of C that contains S. Let us present two very useful properties about minimal faces that were collected and presented in a seminal paper by Pataki [74].

Proposition 2.1.10 (Proposition 3.2.2 of [74]). *Let* C *be a closed convex cone,* $F \trianglelefteq C$, *and* $S \subseteq C$ *be convex. Then:*

1. $F = F_{\min}(S)$ *if, and only if,* $\mathrm{ri}(F) \cap \mathrm{ri}(S) \neq \emptyset$;
2. *If* $S \subseteq F \trianglelefteq C$ *and* $S \cap \mathrm{ri}(F) \neq \emptyset$, *then* $F = F_{\min}(S)$.

In particular, when $S \doteq \{x\}$ is a singleton, then $F_{\min}(S)$ is the unique face of C that contains x in its relative interior.

2.1.4 Metric projection and Moreau's decomposition

For any closed set $C \subseteq \mathbb{Y}$ and any point $x \in \mathbb{Y}$, consider the set

$$\Pi_C(x) \doteq \{y \in C\colon \|x - y\| = \mathrm{dist}(x, C)\},$$

of all points of C that are closest to x, where $\mathrm{dist}(x, C) \doteq \inf\{\|x - y\|\colon y \in C\}$ is the distance between x and C. This is called the *metric projection* (or *orthogonal projection*) of x onto C, and a result originally proved by Bunt [40] tells us that $\Pi_C(x)$ is a singleton for all $x \in \mathbb{Y}$ if, and only if, C is convex. This means that Π_C is well-defined as a function of x when C is convex, and since $\mathrm{dist}(\,\cdot\,, C)$ is continuous in this case, so is Π_C. It also means that $\Pi_C(x) = x$ if, and only if, $x \in C$; that is, the set of all fixed points of Π_C is C itself.

Now, let us recall one of the most useful basic properties of the metric projection:

[4]This nice proof of such a classical result was presented to us by Bruno F. Lourenço, of the Institute of Statistical Mathematics, in a personal communication. Thank you, Bruno!

Theorem 2.1.11 (Theorem 3.1.1 of [54]). *Let $C \subseteq \mathbb{Y}$ be a closed convex set and let $x \in \mathbb{Y}$ and $y \in C$. Then*

$$y = \Pi_C(x) \text{ if, and only if, } \langle x - y, z - y \rangle \leqslant 0 \text{ for all } z \in C. \tag{2.5}$$

Besides providing some geometric intuition on projections, Theorem 2.1.11 tells us that if C is a linear subspace, then $\Pi_C(x)$ is a linear operator because in this case the inequality in (2.5) holds with equality. The converse statement is also true.

Proposition 2.1.12 (Proposition 3.1.3 of [54]). *Let $C \subseteq \mathbb{Y}$ be a closed convex set and let $x_1, x_2 \in \mathbb{Y}$. Then,*

$$\|\Pi_C(x_1) - \Pi_C(x_2)\|^2 \leqslant \langle \Pi_C(x_1) - \Pi_C(x_2),\ x_1 - x_2 \rangle. \tag{2.6}$$

An immediate consequence from Proposition 2.1.12 after applying Cauchy-Schwarz inequality to the right-hand term of (2.6) is that

$$\|\Pi_C(x_1) - \Pi_C(x_2)\| \leqslant \|x_1 - x_2\| \tag{2.7}$$

for every $x_1, x_2 \in \mathbb{Y}$, meaning that Π_C is *nonexpansive*. Furthermore, since Π_C is continuous its "square", $\|\Pi_C(\,\cdot\,)\|^2$, is continuously differentiable. Let us now recall a classical result of Zarantonello [90], as stated by Fitzpatrick and Phelps [46], which computes the derivative of the square of Π_C:

Proposition 2.1.13 (Proposition 2.2 of [46]). *Let $C \subseteq \mathbb{Y}$ be a closed convex set. Then, for every $x \in \mathbb{Y}$,*

$$\nabla \|\Pi_C(x)\|^2 = 2\Pi_C(x).$$

Projections onto closed convex *cones* have a very special property that shall be frequently employed throughout the main results of this thesis, which is known as *Moreau's decomposition*. Let us recall it as stated in [54, Proposition 3.2.3 and Theorem 3.2.5]:

Theorem 2.1.14 (Moreau's decomposition). *Let $C \subseteq \mathbb{Y}$ be a closed convex cone and let $x \in \mathbb{Y}$. Then*

$$y = \Pi_C(x) \text{ if, and only if, } y \in C,\ \ x - y \in C^\circ, \text{ and } \langle x - y, y \rangle = 0. \tag{2.8}$$

Consequently, every element $x \in \mathbb{Y}$ can be decomposed as follows:

$$x = \Pi_C(x) + \Pi_{C^\circ}(x). \tag{2.9}$$

Several pieces of information can be extracted from Theorem 2.1.14. For instance, let C be a cone and observe that $\Pi_{C^\circ}(x) = 0$ if, and only if, $x \in C$; and moreover, that $\|\Pi_{C^\circ}(x)\| = \operatorname{dist}(x, C)$. This fact is especially interesting from the numerical point of view in cases where $\Pi_{C^\circ}(x)$ can be computed, because then it can be used to measure the distance between any given point and C. Also, it follows from (2.8) that for every $x \in \mathbb{Y}$,

$$\Pi_C(\alpha x) = \alpha \Pi_C(x), \text{ for all } \alpha \geqslant 0 \tag{2.10}$$

and also that $\Pi_C(-x) = -\Pi_{-C}(x)$.

2.1.5 Tangent sets and second-order regularity

Mathematical objects that present some sort of linearity tend to be simple, meaning we can understand them. A nonlinear object, on the other hand, may be intricate enough so that the best we can do is to study a grotesque simplification of it, obtained by means of a "linear approximation" in some sense. There are many different notions of what it means to approximate something linearly, but perhaps the most recurrent one is the idea of *tangency*. Although tangent sets (and cones) have an extensive history that touches different fields of mathematics, which eventually yielded several meaningful definitions of tangency, not all of them are equivalent, depending on the context. Let us recall a few of them:

Definition 2.1.3 (Outer, inner, and radial tangent cones). The *Bouligand tangent cone* (also known as *outer tangent cone* and *contingent cone*) to a closed set $C \subseteq \mathbb{Y}$ at a point $x \in C$ is defined as

$$\begin{aligned} T_C(x) &\doteq \left\{ d \in \mathbb{Y} \colon \exists \{\alpha^k\}_{k\in\mathbb{N}} \to 0^+,\ \mathsf{dist}(x + \alpha_k d, C) = o(\alpha_k) \right\} \\ &= \left\{ d \in \mathbb{Y} \colon \exists \{\alpha^k\}_{k\in\mathbb{N}} \to 0^+,\ \exists \{d^k\}_{k\in\mathbb{N}} \to d,\ \forall k \in \mathbb{N},\ x + \alpha^k d^k \in C \right\} \end{aligned} \tag{2.11}$$

The *inner tangent cone* to C at x is given by

$$T_C^{in}(x) = \{ d \in \mathbb{Y} \colon \mathsf{dist}(x + td, C) = o(t),\ t > 0 \}. \tag{2.12}$$

And the *radial cone* to C at x is given by

$$R_C(x) \doteq \{ d \in \mathbb{Y} \colon \exists \tilde{\alpha} > 0,\ \forall \alpha \in [0, \tilde{\alpha}],\ x + \alpha d \in C \}. \tag{2.13}$$

To give some intuition on these cones, observe that the radial cone is simply the set of *directions that point inside* C at x, while the inner and outer tangent cones are sets of *limiting interior directions* to C at x, in the sense that they both consist of limits of sequences of directions that point towards C; roughly speaking, what makes them different is that they consider different notions of limit ($\liminf$ and $\limsup$, respectively). In particular, note that $R_C(x) \subseteq T_C^{in}(x) \subseteq T_C(x)$ for every C and every $x \in C$, but the reverse inclusions are not always true. In fact, it is worth pointing out that $T_C(x)$ and $T_C^{in}(x)$ are always closed, whereas $R_C(x)$ may be not closed. In the presence of convexity, however, the inner and outer notions of tangency coincide with the closure of the radial cone and they are all convex.

Proposition 2.1.15 (Proposition 2.55 of [37]). *Let $C \subseteq \mathbb{Y}$ be a closed convex set and let $x \in C$. Then,*

$$R_C(x) = \mathsf{cone}(C - x) \quad \text{and} \quad T_C(x) = T_C^{in}(x) = \mathrm{cl}(R_C(x)).$$

The polar of the (outer) tangent cone to a closed convex set $C \subseteq \mathbb{Y}$ at $x \in \mathbb{Y}$, denoted by $N_C(x) \doteq T_C(x)^\circ$, is usually called the *normal cone* to C at x. By definition, note that $T_C(x) = N_C(x) = \emptyset$ when $x \notin C$. Normal cones play a pivotal role in optimization theory and in some cases $N_C(x)$ has a friendly characterization, as we will see next:

Proposition 2.1.16 (Equation 2.310 of [37]). *Let $C \subseteq \mathbb{Y}$ be a closed convex set and let $x \in C$. Then,*

$$N_C(x) = \{ y \in \mathbb{Y} \colon \forall z \in C,\ \langle z - x, y \rangle \leqslant 0 \}.$$

In addition, if C is a closed convex cone, then

$$N_C(x) = \{ y \in C^\circ \colon \langle x, y \rangle = 0 \}.$$

Loosely speaking, tangent cones describe the best first-order approximation of a set at a given point. Second-order approximations, that may improve the first-order ones, are reached by means of *parabolic tangent sets* in the sense of [37, Definition 3.28].

Definition 2.1.4 (Outer and inner parabolic tangent sets)**.** The *outer parabolic tangent set* (or *outer second-order tangent set*) to a closed set $C \subseteq \mathbb{Y}$ at a point $x \in C$ in the direction $d \in \mathbb{Y}$ is defined as

$$T_C^2(x,d) \doteq \left\{ w \in \mathbb{Y} \colon \exists \{\alpha^k\}_{k\in\mathbb{N}} \to 0^+,\ \text{dist}\left(x + \alpha_k d + \frac{1}{2}\alpha_k^2 w,\ C\right) = o(\alpha_k^2) \right\} \tag{2.14}$$

and the *inner parabolic tangent set* to C at x along d is the set

$$T_C^{\text{in},2}(x,d) = \left\{ w \in \mathbb{Y} \colon \text{dist}\left(x + \alpha d + \frac{1}{2}\alpha_k w,\ C\right) = o(t^2),\ t > 0 \right\}. \tag{2.15}$$

These sets are nonempty only if $d \in T_C(x)$ and $x \in C$, but they are always closed. Moreover, if C is convex, then $T_C^{\text{in},2}(x,d)$ is also convex, whereas $T_C^2(x,d)$ may be nonconvex [37, Example 3.35]. In particular, this means that they may not coincide even in the presence of convexity. A sufficient condition for these two notions of "parabolicity" to coincide is the so-called *second-order regularity* condition, which is presented in its most general form and discussed in detail around [37, Definition 3.85]. However, since we will only deal with second-order regularity for closed convex sets in finite-dimensional spaces, we employ the following specialized definition of [84, Definition 2.1]:

Definition 2.1.5 (Second-order regularity)**.** Let $C \subseteq \mathbb{Y}$ be a nonempty closed convex set, and let $x \in C$. We say that C is *second-order regular at* x if for every $d \in T_C(x)$ and all sequences $\{\alpha_k\}_{k\in\mathbb{N}} \to 0^+$ and $\{w^k\}_{k\in\mathbb{N}} \subseteq \mathbb{Y}$ such that $\alpha_k w^k \to 0$ and

$$x + \alpha_k d + \frac{1}{2}\alpha_k^2 w^k \in C,\ \forall k \in \mathbb{N}$$

it holds that

$$\lim_{k\to\infty} \text{dist}(w^k,\ T_C^{\text{in},2}(x,d)) = 0.$$

Roughly, second-order regularity at $x \in C$ means that the inner second-order tangent set, $T_C^{\text{in},2}(x,d)$, provides an upper bound for the outer second-order tangent set, $T_C^2(x,d)$, for every $d \in T_C(x)$. Noticing also that $T_C^{\text{in},2}(x,d) \subseteq T_C^2(x,d)$ for every such x and d is enough to see that these sets are equal under second-order regularity. Indeed, following Shapiro [84]:

Proposition 2.1.17. *Let $C \subseteq \mathbb{Y}$ be a closed convex set and let $x \in C$. Also, assume that C is second-order regular at x. Then, for every $d \in T_C(x)$, the equality $T_C^2(x,d) = T_C^{\text{in},2}(x,d)$ holds true and both sets are nonempty.*

Proof. Let $d \in T_C(x)$ and $w \in T_C^2(x,d)$ be arbitrary. Then, there exist sequences $\{\alpha_k\}_{k\in\mathbb{N}} \to 0^+$ and $\{w^k\}_{k\in\mathbb{N}} \to w$ such that $x + \alpha_k d + (\alpha_k^2/2)w^k \in C$ for every $k \in \mathbb{N}$, and if C is second-order regular at x, then $w \in T_C^{\text{in},2}(x,d)$ since the latter is closed. This implies that $T_C^2(x,d) \subseteq T_C^{\text{in},2}(x,d)$, but since the reverse inclusion is always true, equality holds. Furthermore, for each $d \in T_C(x)$ there exist some sequences $\{\alpha_k\}_{k\in\mathbb{N}} \to 0^+$ and $\{d^k\}_{k\in\mathbb{N}} \to d$ such that $x + \alpha_k d^k \in C$ for every $k \in \mathbb{N}$. Define

$$w^k \doteq \frac{2}{\alpha_k}(d^k - d),\ \forall k \in \mathbb{N}$$

and it follows that $\text{dist}(w^k, T_C^{\text{in},2}(x,d)) \to 0$ due to second-order regularity, since $\alpha_k w^k \to 0$. Thus, $T_C^{\text{in},2}(x,d)$ is nonempty and consequently $T_C^2(x,d)$ is also nonempty. □

A useful property of the second-order tangent sets is that, in some situations, they can be seen as tangent sets of tangent sets. The first proof of this result is due to Cominetti [42, Proposition

3.1] but we will present a more complete version of it that can be found in Bonnans and Shapiro's book [37].

Proposition 2.1.18 (Proposition 3.34 of [37]). *Let $C \subseteq \mathbb{Y}$ be a closed convex set, $x \in C$, and $d \in T_C(x)$. Then, the following relations hold:*

$$T_C^{\text{in},2}(x,d) + T_{T_C(x)}(d) \subseteq T_C^{\text{in},2}(x,d) \subseteq T_{T_C(x)}(d)$$

and

$$T_C^2(x,d) + T_{T_C(x)}(d) \subseteq T_C^2(x,d) \subseteq T_{T_C(x)}(d)$$

In particular, if $0 \in T_C^{\text{in},2}(x,d)$, then

$$T_C^{\text{in},2}(x,d) = T_C^2(x,d) = T_{T_C(x)}(d).$$

There are plenty of applications of the notion of second-order regularity, for instance, on the differentiability of the mapping Π_C. Thanks to Zarantonello [90], we know that the metric projection is directionally differentiable at every point of C, whereas Kruskal showed by a counterexample that this conclusion is not necessarily true at points outside C, in general. Nevertheless, Bonnans et al. [35] proved that the metric projection is directionally differentiable at all points $x \in \mathbb{Y}$ such that C is second-order regular at $\Pi_C(x)$. Let us recall their result:

Theorem 2.1.19 (Theorem 7.2 of [35]). *Let $C \subseteq \mathbb{Y}$ be a closed convex set and let $x \in \mathbb{Y}$ be such that C is second-order regular at $\Pi_C(x)$. Then, for every point $y \in \mathbb{Y}$ and every direction $d \in \mathbb{Y}$, the directional derivative $\Pi'_C(y,d)$ exists and it is given by*

$$\Pi'_C(y,d) = \operatorname*{argmin}_{h \in \mathcal{T}(y)} \left\{ \|d-h\|^2 - \sigma\left(\Pi_{C^\circ}(y),\ T_C^2(\Pi_C(y),h)\right) \right\},$$

where $\mathcal{T}(y) \doteq T_C(\Pi_C(y)) \cap \{\Pi_{C^\circ}(y)\}^\perp$, for every $y \in \mathbb{Y}$, and σ is the support function of $T_C^2(\Pi_C(y),h)$ at $\Pi_{C^\circ}(y)$ – see the discussion below for its definition.

The *support function* of a set $C \subseteq \mathbb{Y}$ at a point $x \in \mathbb{Y}$ is defined as

$$\sigma(x,C) \doteq \sup_{y \in C} \langle x,y \rangle \doteq \sup\{\langle x,y \rangle : y \in C\}$$

and it is straightforward to check that $\sigma(\,\cdot\,,C) = \sigma(\,\cdot\,,\mathrm{cl}(C)) = \sigma(\,\cdot\,,\mathrm{conv}(C))$. It also follows from the definition that for every $x_1,x_2 \in \mathbb{Y}$ and every $\alpha \geqslant 0$ we have

$$\sigma(x_1+x_2,C) \leqslant \sigma(x_1,C) + \sigma(x_2,C) \quad \text{and} \quad \sigma(\alpha x_1,C) = \alpha\sigma(x_1,C).$$

Moreover, we see from [37, Proposition 2.116] that for any pair of convex sets $C_1,C_2 \subseteq \mathbb{Y}$, it holds that

$$\sigma(\,\cdot\,,C_1+C_2) = \sigma(\,\cdot\,,C_1) + \sigma(\,\cdot\,,C_2)$$

and $C_1 \subseteq C_2$ if, and only if, $\sigma(\,\cdot\,,C_1) \leqslant \sigma(\,\cdot\,,C_2)$. Consequently, two sets coincide if, and only if, their support functions coincide.

Now consider the *indicator function* of C at $x \in \mathbb{Y}$, given by

$$\delta(x,C) \doteq \begin{cases} 0, & \text{if } x \in C \\ +\infty, & \text{otherwise,} \end{cases} \tag{2.16}$$

and note that we may rewrite the support function as

$$\begin{aligned}\sigma(x, C) &= \sup_{y \in \mathbb{Y}} \{\langle x, y\rangle - \delta(y, C)\} \\ &= -\inf_{y \in \mathbb{Y}} \{\delta(y, C) - \langle x, y\rangle\}.\end{aligned} \tag{2.17}$$

Rockafellar [81] and many other authors call this property *conjugacy*; that is, the indicator function is the *conjugate* of the support function in this sense. In particular, if C is a closed convex cone, then $\sigma(x, C) = \delta(x, C^\circ)$ for every $x \in \mathbb{Y}$. In the next section, we will use support functions in the study of the problem of optimizing a linear function over the preimage of a cone by an affine mapping, which is in particular a generalization of evaluating the support function itself.

2.2 Linear optimization theory

The general form of a (finite-dimensional) *linear optimization* problem – or *Linear Programming* (LP) problem – is the following:

$$\begin{aligned}\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad & \langle c, x\rangle, \\ \text{subject to} \quad & Ax + b \in \mathcal{K},\end{aligned} \tag{LP}$$

where $\mathcal{K} \subseteq \mathbb{Y}$ is a nonempty closed convex *set* that shall remain fixed from this point onwards, $\mathbb{Y}$ is a finite-dimensional space with an inner product, $A\colon \mathbb{R}^n \to \mathbb{Y}$ is a linear mapping, and $b \in \mathbb{Y}$. We will use the same notation $\langle \cdot, \cdot\rangle$ for the standard inner product of $\mathbb{R}^n$ and the inner product of $\mathbb{Y}$, unless the space is not clear from the context, in which case we will specify it as in $\langle \cdot, \cdot\rangle_{\mathbb{Y}}$. The *feasible set* of (LP) will be denoted by

$$\Omega \doteq \{x \in \mathbb{R}^n \colon Ax + b \in \mathcal{K}\}.$$

Needless to say, when $\mathcal{K} = \mathbb{R}^m_+$ Problem (LP) becomes a classical Linear Programming problem. However, saying that the general (LP) is also a *linear* problem is somewhat an abuse of terminology, since $\mathcal{K}$ is not necessarily assumed to be a "linear set" (i.e., a polyhedral set or a polytope). The effects of such "nonlinearity" are very apparent in one the most fundamental aspects of optimization theory, which is *duality*.

2.2.1 Parametric duality

There are several different ways of deriving duality results for conic optimization, motivated by different interpretations of duality; all of them are meaningful. In fact, duality is one of those concepts that are basic enough so everyone knows a bit of it, but also extensive enough to surprise us with something new every once in a while.

The most intuitive notion of duality is the one presented by Rockafellar [81] and also by Bonnans and Shapiro [37, Section 2.5.3] which is called *parametric duality*. The idea is to perturb (LP) by introducing a parameter $y \in \mathbb{Y}$ and then consider the parametric problem

$$\begin{aligned}\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad & \langle c, x\rangle, \\ \text{subject to} \quad & Ax + b + y \in \mathcal{K}.\end{aligned} \tag{Par-LP}$$

Although this process may seem rather artificial at first sight, it gives us a bit more of structure to study the solution of (LP). For instance, our problem is now reduced to simply evaluating the function

$$\mathcal{P}(y) \doteq \inf_{x\in\mathbb{R}^n}\{\langle c,x\rangle + \delta(Ax+b+y,\mathcal{K})\}$$

at $y=0$. Indeed, notice that for every $y\in\mathbb{Y}$, we have that $\delta(Ax+b+y,\mathcal{K})=0$ if x is feasible for (Par-LP) and it is infinity-valued otherwise – see (2.16). Then, our next step is to find the best linear function that bounds $\mathcal{P}$ inferiorly around a given y, which is roughly the same as finding the best supporting hyperplane to the *epigraph* of $\mathcal{P}$ at the point $(y,\mathcal{P}(y))$. In other words, we want to find the best $\mu\in\mathbb{Y}$ and $\gamma\in\mathbb{R}$ such that

$$\mathcal{P}(y) \geqslant \langle y,\mu\rangle + \gamma,\ \forall y\in\mathbb{Y} \tag{2.18}$$

so for each slope μ clearly the best $\gamma\doteq\gamma(\mu)$ satisfying (2.18) is given by

$$\begin{aligned}
\gamma(\mu) &\doteq \inf_{y\in\mathbb{Y}}\{\mathcal{P}(y)-\langle y,\mu\rangle\}\\
&= \inf_{y\in\mathbb{Y}}\left\{\inf_{x\in\mathbb{R}^n}\{\langle c,x\rangle+\delta(Ax+b+y,\mathcal{K})\}-\langle y,\mu\rangle\right\}\\
&= \inf_{x\in\mathbb{R}^n,y\in\mathbb{Y}}\{\langle c,x\rangle+\delta(Ax+b+y,\mathcal{K})-\langle Ax+b+y,\mu\rangle+\langle Ax+b,\mu\rangle\}\\
&= \inf_{x\in\mathbb{R}^n}\left\{\langle c,x\rangle+\langle Ax+b,\mu\rangle+\inf_{y\in\mathbb{Y}}\{\delta(Ax+b+y,\mathcal{K})-\langle Ax+b+y,\mu\rangle\}\right\}\\
&= \inf_{x\in\mathbb{R}^n}\{\langle c,x\rangle+\langle Ax+b,\mu\rangle-\sigma(\mu,\mathcal{K})\},
\end{aligned}\tag{2.19}$$

where in the third equation we added the null term $\langle Ax+b,\mu\rangle-\langle Ax+b,\mu\rangle$ and in the last equation we used (2.17) with a change of variables $y\doteq Ax+b+y$ that maps $\mathbb{Y}$ to itself. Then, the best lower linear approximation of $\mathcal{P}$ at y is given by $\overline{\mu}\in\mathbb{Y}$ such that

$$\langle y,\overline{\mu}\rangle+\gamma(\overline{\mu}) = \sup_{\mu\in\mathbb{Y}}\{\langle y,\mu\rangle+\gamma(\mu)\},$$

assuming that it exists. In particular, at $y=0$, the problem of our interest is to find

$$\begin{aligned}
\sup_{\mu\in\mathbb{Y}}\gamma(\mu) &= \sup_{\mu\in\mathbb{Y}}\inf_{x\in\mathbb{R}^n}\{\langle c,x\rangle+\langle Ax+b,\mu\rangle-\sigma(\mu,\mathcal{K})\}\\
&= \sup_{\mu\in\mathbb{Y}}\{\langle b,\mu\rangle-\sigma(\mu,\mathcal{K})+\inf_{x\in\mathbb{R}^n}\langle x,A^*[\mu]+c\rangle\}\\
&= \sup_{\mu\in\mathbb{Y}}\{\langle b,\mu\rangle-\sigma(\mu,\mathcal{K})-\sigma(-A^*[\mu]-c,\mathbb{R}^n)\}
\end{aligned}\tag{2.20}$$

where we recall that A^* denotes the *adjoint* operator of A. However, note that $\sigma(-A^*[\mu]-c,\mathbb{R}^n)=\delta(-A^*[\mu]-c,\{0\})$ since $(\mathbb{R}^n)^\circ=\{0\}$, meaning that evaluating (2.20) can be rephrased as an optimization problem in the form:

$$\begin{aligned}
&\underset{\mu\in\mathbb{Y}}{\text{Maximize}} && \langle b,\mu\rangle-\sigma(\mu,\mathcal{K}),\\
&\text{subject to} && A^*[\mu]+c=0.
\end{aligned}\tag{LD}$$

Problem (LD) is known as the (parametric) *dual problem* of (LP). If $\mathcal{K}$ is a closed convex cone, then $\sigma(\mu,\mathcal{K})=\delta(\mu,\mathcal{K}^\circ)$ and in this case (LD) can be equivalently reformulated by dropping the "sigma-term" $-\sigma(\mu,\mathcal{K})$ from the objective function and adding the constraint $\mu\in\mathcal{K}^\circ$. Moreover, the inequality

$$\mathcal{P}(0) \geqslant \sup_{\mu \in \mathbb{Y}} \gamma(\mu),$$

which holds by the construction of γ, is called *weak duality* and the difference between the optimal values of (LP) and (LD) is called the *duality gap* between these problems. If either (LP) or (LD) is infeasible, then we can see from the definition of $\mathcal{P}$ and (2.20), respectively, that the duality gap must be infinite. On the other hand, if the following conditions hold:

- The duality gap is zero; that is, $\mathcal{P}(0) = \sup\{\gamma(\mu) \colon \mu \in \mathbb{Y}\}$;
- Both optimal values, of (LP) and (LD), are attained;

then we say that *strong duality* holds between (LP) and (LD). Relying on strong duality, one can solve (LP) by means of solving (LD) which may be simpler in some situations.

It is crucial to remark that, contrary to classical Linear Programming, the duality gap for the general (LP) may be zero but strong duality may not hold anyway due to the lack of attainability. Let us illustrate this with an example by Pataki [77][5].

Example 2.2.1 (Example 1 of [77]). *Consider the following dual pair of problems:*

$$\begin{aligned} &\underset{x \in \mathbb{R}^2}{\text{Minimize}} && x_1, \\ &\text{subject to} && \begin{bmatrix} x_1 & 1 \\ 1 & x_2 \end{bmatrix} \in \mathbb{S}^2_+, \end{aligned} \qquad \text{(P)}$$

$$\begin{aligned} &\underset{\mu \in -\mathbb{S}^2_+}{\text{Maximize}} && \left\langle \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{12} & \mu_{22} \end{bmatrix} \right\rangle, \\ &\text{subject to} && \mu_{11} = -1, \ \mu_{22} = 0, \end{aligned} \qquad \text{(D)}$$

where $\mathbb{S}^2_+$ is the set of 2×2 symmetric positive semidefinite matrices. Notice that the feasible set of (P) is $\Omega = \{x \in \mathbb{R}^2_+ \colon x_1 x_2 \geqslant 1\}$, whereas the feasible set of (D) contains only the point $\overline{\mu} \doteq \operatorname{Diag}(-1, 0)$.

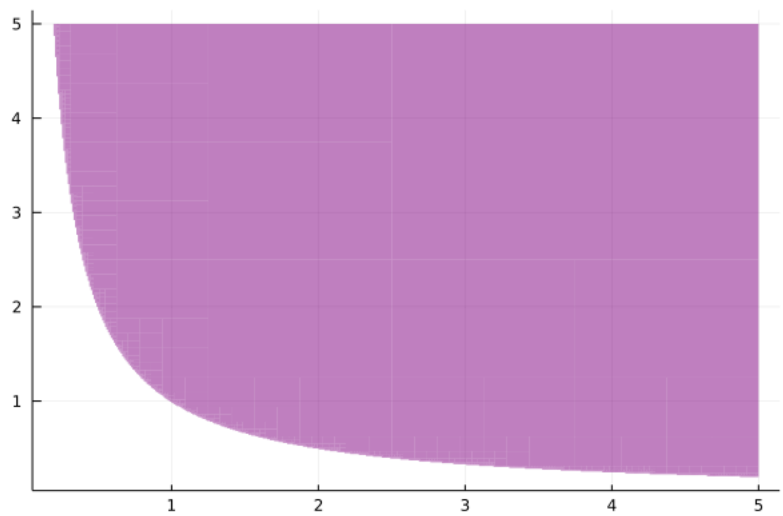


Figure 2.2: *Feasible set of (P).*

The dual problem (D) is bounded above by zero and attains this optimal value at $\overline{\mu}$. The primal problem (P) is also bounded below by zero, which is also its infimum value since $x_\varepsilon \doteq (\varepsilon, 1/\varepsilon) \in \Omega$ for all $\varepsilon > 0$. However, note that it does not attain it. Thus, the duality gap in this case is zero but strong duality does not hold nonetheless.

[5] We are grateful to Gábor Pataki for pointing us to this fantastic paper [77].

Any sufficient condition over A, b, and $\mathcal{K}$, for strong duality to hold provided that (LP) is feasible and attains its solution, say at $\overline{x} \in \Omega$, is called a *Constraint Qualification* (CQ) for (LP). For instance:

- $\mathcal{K}$ is polyhedral;
- **Slater's CQ**: There exists some $x \in \mathbb{R}^n$ such that $Ax + b \in \text{ri}(\mathcal{K})$;
- **Mangasarian-Fromovitz' CQ**: There is some $d \in \mathbb{R}^n$ such that $Ad \in \text{ri}(T_{\mathcal{K}}(A\overline{x} + b))$;
- **Guignard's CQ**: There exists a solution $\overline{x}$ of (LP) such that the set

$$H(\overline{x}) \doteq A^*[N_{\mathcal{K}}(A\overline{x} + b)] \tag{2.21}$$

  is closed.

are examples of Constraint Qualifications. Let us prove this.

Proposition 2.2.1. *If Slater's CQ holds, then every feasible point of* (LP) *satisfies Mangasarian-Fromovitz' CQ. Moreover, Mangasarian-Fromovitz' CQ implies Guignard's CQ at any given feasible point of* (LP).

Proof. Suppose that Slater's CQ holds, so there exists some $x \in \mathbb{R}^n$ such that $Ax + b \in \text{ri}(\mathcal{K})$, and let $y \in \Omega$ be any other feasible point. Then,

$$Ay + b + A(x - y) \in \text{ri}(\mathcal{K}).$$

so define $d \doteq x - y$ and we have $Ad \in \text{ri}(\mathcal{K} - Ay - b)$, but notice that

$$\text{ri}(\mathcal{K} - Ay - b) \subseteq \text{ri}(\text{cone}(\mathcal{K} - Ay - b)) = \text{ri}(T_{\mathcal{K}}(Ay + b)),$$

because the first inclusion comes from Lemma 2.1.2 after realizing that $\text{ri}(\mathcal{K} - Ax - b) \cap \text{ri}(\text{cone}(\mathcal{K} - Ax - b)) \neq \emptyset$ since otherwise they would be properly separable; and the equality comes from Proposition 2.1.15[6]. Then, $Ad \in \text{ri}(T_{\mathcal{K}}(Ay + b))$ and Mangasarian-Fromovitz' CQ holds at the arbitrarily chosen y. Furthermore, for any $y \in \Omega$ the existence of some d such that $Ad \in \text{ri}(T_{\mathcal{K}}(Ay + b))$ implies through Proposition 2.1.7 that $A^*[T_{\mathcal{K}}(Ay + b)^\circ] = H(y)$ is closed. □

It also follows from Proposition 2.1.7 that $H(y)$ is closed when $\mathcal{K}$ is a polyhedral cone, since in this case $N_{\mathcal{K}}(Ay + b)$ is also polyhedral for every $y \in \Omega$. Moreover:

Theorem 2.2.2. *If* (LP) *admits a solution $\overline{x} \in \Omega$ that satisfies Guignard's CQ, then strong duality holds between* (LP) *and* (LD).

Proof. Although the general idea of this proof is somewhat classical, we will exhibit a mild adaptation of the original result by Guignard [53] that better suits our context, for completeness. Let $\overline{x}$ be a solution of (LP) satisfying Guignard's CQ, so we have $\langle c, \overline{x} + \alpha d\rangle \geqslant \langle c, \overline{x}\rangle$ for every $d \in \mathbb{R}^n$ and every $\alpha \geqslant 0$ such that $A(\overline{x} + \alpha d) + b \in \mathcal{K}$. By definition, this implies the following:

$$\langle c, d\rangle \geqslant 0 \text{ for every } d \in A^{-1}(T_{\mathcal{K}}(A\overline{x} + b)). \tag{2.22}$$

[6]Recall that $\text{ri}(C) = \text{ri}(\text{cl}(C))$ for every closed convex set $\mathcal{C} \subseteq \mathbb{Y}$.

We will now prove that $A^{-1}(T_{\mathcal{K}}(A\overline{x}+b)) = H(\overline{x})^{\circ}$. For any $d \in \mathbb{R}^n$ such that $Ad \in T_{\mathcal{K}}(A\overline{x}+b)$ and any $y \in N_{\mathcal{K}}(A\overline{x}+b)$ it holds that $\langle d, A^*y\rangle = \langle Ad, y\rangle \leqslant 0$ and hence

$$A^{-1}(T_{\mathcal{K}}(A\overline{x}+b)) \subseteq H(\overline{x})^{\circ}.$$

Conversely, if there exists some $d \in H(\overline{x})^{\circ}$ such that $Ad \notin T_{\mathcal{K}}(A\overline{x}+b)$, then Theorem 2.1.4 tells us that there exists a vector w such that $\langle w, Ad\rangle > 0$ and $\langle w, z\rangle < 0$, for all $z \in T_{\mathcal{K}}(A\overline{x}+b)$, which in turn implies that $w \in N_{\mathcal{K}}(A\overline{x}+b)$. Then, it follows that $A^*[w] \in H(\overline{x})$, but since $\langle A^*[w], d\rangle = \langle w, Ad\rangle > 0$, this contradicts $d \in H(\overline{x})^{\circ}$. This proves that

$$H(\overline{x})^{\circ} \subseteq A^{-1}(T_{\mathcal{K}}(A\overline{x}+b))$$

also. With this in mind, observe that (2.22) holds if, and only if,

$$-c \in (H(\overline{x})^{\circ})^{\circ} = \mathrm{cl}(H(\overline{x})) = H(\overline{x})$$

where the last equality is due to Guignard's CQ. Then, there exists some $\overline{\mu} \in N_{\mathcal{K}}(A\overline{x}+b)$ such that $-c = A^*[\overline{\mu}]$, but recalling Proposition 2.1.16 we see that $\langle z - A\overline{x} - b, \overline{\mu}\rangle \leqslant 0$ for every $z \in \mathcal{K}$, which implies that

$$\sigma(\overline{\mu}, \mathcal{K}) \leqslant \langle \overline{x}, A^*[\overline{\mu}]\rangle + \langle b, \overline{\mu}\rangle = \langle \overline{x}, -c\rangle + \langle b, \overline{\mu}\rangle$$

and hence $\langle \overline{x}, c\rangle \leqslant \langle b, \overline{\mu}\rangle - \sigma(\overline{\mu}, \mathcal{K})$. Recall weak duality, which implies that $\langle \overline{x}, c\rangle \geqslant \langle b, \mu\rangle - \sigma(\mu, \mathcal{K})$ for every μ such that $A^*[\mu] + c = 0$ (in particular, $\overline{\mu}$), and we conclude that strong duality holds. □

Even though Slater's CQ and Mangasarian-Fromovitz' CQ are very well known, Guignard's CQ is rarely mentioned in the literature of finite-dimensional conic optimization. In fact this is quite surprising since Guignard's CQ is popular in the infinite-dimensional world, where it was originally presented [53], and also in *Nonlinear Programming* (NLP). Besides, Guignard's CQ has a very interesting property: it is the weakest possible Constraint Qualification for Conic Programming (see [51]), in the sense that it must be implied by any other Constraint Qualification. Observe that this is consistent with Proposition 2.2.1.

To finish this discussion, we exhibit an example extracted from Andersen et al. [2, Subsection 2.1] that illustrates a conic problem whose primal solution is attained, but the dual problem is infeasible and hence the duality gap is infinite; also, we show that no Constraint Qualification holds at its solution.

Example 2.2.2. *Consider the following problem:*

$$\begin{aligned} \underset{x\in\mathbb{R}^2}{\text{Minimize}} \quad & -x_2, \\ \text{subject to} \quad & (x_1, x_1, x_2) \in \mathbb{L}^3, \end{aligned}$$

where $\mathbb{L}^3 = \{(x_0, \widehat{x}) \in \mathbb{R} \times \mathbb{R}^2 \colon x_0 \geqslant \|\widehat{x}\|_2\}$ *is the three-dimensional Lorentz cone. In this case, we have* $\Omega = \{x \in \mathbb{R}^2 \colon x_1 \geq 0,\ x_2 = 0\}$ *and the point* $\overline{x} = (0,0) \in \mathbb{R}^2$ *is among its solutions. Additionally, note that the set* $H(\overline{x})$ *defined in* (2.21) *is given by*

$$H(\overline{x}) = \{(y_1 + y_2, y_3) \in \mathbb{R}^2 \colon (y_1, y_2, y_3) \in (\mathbb{L}^3)^{\circ}\}$$

is not closed, since it is elementary to see that the sequence

$$\left\{\left(-\frac{1}{k}, -1\right)\right\}_{k\in\mathbb{N}} \to (0,-1)$$

is contained in $H(\overline{x})$, but $(0,-1)$ is not. Thus, Guignard's CQ is violated in this example, and since it is the weakest possible CQ, no other CQ holds. The dual problem in this case is given by

$$\begin{aligned} &\underset{\mu\in -\mathbb{L}^3}{\text{Maximize}} && 0,\\ &\text{subject to} && \begin{bmatrix}1 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\mu = \begin{bmatrix}0\\ -1\end{bmatrix}, \end{aligned}$$

which is infeasible, meaning that the duality gap is infinite.

Problems that do not satisfy any CQ are called *degenerate*. An outstanding advance to "treat" degenerate problems in order to "make them solvable" using duality tools was made by Wolkowicz and Borwein [38] by means of a preprocessing technique called *facial reduction*, which stood out for allowing strong duality results for (LP) to be obtained without CQs by focusing on a nondegenerate but equivalent reformulation of the problem. Next, we review some aspects of facial reduction for a subclass of (LP).

2.2.2 Notions of facial reduction

Before we start, we must stress that in this section we assume that $\mathcal{K}$ is a closed convex *cone* in Problem (LP). This particular case is often called in the literature a (linear) *Conic Programming* (CP) problem. For simplicity, because this subsection is adapted from Pataki [76], we will also assume that $\mathcal{K}$ is *nice* or *facially dual complete*, which means that $\mathcal{K}^\circ + F^\perp$ is closed for every $F \trianglelefteq \mathcal{K}$.

Roughly speaking, the idea behind the facial reduction approach of Wolkowicz and Borwein [38] is to find the smallest face F of $\mathcal{K}$ such that

$$\Omega = \{x \in \mathbb{R}^n \colon Ax + b \in F\},$$

but note that, for every face $F \trianglelefteq \mathcal{K}$ that contains the set $\Gamma \doteq A[\Omega] + b$, we have that

$$Ax + b \in \mathcal{K} \;\Leftrightarrow\; Ax + b \in \Gamma \;\Leftrightarrow\; Ax + b \in F,$$

which means that finding such a minimal face representation of Ω amounts to finding the minimal face of $\mathcal{K}$ that contains Γ, i.e., $\overline{F} \doteq F_{\min}(\Gamma)$. The main reason why this is interesting is that even if the original constraint $Ax + b \in \mathcal{K}$ is degenerate, the "facially reduced" constraint $Ax + b \in \overline{F}$ always satisfies Slater's CQ. Indeed, by item 1 of Proposition 2.1.10, which follows from the very definition of minimal face, we see that $\operatorname{ri}(\overline{F}) \cap \Gamma \neq \emptyset$, so there exists some $x \in \mathbb{R}^n$ such that $Ax + b \in \operatorname{ri}(\overline{F})$.

The work of Wolkowicz and Borwein was later revisited and improved by Pataki [76], and Waki and Muramatsu [87], both of whom presented very elementary derivations of simple algorithms for computing $\overline{F}$; i.e., *facial reduction algorithms*. For completeness, we will present the former. The essence of Pataki's algorithm lies in the following self-explanatory lemma, sometimes called the *facial reduction lemma*.

Lemma 2.2.3. *Suppose that $\Omega \neq \emptyset$ and let $\Gamma \doteq A[\Omega] + b$. Then, for every $F \trianglelefteq \mathcal{K}$ such that $\Gamma \subseteq F$, it holds that:*

1. *For every* $y \in F^\circ \cap \mathrm{Ker}(A^*) \cap \{b\}^\perp$ *we have* $F_{\min}(\Gamma) \subseteq F \cap \{y\}^\perp \subseteq F$*;*

2. *There exists some* $y \in F^\circ \cap \mathrm{Ker}(A^*) \cap \{b\}^\perp$ *such that* $F \cap \{y\}^\perp \neq F$ *if, and only if,* $F \neq F_{\min}(\Gamma)$*.*

Proof. For item 1, note that for every $x \in \Omega$, we have $\langle Ax + b, y\rangle = \langle x, A^*[y]\rangle + \langle b, y\rangle = 0$, so $\Gamma = A[\Omega] + b \subseteq F \cap \{y\}^\perp$ and since $y \in F^\circ$ the latter is a face of F (and by extension a face of $\mathcal{K}$), implying that $F_{\min}(\Gamma) \subseteq F \cap \{y\}^\perp$.

The proof of item 2 that we present is a small adaptation of the proof of Pataki [76, Lemma 1]. Take any $w \in \mathrm{ri}(F)$ and consider the following dual pair of problems:

$$\begin{aligned} &\underset{(t,x)\in\mathbb{R}\times\mathbb{R}^n}{\text{Minimize}} && t && \text{(Aux-P)} \\ &\text{subject to} && Ax + b + tw \in F \end{aligned} \qquad \begin{aligned} &\underset{y\in F^\circ}{\text{Maximize}} && \langle y, b\rangle && \text{(Aux-D)} \\ &\text{subject to} && \langle w, y\rangle = -1, \\ & && A^*[y] = 0. \end{aligned}$$

By Proposition 2.1.10 item 2, we know that if $F \neq F_{\min}(\Gamma)$, then $\Gamma \cap \mathrm{ri}(F) = \emptyset$, which in turn implies that every pair $(t, x) \in \mathbb{R} \times \mathbb{R}^n$ satisfying $Ax + b + tw \in F$ must also satisfy $t \geqslant 0$. Indeed, it follows from Theorem 2.1.1 that if $(t, x) \in \mathbb{R} \times \mathbb{R}^n$ satisfies $Ax + b + tw \in F$ and $t < 0$, then

$$\mu(Ax + b + tw) - (1 - \mu)tw \in \mathrm{ri}(F)$$

for every $\mu \in [0, 1)$, because $-tw \in \mathrm{ri}(F)$. In particular, for $\mu = 1/2$, we obtain that $(Ax+b)/2 \in \mathrm{ri}(F)$ and, therefore, $Ax + b \in \mathrm{ri}(F)$ as well, so $Ax + b \in \Gamma \cap \mathrm{ri}(F) \neq \emptyset$. Moreover, note that since $\Omega \neq \emptyset$, then (Aux-P) satisfies Slater's CQ regardless of its fulfilment for (LP). Thus, if $F \neq F_{\min}(\Gamma)$, then problem (Aux-P) is bounded below, attains its optimal value at $t = 0$, and due to Slater's CQ (Aux-D) also attains the same optimal value. The dual solution y then belongs to $F^\circ \cap \mathrm{Ker}(A^*) \cap \{b\}^\perp$, but $w \in F \setminus (F \cap \{y\}^\perp)$. □

This naturally leads to a simple (theoretical) algorithm for computing $F_{\min}(\Gamma)$ that consists of solving (Aux-P) and (Aux-D) successively, but replacing F with $F \cap \{y\}^\perp$ after each step. If (Aux-P) turns out to be infeasible, then one concludes that $\Omega = \emptyset$, implying $\Gamma = F_{\min}(\Gamma) = \emptyset$ as well. If (Aux-P) is feasible but unbounded, then $F = F_{\min}(\Gamma)$ and the algorithm terminates.

2.3 Nonlinear conic optimization theory

In this section, we will use some results on linear Conic Programming to study a more general conic optimization problem, that admits nonlinearity both in the cone and the functions of the problem. The general form of such a *Nonlinear Conic Programming* (NCP) problem is the following:

$$\begin{aligned} &\underset{x\in\mathbb{R}^n}{\text{Minimize}} && f(x), \\ &\text{subject to} && G(x) \in \mathcal{K}, \end{aligned} \tag{NCP}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $G\colon \mathbb{R}^n \to \mathbb{Y}$ are at least twice continuously differentiable functions. Just as in (LP) the set of all feasible points of this problem will be denoted by Ω.

We denote the first derivative (or *Jacobian*) of G at a given point $x \in \mathbb{R}^n$ by $DG(x)\colon \mathbb{R}^n \to \mathbb{Y}$. As a standard, we will use its representation in the canonical basis of $\mathbb{R}^n$, namely $e_1, \ldots, e_n$, which yields

$$DG(x)[d] = \sum_{i=1}^{n} \partial_i G(x) d_i$$

for any $d \in \mathbb{R}^n$, where $\partial_i G(x) \doteq G'(x, e_i)$ is the i-th partial derivative of G at x. Moreover, it is elementary to see that for every $x \in \mathbb{R}^n$ and every $y \in \mathbb{Y}$, the adjoint of $DG(x)$ satisfies the relation

$$DG(x)^*[y] = [\langle \partial_i G(x), y \rangle]_{i \in \{1,\ldots,n\}}.$$

Similarly, it is possible to define a "second-order ajoint operator" for the second derivative of G at x, that is, $D^2G(x) \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{Y}$, as follows:

$$D^2G(x)^*[y] \doteq [\langle \partial_i \partial_j G(x), y \rangle]_{\substack{i \in \{1,\ldots,n\} \\ j \in \{1,\ldots,n\}}},$$

for every $y \in \mathbb{Y}$, which is derived from the adjoints of $D^2G(x)[d, \cdot]$ and $D^2G(x)[\cdot, d]$ for each fixed $d \in \mathbb{R}^n$ hence, by construction it satisfies the relation

$$\langle D^2G(x)[d, d], y \rangle = d^\top (D^2G(x)^*[y])d$$

for every $y \in \mathbb{Y}$ and every $d \in \mathbb{R}^n$.

2.3.1 First-order optimality conditions

The ideal and perhaps most mathematical way of solving an NCP problem is by characterizing its solutions algebraically, in terms of the problem data. For instance, the local minimizers of a continuously differentiable convex function $\phi \colon \mathbb{R}^n \to \mathbb{R}$ are simply the points $\overline{x} \in \mathbb{R}^n$ such that $\nabla \phi(\overline{x}) = 0$. However, such an ideal approach is not always applicable to the general nonconvex case and weaker descriptions of optimality must come into play. Any statement that must be satisfied by all local solutions of (NCP) is called a *necessary optimality condition*, whereas conditions that imply optimality are called *sufficient optimality conditions*. Not all necessary conditions are sufficient and vice-versa. Besides, they usually play different roles. For example, in the design of iterative algorithms for solving (NCP), necessary conditions are useful for defining stopping criteria since their violation at a given iteration ensures that the current point is not optimal, in which case the next iteration is well-defined and the algorithm must keep running; sufficient conditions, on the other hand, are useful for proving convergence rates and stability results for they are usually sufficient for "strict optimality" also (that is, they usually characterize isolated optima).

In this section, we will review *first-order* optimality conditions, which are mostly of the necessary type since in general sufficient conditions demand information on the curvature of the problem. Using "first-order" information roughly means approximating the problem linearly around a given point of interest and trying to take conclusions for the original problem based on such a linear approximation. For instance, let $\overline{x} \in \Omega$ be a local minimizer of (NCP), consider the tangent approximation of Ω at $\overline{x}$ given by $T_\Omega(\overline{x})$, and let us define the following "tangent problem" around $\overline{x}$:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{Minimize}} & \langle \nabla f(\overline{x}), d \rangle \\ \text{subject to} & d \in T_\Omega(\overline{x}). \end{array} \tag{Tan-P}$$

Recall Definition 2.1.3 and notice that for each $d \in T_\Omega(\overline{x})$ there are sequences $\{\alpha_k\}_{k \in \mathbb{N}} \to 0^+$ and $\{d^k\}_{k \in \mathbb{N}} \to d$ such that $\overline{x} + \alpha_k d^k \in \Omega$ for all $k \in \mathbb{N}$, and by the local optimality of $\overline{x}$ over Ω we must have

$$f(\overline{x} + \alpha_k d^k) = f(\overline{x}) + \alpha_k \langle \nabla f(\overline{x}), d^k \rangle + o(\alpha^k) \geqslant f(\overline{x})$$

for every k large enough, which implies that $\langle \nabla f(\overline{x}), d^k \rangle \geqslant 0$ for all such k and, consequently,

$\langle \nabla f(\overline{x}), d\rangle \geqslant 0$. Because $0 \in T_\Omega(\overline{x})$, the point $d = 0$ must be a (global) solution of (Tan-P). This is one of the simplest necessary optimality conditions, sometimes called the *geometric optimality condition*.

- **Geometric optimality condition:** If $\overline{x} \in \Omega$ is a local solution of (NCP), then the global solution of (Tan-P) is $d = 0$; equivalently, $-\nabla f(\overline{x}) \in T_\Omega(\overline{x})^\circ$.

Conversely, if $d = 0$ is the unique solution of (Tan-P), then

$$\langle \nabla f(\overline{x}), d\rangle > 0,\ \forall d \in T_\Omega(\overline{x}) \setminus \{0\}; \tag{2.23}$$

and going a bit further we claim that (2.23) implies that $\overline{x}$ must be a local minimizer of (NCP). Indeed, if $\overline{x}$ was not a local minimizer of (NCP), then there would exist a sequence of feasible points $\{x^k\}_{k\in\mathbb{N}} \subseteq \Omega$, converging to $\overline{x}$, such that $f(x^k) < f(\overline{x})$ for every $k \in \mathbb{N}$ (in particular, note that $x^k \neq \overline{x}$); then, every limit point d of the sequence $\{d^k\}_{k\in\mathbb{N}}$ given by

$$d^k \doteq \frac{x^k - \overline{x}}{\|x^k - \overline{x}\|}$$

belongs to $T_\Omega(\overline{x}) \setminus \{0\}$, but taking limits in the relation

$$f(x^k) = f(\overline{x} + \|x^k - \overline{x}\| d^k) = f(\overline{x}) + \|x^k - \overline{x}\|\langle \nabla f(\overline{x}), d^k\rangle + o(\|x^k - \overline{x}\|) < f(\overline{x})$$

leads to $\langle \nabla f(\overline{x}), d\rangle \leqslant 0$. As a matter of fact (2.23) can be seen as the "sufficient counterpart" of the geometric condition.

The major drawback of the geometric condition is that the set $T_\Omega(\overline{x})$ may be somewhat intricate, at least to the point of not admitting a straightforward closed form. Our response to such adversity is, of course, "linearizing" $T_\Omega(\overline{x})$ even further by exploiting the structure of Ω as the preimage of $\mathcal{K}$ by G.

Definition 2.3.1. Let $\overline{x} \in \Omega$. The *linearized tangent cone* to Ω at $\overline{x}$ is the set

$$L_\Omega(\overline{x}) \doteq DG(x)^{-1}(T_\mathcal{K}(G(\overline{x}))) = \{d \in \mathbb{R}^n \colon DG(x)[d] \in T_\mathcal{K}(G(\overline{x}))\}.$$

Observe that the linearized tangent cone of Definition 2.3.1, which also appears in the classical textbook of Bonnans and Shapiro [37] and in the seminal papers of Kawasaki [62] and Cominetti [42], for instance, is essentially a linear approximation of $G^{-1}(\mathcal{K})$ at $\overline{x}$ obtained by simply approximating each part of it at a time. Note also that $T_\Omega(\overline{x}) \subseteq L_\Omega(\overline{x})$ for every $\overline{x} \in \Omega$, but the opposite inclusion is not always true: take, for instance, $G(x) \doteq x_1x_2$ and $\mathcal{K} \doteq \{0\}$ at $\overline{x} = 0$; then $T_\Omega(\overline{x}) = \Omega$ but $L_\Omega(\overline{x}) = \mathbb{R}^2$. Furthermore, Definition 2.3.1 defines a relaxed linear approximation of the tangent problem (Tan-P) as follows:

$$\begin{aligned} &\underset{d\in\mathbb{R}^n}{\text{Minimize}} && \langle \nabla f(\overline{x}), d\rangle \\ &\text{subject to} && DG(\overline{x})[d] \in T_\mathcal{K}(G(\overline{x})). \end{aligned} \tag{Lin-P}$$

Contrary to (Tan-P), the optimality of $\overline{x}$ does not imply that $d = 0$ is the solution of (Lin-P), unless of course $T_\Omega(\overline{x})^\circ = L_\Omega(\overline{x})^\circ$, in which case the conclusion should follow from the geometric optimality condition. Nevertheless, the utility of (Lin-P) lies in the fact it is a linear Conic Programming problem, which allows us to derive a "dual form" of the geometric condition by means of the (parametric) dual problem of (Lin-P):

$$\begin{aligned} &\underset{\mu \in N_{\mathcal{K}}(G(\bar{x}))}{\text{Maximize}} && 0 \\ &\text{subject to} && \nabla f(\bar{x}) + DG(\bar{x})^*[\mu] = 0. \end{aligned} \tag{Lin-D}$$

Thus, if $T_\Omega(\bar{x})^\circ = L_\Omega(\bar{x})^\circ$ and some Constraint Qualification holds for (Lin-P), then the optimality of $\bar{x}$ ensures that (Lin-P) is feasible and attains its solution $d = 0$, which in turn implies that (Lin-D) is also feasible. The feasible points of (Lin-D) are called *Lagrange multipliers* associated with $\bar{x}$ and the feasible set of (Lin-D) will be denoted by $\Lambda(\bar{x})$ throughout the text. Noteworthy, the constraints of (Lin-D) are widely known as the *Karush-Kuhn-Tucker* (KKT) conditions.

Definition 2.3.2 (KKT conditions). Let $\bar{x} \in \Omega$. We say that $\bar{x}$ satisfies the *Karush-Kuhn-Tucker conditions* if there exists some $\bar{\mu} \in N_{\mathcal{K}}(G(\bar{x})) = \{\mu \in \mathcal{K}^\circ \colon \langle G(\bar{x}), \mu \rangle = 0\}$ such that $\nabla f(\bar{x}) + DG(\bar{x})^*[\bar{\mu}] = 0$.

Points that satisfy the KKT conditions are usually called *stationary points* or, sometimes, *KKT points* of (NCP). It is clear from the previous paragraphs that stationarity can be roughly interpreted as some sort of "local strong duality," hence we will abuse terminology to call *any condition that ensures* $T_\Omega(\bar{x})^\circ = L_\Omega(\bar{x})^\circ$ *and strong duality between* (Lin-P) *and* (Lin-D) as a *Constraint Qualification* for (NCP) at $\bar{x} \in \Omega$. For instance, here are some examples of Constraint Qualifications for (NCP) that are inherited from the linear case:

- **Slater's CQ and Mangasarian-Fromovitz' CQ:** $T_\Omega(\bar{x})^\circ = L_\Omega(\bar{x})^\circ$ and there exists some $d \in \mathbb{R}^n$ such that
$$DG(\bar{x})[d] \in \mathrm{ri}(T_{\mathcal{K}}(G(\bar{x})));$$
- **Guignard's CQ:** $T_\Omega(\bar{x})^\circ = L_\Omega(\bar{x})^\circ$ and the set
$$H(\bar{x}) \doteq DG(\bar{x})^*[N_{\mathcal{K}}(G(\bar{x}))] \tag{2.24}$$
is closed, which comes from (2.21) since $N_{T_{\mathcal{K}}(G(\bar{x}))}(0) = N_{\mathcal{K}}(G(\bar{x}))$;

Another well-known condition that is very similar to Guignard's CQ, although it was defined independently of it, is the following:

- **Abadie's CQ:** $T_\Omega(\bar{x}) = L_\Omega(\bar{x})$ and $H(\bar{x})$ is closed [1];

We remark that Abadie's CQ and Slater's CQ (which is equivalent to Mangasarian-Fromovitz' CQ in this case) are both stronger than Guignard's CQ due to Proposition 2.2.1. In fact, even in the nonlinear conic optimization context Guignard's CQ is still the weakest possible Constraint Qualification [51]. It is also worth remarking that since $H(\bar{x})$ is always closed when $\mathcal{K}$ is a polyhedral cone, Abadie's CQ and Guignard's CQ are usually defined in the NLP literature as merely the equality between $T_\Omega(\bar{x})$ and $L_\Omega(\bar{x})$ and the equality between their polars, respectively.

Although stationarity is not a necessary optimality condition on its own since it depends on the fulfilment of a Constraint Qualification at local minimizers, observe that it defines a class of optimality conditions in the form "KKT or not-CQ" for any given CQ. For instance, every local minimizer must either satisfy KKT or violate Slater's CQ. In addition, weaker CQs define stronger optimality conditions, in the sense of logical implication; for example, since Guignard's CQ is weaker than (i.e., it is implied by) Slater's CQ, then the optimality condition "KKT or not-Guignard's CQ" is stronger than (i.e., it implies) "KKT or not-Slater's CQ." Stronger optimality conditions are generally more desirable for some applications, such as defining stopping criteria for algorithms, than weaker conditions because they are able to rule out a larger set of non-optimal points. In some sense, using stronger optimality conditions for this purpose makes such algorithms more *reliable*.

On the other hand, stronger Constraint Qualifications may also be interesting when they are able to describe special properties of $\Lambda(\overline{x})$. For instance, one of the most relevant Constraint Qualifications in the literature of Nonlinear Conic Programming is known as *Robinson's CQ*:

Definition 2.3.3 (Robinson's CQ). Let $\overline{x} \in \Omega$. We say that *Robinson's CQ* holds at $\overline{x}$ if

$$0 \in \text{int}\left(\text{Im}(DG(\overline{x})) - \mathcal{K} + G(\overline{x})\right).$$

By [37, Proposition 2.97 and Corollary 2.98], using the fact $\mathbb{Y}$ is finite-dimensional, we get that Robinson's CQ can be equivalently characterized as

$$\text{Im}(DG(\overline{x})) - T_{\mathcal{K}}(G(\overline{x})) = \mathbb{Y}, \tag{2.25}$$

and taking the polar of both sides wee see that (2.25) is in turn equivalent to

$$\text{Ker}(DG(\overline{x})^*) \cap N_{\mathcal{K}}(G(\overline{x})) = \{0\}. \tag{2.26}$$

If Robinson's CQ is fulfilled at a given point $\overline{x} \in \Omega$, then $\Lambda(\overline{x})$ is bounded [37, Theorem 3.9] besides being nonempty, which follows from the fact it implies Abadie's CQ[7]. Moreover, if $\Lambda(\overline{x})$ is nonempty and bounded, then $\overline{x}$ satisfies Robinson's CQ.

Lemma 2.3.1 (Lemma 2.99 of [37]). *Suppose that* $\text{int}(\mathcal{K}) \neq \emptyset$ *and let* $\overline{x} \in \Omega$. *Then, Robinson's CQ holds at* $\overline{x}$ *if, and only if, there exists some* $d \in \mathbb{R}^n$ *such that* $G(\overline{x}) + DG(\overline{x})[d] \in \text{int}(\mathcal{K})$.

Observe that Lemma 2.3.1 provides an alternative interpretation of Robinson's CQ as some analogue of Slater's CQ for the alternative linearized constraint $G(\overline{x}) + DG(\overline{x})[d] \in \mathcal{K}$ in the case $\text{int}(\mathcal{K}) \neq \emptyset$. It is also worth mentioning that in the NLP context Robinson's CQ coincides with Slater's CQ and Mangasarian-Fromovitz' CQ. We now recall another very relevant condition from the literature:

Definition 2.3.4 (Nondegeneracy). Let $\overline{x} \in \Omega$. We say that the *Nondegeneracy* (or *transversality*) condition holds at $\overline{x}$ if

$$\text{Im}(DG(\overline{x})) + \text{lin}(T_{\mathcal{K}}(G(\overline{x}))) = \mathbb{Y}.$$

Also, points that satisfy the Nondegeneracy condition are called *nondegenerate*.

The first definition of Nondegeneracy for a Nonlinear Conic Programming problem is due to Shapiro and Fan [85] for Semidefinite Programming, and a complete discussion on it can be found in the vicinity of [37, Definition 4.70]. Let us briefly summarize some fundamental aspects of it: first, it should be noted that since $\text{lin}(T_{\mathcal{K}}(G(\overline{x}))) \subseteq T_{\mathcal{K}}(G(\overline{x}))$, then Nondegeneracy implies Robinson's CQ, meaning it is also a Constraint Qualification. Moreover, [37, Proposition 4.75] states that if $\overline{x} \in \Omega$ is nondegenerate, then $\Lambda(\overline{x})$ is a singleton; conversely, if $\Lambda(\overline{x}) = \{\overline{\mu}\}$ and $\overline{\mu} \in \text{ri}(N_{\mathcal{K}}(G(\overline{x})))$, then $\overline{x}$ is nondegenerate. By the way, the existence of some Lagrange multiplier $\overline{\mu} \in \Lambda(\overline{x})$ such that $\overline{\mu} \in \text{ri}(N_{\mathcal{K}}(G(\overline{x})))$ is called the *strict complementarity* condition, which is essentially Slater's CQ for (Lin-D) when $\mathcal{K}$ is a cone.

[7] See, for instance, [37, Proposition 2.98] and the discussion below Definition 3.1.2 together with Theorem 3.1.4, or Serranoni's dissertation (in preparation), for a detailed analysis on this topic.

2.3.2 Second-order optimality conditions

Let us return to the linearized problem (Tan-P), built around an arbitrary point $\overline{x} \in \Omega$, and recall the geometric optimality conditions:

- If $\overline{x}$ is a local minimizer of (NCP), then $\langle \nabla f(\overline{x}), d\rangle \geqslant 0$ for every $d \in T_\Omega(\overline{x})$ (that is, $d = 0$ is a solution of (Tan-P));
- If $\langle \nabla f(\overline{x}), d\rangle > 0$ for every nonzero $d \in T_\Omega(\overline{x})$ (that is, if $d = 0$ is the unique solution of (Tan-P)), then $\overline{x}$ is a local minimizer of (NCP).

We invite the reader to give a more critic look at them. On the one hand, note that the sufficient condition is too strong – and sometimes meaningless – for instance, for unconstrained problems, where $\Omega = T_\Omega(\overline{x}) = \mathbb{R}^n$, it is never fulfilled. On the other hand, the necessary condition is too weak; again, for unconstrained problems for example we have $T_\Omega(\overline{x})^\circ = \{0\}$, so besides local minimizers, all saddle point and local maximizers of f will satisfy it.

To improve the "first-order conditions" of the previous subsection, we need additional information over the nonzero solutions of (Tan-P) which are usually called *critical tangent directions* of the original problem (NCP). The collection of such directions,

$$C_T(\overline{x}) \doteq \{d \in \mathbb{R}^n \colon d \in T_\Omega(\overline{x}),\ \langle \nabla f(\overline{x}), d\rangle = 0\}, \tag{2.27}$$

forms the so-called *critical tangent cone* of (NCP). The reason why they are "critical" is that they are somewhat "invisible" to first-order analysis; that is, the optimality of $\overline{x}$ ensures that every $d \in T_\Omega(\overline{x}) \setminus C_T(\overline{x})$ is an increasing direction for f at $\overline{x}$, but nothing can be concluded for the directions in $C_T(\overline{x})$. This is where the second-order tangent set to Ω at $\overline{x}$ comes into play. Following Ben-Tal and Zowe [30], if $d \in C_T(\overline{x})$ and $w \in T_\Omega^{\mathrm{in},2}(\overline{x}, d)$, then for every $\alpha > 0$ we have

$$\overline{x} + \alpha d + \frac{\alpha^2}{2} w + o(\alpha^2) \in \Omega$$

and if $\overline{x} \in \Omega$ is a local solution of (NCP), then after writing two consecutive Taylor expansions, we obtain

$$\begin{aligned} f\left(\overline{x} + \alpha d + \frac{\alpha^2}{2} w + o(\alpha^2)\right) &= f(\overline{x} + \alpha d) + \frac{\alpha^2}{2}\langle \nabla f(\overline{x}), w\rangle + o(\alpha^2) \\ &= f(\overline{x}) + \frac{\alpha^2}{2}\nabla^2 f(\overline{x})[d,d] + \frac{\alpha^2}{2}\langle \nabla f(\overline{x}), w\rangle + o(\alpha^2) \\ &\geqslant f(\overline{x}), \end{aligned}$$

which yields

$$\nabla^2 f(\overline{x})[d,d] + \langle \nabla f(\overline{x}), w\rangle \geqslant 0$$

for all such d and w. Thus,

- **Second-order geometric optimality condition:** If $\overline{x} \in \Omega$ is a local solution of (NCP), then $\langle \nabla f(\overline{x}), w\rangle + \nabla^2 f(\overline{x})[d,d] \geqslant 0$ for every $d \in C_T(\overline{x})$ and every $w \in T_\Omega^{\mathrm{in},2}(\overline{x}, d)$.

Just as KKT can be seen as a "friendly dual version" of the first-order geometric optimality condition, we may seek a dual version of the second-order condition as well. To do so, we will resort to a Constraint Qualification, of course. Let us recall from the seminal paper of Bonnans et al. [36], a characterization of the inner (and outer) second-order tangent sets to Ω at $\overline{x}$ in terms of the inner (respectively, outer) second-order tangent sets to $\mathcal{K}$ at $G(\overline{x})$.

Proposition 2.3.2 (Proposition 2.3 of [36]). *Let $\overline{x} \in \Omega$ satisfy Robinson's CQ*[8]*. Then, for every $d \in \mathbb{R}^n$ it holds that*

$$T_{\Omega}^{\mathsf{in},2}(\overline{x}, d) = DG(\overline{x})^{-1}\left(T_{\mathcal{K}}^{\mathsf{in},2}(G(\overline{x}), DG(\overline{x})[d]) - D^2G(\overline{x})[d,d]\right)$$

and

$$T_{\Omega}^{2}(\overline{x}, d) = DG(\overline{x})^{-1}\left(T_{\mathcal{K}}^{2}(G(\overline{x}), DG(\overline{x})[d]) - D^2G(\overline{x})[d,d]\right).$$

Clearly, Proposition 2.3.2 looks like a second-order analogue of Abadie's CQ, and it indeed plays the same role as Abadie's CQ for second-order analysis. However, the dual version of the second-order geometric condition does not follow immediately from strong duality as in the linear case. For the sake of understanding we will present next the result of Bonnans et al. [36] together with its full proof[9].

Theorem 2.3.3 (Theorem 3.1 of [36]). *Let $\overline{x} \in \Omega$ be a local solution of (NCP) that satisfies Robinson's CQ. Then, for every $d \in C_T(\overline{x})$ it holds that*

$$\sup_{\mu \in \Lambda(\overline{x})} \left(\nabla^2 f(\overline{x})[d,d] + \langle D^2G(\overline{x})[d,d], \mu\rangle - \sigma(\mu, T_{\mathcal{K}}^{\mathsf{in},2}(G(\overline{x}), DG(\overline{x})[d]))\right) \geqslant 0. \tag{2.28}$$

Proof. Let $d \in C_T(\overline{x})$ and by Robinson's CQ we have $DG(\overline{x})[d] \in T_{\mathcal{K}}(G(\overline{x}))$. Since $\mathcal{K}$ is closed and convex, this also implies that $T_{\mathcal{K}}^{\mathsf{in},2}(G(\overline{x}), DG(\overline{x})[d]) \neq \emptyset$. The idea of this proof is essentially the same as the derivation of the KKT conditions via strong duality of (Lin-P) and (Lin-D). For this, consider the linear approximation of (NCP) around $\overline{x}$ with a second-order displacement along d given by

$$\begin{aligned} &\underset{w \in \mathbb{R}^n}{\text{Minimize}} && \langle \nabla f(\overline{x}), w\rangle + \nabla^2 f(\overline{x})[d,d] \\ &\text{subject to} && DG(\overline{x})[w] + D^2G(\overline{x})[d,d] \in T_{\mathcal{K}}^{\mathsf{in},2}(G(\overline{x}), DG(\overline{x})[d]). \end{aligned} \tag{2Lin-P}$$

By Proposition 2.3.2 we observe that the feasible set of (2Lin-P) is precisely $T_{\Omega}^{\mathsf{in},2}(\overline{x}, d)$. Then, by the optimality of $\overline{x}$ and the second-order geometric optimality condition, it follows that (2Lin-P) is bounded below by zero. We can also compute the dual problem of (2Lin-P) analogously to the dual of (LP) – the extra constant term $\nabla^2 f(\overline{x})[d,d]$ can be carried through the whole computation – which yields:

$$\begin{aligned} &\underset{\mu \in \mathbb{Y}}{\text{Maximize}} && \langle D^2G(\overline{x})[d,d], \mu\rangle + \nabla^2 f(\overline{x})[d,d] - \sigma(\mu, T_{\mathcal{K}}^{\mathsf{in},2}(G(\overline{x}), DG(\overline{x})[d])) \\ &\text{subject to} && \nabla f(\overline{x}) + DG(\overline{x})^*[\mu] = 0. \end{aligned} \tag{2Lin-D}$$

Note that the "sigma-term" is inherited from (LD) because $T_{\mathcal{K}}^{\mathsf{in},2}(G(\overline{x}), DG(\overline{x})[d])$ is not necessarily a cone. To proceed, observe that

$$T_{\mathcal{K}}(G(\overline{x})) \subseteq T_{T_{\mathcal{K}}(G(\overline{x}))}(DG(\overline{x})[d]) = \operatorname{cl}(T_{\mathcal{K}}(G(\overline{x})) + \mathsf{span}(DG(\overline{x})[d])),$$

so it follows from Proposition 2.1.18 that for every $\mu \in T_{\mathcal{K}}^{\mathsf{in},2}(G(\overline{x}), DG(\overline{x})[d])$ we have

$$\mu + T_{\mathcal{K}}(G(\overline{x})) \subseteq T_{\mathcal{K}}^{\mathsf{in},2}(G(\overline{x}), DG(\overline{x})[d]) \tag{2.29}$$

[8] A recent work [70, Section 4] shows that this result can also be obtained under Metric Subregularity CQ, which is strictly weaker than Robinson's CQ. Thanks to Patrick Mehlitz for pointing this out!

[9] *Another personal note*: I strongly recommend carefully reading the proof of Theorem 2.3.3 because, in my opinion, it may appear to be much more complicated than it actually is, if you don't read its full proof.

which implies

$$\sigma(\mu, \mu + T_{\mathcal{K}}(G(\overline{x}))) = \sigma(\mu, \mu) + \sigma(\mu, T_{\mathcal{K}}(G(\overline{x}))) \leqslant \sigma(\mu, T^{\text{in},2}_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d]))$$

but since $\sigma(\mu, T_{\mathcal{K}}(G(\overline{x}))) = \delta(\mu, N_{\mathcal{K}}(G(\overline{x})))$ we get that $\sigma(\mu, T^{\text{in},2}_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d])) = \infty$ for every $\mu \notin N_{\mathcal{K}}(G(\overline{x}))$ and, by consequence, the effective domain of (2Lin-D) is included in $N_{\mathcal{K}}(G(\overline{x}))$ hence its feasible set is contained in $\Lambda(\overline{x})$[10].

Finally, we point out that Robinson's CQ (2.25) together with (2.29) imply

$$\mathbb{Y} \subseteq \mathbb{Y} - \mu \subseteq \text{Im}(DG(\overline{x})) - T_{\mathcal{K}}(G(\overline{x})) - \mu \subseteq \text{Im}(DG(\overline{x})) - T^{\text{in},2}_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d])$$

for every $\mu \in T^{\text{in},2}_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d]) \subseteq \mathbb{Y}$, which in turn implies that

$$-D^2G(\overline{x})[d, d] \in \text{int}(\text{Im}(DG(\overline{x})) - T^{\text{in},2}_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d]))$$

and Robinson's CQ holds at every feasible point of (2Lin-D). Thus, strong duality holds between (2Lin-P) and (2Lin-D) but since the former is bounded below by zero, we can conclude that

$$\sup_{\mu \in \Lambda(\overline{x})} \Big(\nabla^2 f(\overline{x})[d, d] + \langle D^2G(\overline{x})[d, d], \mu\rangle - \sigma(\mu, T^{\text{in},2}_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d]))\Big) \geqslant 0.$$

□

As a technical note, it follows from Proposition 2.1.18 that if $G(\overline{x}) \in \mathcal{K}$, $DG(\overline{x})[d] \in T_{\mathcal{K}}(G(\overline{x}))$, and $0 \in T^{\text{in},2}_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d])$ – a condition that is always satisfied when $\mathcal{K}$ is a polyhedral cone, for instance – then

$$T^{\text{in},2}_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d]) = T_{T_{\mathcal{K}}(G(\overline{x}))}(DG(\overline{x})[d]), \tag{2.30}$$

meaning it is a closed convex cone and, in this case,

$$\sigma(\mu, T^{\text{in},2}_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d])) = \delta(\mu, N_{\mathcal{K}}(G(\overline{x})) \cap \{DG(\overline{x})[d]\}^{\perp}) = 0$$

because $\mu \in N_{\mathcal{K}}(G(\overline{x}))$ and $\langle \mu, DG(\overline{x})[d]\rangle = \langle -\nabla f(\overline{x}), d\rangle = 0$ for every feasible point μ of (2Lin-D).

With the proof of Theorem 2.3.3 in mind we notice that any condition at $\overline{x} \in \Omega$ that ensures the equality $T_{\Omega}(\overline{x}) = L_{\Omega}(\overline{x})$ and the equality of Proposition 2.3.2 as well as strong duality for both linearized problems, (Lin-P) and (2Lin-P), would in turn imply (2.28). For instance:

- **Second-order Abadie's CQ:** $T_{\Omega}(\overline{x}) = L_{\Omega}(\overline{x})$, $H(\overline{x})$ is closed and, for every $d \in C_T(\overline{x})$, it holds that

  $$T^{\text{in},2}_{\Omega}(\overline{x}, d) = L^{\text{in},2}_{\Omega}(\overline{x}, d) \doteq DG(\overline{x})^{-1}\left(T^{\text{in},2}_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d]) - D^2G(\overline{x})[d, d]\right)$$

  and the set

  $$H^{\text{in},2}(\overline{x}, d) = DG(\overline{x})^*\left[N_{T^{\text{in},2}_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d])}(D^2G(\overline{x})[d, d])\right]$$

  is closed.

[10] *Future work proposal:* It may be interesting to investigate under which conditions the effective domain of (2Lin-D) is equal to $\Lambda(\overline{x})$.

Furthermore, note that the sigma-term can be dropped if $T_{\mathcal{K}}^{\text{in},2}(G(\overline{x}), DG(\overline{x})[d])$ is a cone; for instance, this is always true if $\mathcal{K}$ is a polyhedral cone due to (2.30). As observed by Bonnans et al. [36], it is not mandatory to consider the inner second-order tangent set to derive Theorem 2.3.3; in fact, the same result could be obtained in terms of any convex subset of the outer second-order tangent set, say $S(\overline{x}, d) \subseteq T_{\mathcal{K}}^2(G(\overline{x}), DG(\overline{x})[d])$. In this case, the sharpest result of this kind is given by the largest $S(\overline{x}, d)$ by inclusion. Under second-order regularity (Definition 2.1.5) the inner and outer sets coincide and Theorem 2.3.3 already gives such a "sharpest result."

As a final comment, we remark that if $T_\Omega(\overline{x}) = L_\Omega(\overline{x})$ and strict complementarity holds; that is, if there exists a Lagrange multiplier $\overline{\mu} \in \Lambda(\overline{x})$ such that $\overline{\mu} \in \text{ri}(N_{\mathcal{K}}(G(\overline{x})))$, then $C_T(\overline{x})$ reduces to a subspace. Indeed, note that for every $d \in C_T(\overline{x})$,

$$\langle \nabla f(\overline{x}), d \rangle = \langle \overline{\mu}, DG(\overline{x})[d] \rangle = 0,$$

hence $\overline{\mu} \in N_{\mathcal{K}}(G(\overline{x})) \cap \{DG(\overline{x})[d]\}^\perp$ which is a face of $N_{\mathcal{K}}(G(\overline{x}))$; but since $N_{\mathcal{K}}(G(\overline{x}))$ itself is its smallest face that contains $\overline{\mu}$, it must hold that $N_{\mathcal{K}}(G(\overline{x})) \subseteq \{DG(\overline{x})[d]\}^\perp$ and, consequently, $\text{span}(DG(\overline{x})[d]) \subseteq T_{\mathcal{K}}(G(\overline{x}))$. Then, $DG(\overline{x})[d] \in \text{lin}(T_{\mathcal{K}}(G(\overline{x})))$ and $d \in \text{lin}(C_T(\overline{x}))$. Since d is arbitrary, we conclude that $C_T(\overline{x}) = \text{lin}(C_T(\overline{x}))$ in this case. This means that in the presence of strict complementarity, condition (2.28) needs to be evaluated over a subspace instead of a cone, which is a simpler task from the computational point of view.

In this work, we will make distinction among all these variants of second-order conditions due to the different natures of the assumptions required to obtain each of them. With this in mind:

Definition 2.3.5. Let $\overline{x} \in \Omega$ satisfy the KKT conditions and, for any $d \in \mathbb{R}^n$ and any $\mu \in \mathbb{Y}$ consider the inequality

$$\nabla^2 f(\overline{x})[d, d] + \langle D^2 G(\overline{x})[d, d], \mu \rangle - \sigma(\mu, T_{\mathcal{K}}^{\text{in},2}(G(\overline{x}), DG(\overline{x})[d])) \geqslant 0. \qquad (2.31)$$

We say that $\overline{x}$ satisfies the:

1. **Basic Second-Order Optimality Condition** (BSOC) if for each $d \in C_T(\overline{x})$ there exists some $\mu \in \Lambda(\overline{x})$ (possibly dependent on d) such that (2.31) holds;

2. **Semi-Strong Second-Order Optimality Condition** (Semi-SSOC) if there exists some $\mu \in \Lambda(\overline{x})$ such that (2.31) holds for every $d \in C_T(\overline{x})$;

3. **Strong Second-Order Optimality Condition** (SSOC) if (2.31) holds for every $d \in C_T(\overline{x})$ and every $\mu \in \Lambda(\overline{x})$;

4. **Weak Second-Order Optimality Condition** (WSOC) if there exists some $\mu \in \Lambda(\overline{x})$ such that (2.31) holds for every $d \in \text{lin}(C_T(\overline{x}))$.

Clearly, SSOC implies Semi-SSOC which implies both WSOC and BSOC which in turn seem unrelated. To illustrate how these conditions may be different, it is known from the literature that Robinson's CQ is sufficient for obtaining BSOC at local minimizers, but an example by Baccari [28] showed that is not enough for WSOC, Semi-SSOC, or SSOC. In NLP only, Andreani et al. [5] were able to prove that WSOC may hold when Robinson's CQ is fulfilled together with the so-called *Weak Constant Rank* (WCR) property, which will be discussed in details in a future section. However, it is not known whether these conditions together ensure SSOC and Semi-SSOC or not. To obtain results regarding SSOC (and consequently, Semi-SSOC) Andreani et al. [22] employed a stronger constant rank-type condition introduced by Janin [57], called *Constant Rank Constraint Qualification* (CRCQ). Prior to their work, such a result was only known to be obtainable under

Nondegeneracy, even though this is quite obvious since the uniqueness of Lagrange multipliers makes the second-order conditions listed as 1–3 in Definition 2.3.5 equivalent.

2.3.3 Cone reducibility

The so-called *reduction approach* of Bonnans and Shapiro [37] can be used to extend all of the conclusions of the previous two subsections for the case where $\mathcal{K}$ is not necessarily a cone, as long as it *looks like* a cone in a neighborhood of $G(\overline{x})$ where $\overline{x} \in \Omega$ is a given point of interest. When K is itself a cone, the reduction approach can also be conveniently employed to simplify the problem since it allows us to roughly translate $G(\overline{x})$ to the vertex of $\mathcal{K}$ without losing structure. Let us recall the general definition of reducibility:

Definition 2.3.6 (Definition 3.135 of [37]). Let $\mathbb{Y}$ and $\mathbb{F}$ be finite-dimensional linear spaces, and let $\mathcal{K} \subseteq \mathbb{Y}$ and $\mathcal{C} \subseteq \mathbb{F}$ be nonempty closed convex sets; moreover, assume that $\mathcal{C}$ is a pointed cone. We say that $\mathcal{K}$ is *$\mathscr{C}^2$-reducible* (or simply *reducible*) to $\mathcal{C}$ at a point $y \in \mathcal{K}$ if there exists a neighborhood $\mathcal{N}$ of y and a twice continuously differentiable reduction function $\Xi\colon \mathcal{N} \to \mathbb{F}$ (possibly depending on y) such that:

1. $\Xi(y) = 0$;
2. $D\Xi(y)$ is surjective;
3. $\mathcal{K} \cap \mathcal{N} = \{z \in \mathcal{N} \colon \Xi(z) \in \mathcal{C}\}$.

In the context of optimization problems such as (NCP), given any point $\overline{x} \in \Omega$ such that $\mathcal{K}$ is reducible to $\mathcal{C}$ at $G(\overline{x})$ by the reduction function Ξ, we can define a new constraint function

$$\mathcal{G} \doteq \Xi \circ G$$

and the reduced constraint $\mathcal{G}(x) \in \mathcal{C}$ turns out to be equivalent to the original constraint $G(x) \in \mathcal{K}$ in a sufficiently small neighborhood of $\overline{x}$.

Example 2.3.1 (Nonlinear Programming). *For the particular case of NLP, where $\mathbb{Y} = \mathbb{R}^m$, $\mathcal{K} = \mathbb{R}^m_+$, and $G(x) \doteq (g_1(x), \dots, g_m(x))$, an example of reduction is to simply isolate* active constraints*; that is, the constraints indexed by*

$$\mathcal{A}(\overline{x}) \doteq \{j \in \{1, \dots, m\} \colon g_j(\overline{x}) = 0\}$$

in a neighborhood of $G(\overline{x})$ for a given point $\overline{x} \in \Omega$. This can be done by setting $\mathcal{C} = \mathbb{R}^{|\mathcal{A}(\overline{x})|}_+$ and the reduction mapping is given by $\Xi(y_1, \dots, y_m) \doteq [y_i]_{i \in \mathcal{A}(\overline{x})}$. Then, at least in a neighborhood of $\overline{x}$, the original constraint $G(x) \in \mathbb{R}^m_+$ is equivalent to the reduced constraint

$$\mathcal{G}(x) \doteq [g_i(x)]_{i \in \mathcal{A}(\overline{x})} \in \mathbb{R}^{|\mathcal{A}(\overline{x})|}_+. \tag{2.32}$$

With this in mind, in this section we will focus on the reduced problem below and clarify which results from the previous sections are kept and which are lost in the reduction process.

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & \mathcal{G}(x) \in \mathcal{C}. \end{aligned} \tag{Red-NCP}$$

Let us begin by recalling some classical correspondences between (NCP) and (Red-NCP) from Bonnans and Shapiro's book [37, Section 3.4.4]. First, observe that

$$D\Xi(G(\overline{x}))[T_{\mathcal{K}}(G(\overline{x}))] = T_{\mathcal{C}}(\mathcal{G}(\overline{x})) = \mathcal{C}$$

because $\mathcal{G}(\overline{x}) = 0$ whence follows that $N_{\mathcal{K}}(G(\overline{x})) = D\Xi(G(\overline{x}))^*[\mathcal{C}^\circ]$ as it is also stated in [37, Equation 3.267]. Then, we use the chain rule

$$D\mathcal{G}(\overline{x})[\,\cdot\,] = D\Xi(G(\overline{x}))[DG(\overline{x})[\,\cdot\,]]$$

to conclude that

$$\begin{aligned} L_\Omega(\overline{x}) &= \bigcap_{\mu\in N_{\mathcal{K}}(G(\overline{x}))} \{d \in \mathbb{R}^n \colon \langle DG(\overline{x})[d], \mu\rangle \leqslant 0\} \\ &= \bigcap_{\eta\in\mathcal{C}^\circ} \{d \in \mathbb{R}^n \colon \langle DG(\overline{x})[d], D\Xi(G(\overline{x}))^*[\eta]\rangle \leqslant 0\} \\ &= \bigcap_{\eta\in\mathcal{C}^\circ} \{d \in \mathbb{R}^n \colon \langle D\mathcal{G}(\overline{x})[d], \eta\rangle \leqslant 0\} \\ &= \{d \in \mathbb{R}^n \colon D\mathcal{G}(\overline{x})[d] \in \mathcal{C}\}; \end{aligned} \tag{2.33}$$

that is, the linearized cone of (NCP) coincides with the linearized cone of the reduced problem (Red-NCP). Similarly, recall the set $H(\overline{x})$ defined in (2.24) and notice that

$$\begin{aligned} H(\overline{x}) &= DG(\overline{x})[N_{\mathcal{K}}(G(\overline{x}))] \\ &= DG(\overline{x})^*[D\Xi(G(\overline{x}))^*[\mathcal{C}^\circ]] \\ &= D\mathcal{G}(\overline{x})^*[\mathcal{C}^\circ], \end{aligned} \tag{2.34}$$

meaning it also coincides with its counterpart defined for the reduced problem (Red-NCP). Moreover, the Lagrange multiplier sets $\Lambda(\overline{x})$ and $\mathcal{M}(\overline{x})$, of (NCP) and (Red-NCP) respectively, both associated with $\overline{x}$, satisfy the relation

$$\Lambda(\overline{x}) = D\Xi(G(\overline{x}))^*[\mathcal{M}(\overline{x})],$$

because $N_{\mathcal{K}}(G(\overline{x})) = D\Xi(G(\overline{x}))^*[\mathcal{C}^\circ]$.

Equations (2.33) and (2.34) imply that Guignard's CQ (and Abadie's CQ) is invariant to reduction mappings. In fact, Robinson's CQ is also invariant to reductions since $\mathrm{Im}(DG(\overline{x})) - T_{\mathcal{K}}(G(\overline{x})) = \mathbb{Y}$ holds if, and only if, $\mathrm{Im}(D\mathcal{G}(\overline{x})) - \mathcal{C} = \mathbb{F}$ due to the surjectivity of $D\Xi(G(\overline{x}))$. Similarly, the Nondegeneracy condition holds for (NCP) at $\overline{x}$ if, and only if, the linear mapping $D\mathcal{G}(\overline{x})$ is surjective.

Another useful property of reductions is the following:

Proposition 2.3.4 (Proposition 3.136 of [37]). *If $\mathcal{K}$ is reducible to $\mathcal{C}$ at $y \in \mathcal{K}$ by the reduction mapping Ξ, then for every $h \in \mathcal{C}$ it holds that*

$$T^{\text{in},2}_{\mathcal{C}}(0,h) = T^2_{\mathcal{C}}(0,h) = T_{\mathcal{C}}(h)$$

and, in particular, $\mathcal{C}$ is second-order regular at $\Xi(y) = 0$ and $\mathcal{K}$ is second-order regular at y. In addition, for every $d \in T_{\mathcal{K}}(y)$ it holds that

$$\begin{aligned} T^{\text{in},2}_{\mathcal{K}}(y,d) = T^2_{\mathcal{K}}(y,d) &= D\Xi(y)^{-1}\left(T^2_{\mathcal{C}}(0, D\Xi(y)[d]) - D^2\Xi(y)[d,d]\right) \\ &= D\Xi(y)^{-1}\left(T_{\mathcal{C}}(D\Xi(y)[d]) - D^2\Xi(y)[d,d]\right). \end{aligned} \tag{2.35}$$

Still following Bonnans and Shapiro [37, Section 3.4.4], it follows from (2.33) that the critical cone of (NCP) is the same as the critical of (Red-NCP); furthermore, we obtain from Proposition 2.3.4 the following:

Lemma 2.3.5. *Let $\overline{x} \in \Omega$ be a point such that $\Lambda(\overline{x}) \neq \emptyset$ and suppose that $\mathcal{K}$ is reducible to $\mathcal{C}$ at $G(\overline{x})$ by the reduction mapping Ξ. Then, for every $\overline{\mu} \in \Lambda(\overline{x})$, every $d \in C_T(\overline{x})$, and every $\overline{\eta} \in \mathcal{C}^{\circ}$ such that $\overline{\mu} = D\Xi(G(\overline{x}))^*[\overline{\eta}]$ the relation*

$$\sigma(\overline{\mu},\ T_{\mathcal{K}}^{\mathrm{in},2}(G(\overline{x}), DG(\overline{x})[d])) = -\left\langle \overline{\eta},\ D^2\Xi(G(\overline{x}))[DG(\overline{x})[d], DG(\overline{x})[d]] \right\rangle \tag{2.36}$$

holds true.

Proof. By direct computation:

$$\begin{aligned}\sigma(\overline{\mu},\ T_{\mathcal{K}}^{\mathrm{in},2}(G(\overline{x}), DG(\overline{x})[d])) &= \sup\left\{\langle \overline{\eta}, D\Xi(G(\overline{x}))[z]\rangle \colon z \in T_{\mathcal{K}}^{\mathrm{in},2}(G(\overline{x}), DG(\overline{x})[d])\right\} \\ &= \delta(\overline{\eta},\ N_{\mathcal{C}}(D\Xi(G(\overline{x}))[DG(\overline{x})[d]]))+ \\ &\qquad + \sigma(\overline{\eta},\ -D^2\Xi(G(\overline{x}))[DG(\overline{x})[d], DG(\overline{x})[d]]) \\ &= -\left\langle \overline{\eta},\ D^2\Xi(G(\overline{x}))[DG(\overline{x})[d], DG(\overline{x})[d]] \right\rangle.\end{aligned}$$

To get the third equation from the second, we used that

$$\langle \overline{\eta},\ D\Xi(G(\overline{x}))[DG(\overline{x})[d]]\rangle = \langle \overline{\mu}, DG(\overline{x})[d])\rangle = \langle -\nabla f(\overline{x}), d\rangle = 0$$

so $\overline{\eta} \in N_{\mathcal{C}}(D\Xi(G(\overline{x}))[DG(\overline{x})[d]])$ and consequently $\delta(\overline{\eta},\ N_{\mathcal{C}}(D\Xi(G(\overline{x}))[DG(\overline{x})[d]])) = 0$. □

Thus, the second-order optimality condition for (NCP) can be explicitly recovered from the second-order condition for (Red-NCP) through Equation (2.36).

Chapter 3

New optimality conditions for Nonlinear Conic Programming

A huge part of the theory of optimality for nonlinear conic programs relies on Constraint Qualifications to be derived, which is reasonable given that even the (very strong) Nondegeneracy condition is *generic* in the sense that it holds almost everywhere [78]. Consequently, all of the Constraint Qualifications and theorems that follow from Nondegeneracy are also generic, at least theoretically. However, reality is a bit cruel and it is known that many important optimization models of real world problems happen to have degenerate solutions [83]. Besides, deciding whether a given Constraint Qualification holds or not at a particular point can be quite hard, from the computational point of view. This is where the so-called *sequential optimality conditions* come into play.

In summary, sequential optimality conditions are *optimality conditions certified by sequences that can be computed in practice*. To explain the meaning of our previous sentence, let us show how sequential optimality conditions are born[1].

3.1 Approximate Karush-Kuhn-Tucker

We begin this section with an intuitive construction of the simplest sequential optimality condition presented in our first paper [9], called the *Approximate KKT* (AKKT) condition. To do so, let $\overline{x} \in \Omega$ be a local minimizer of (NCP) and recall the linearized problem around $\overline{x}$, given by:

$$\begin{aligned} \underset{d\in\mathbb{R}^n}{\text{Minimize}} \quad & \langle \nabla f(\overline{x}), d\rangle \\ \text{subject to} \quad & DG(\overline{x})[d] \in T_{\mathcal{K}}(G(\overline{x})). \end{aligned} \tag{Lin-P}$$

Also, recall that $\overline{x}$ satisfies the KKT conditions if, and only if, strong duality holds for (Lin-P), and in this case its dual solutions are called Lagrange multipliers associated with $\overline{x}$. However, even if strong duality does not hold, we can at least prove existence of a sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and some sequences of perturbations, $\{\delta^k\}_{k\in\mathbb{N}} \to 0$ and $\{\Delta^k\}_{k\in\mathbb{N}} \to 0$, such that for each $k \in \mathbb{N}$ the approximate problem:

$$\begin{aligned} \underset{d\in\mathbb{R}^n}{\text{Minimize}} \quad & \langle \nabla f(x^k) + \delta^k, d\rangle \\ \text{subject to} \quad & DG(x^k)[d] \in T_{\mathcal{K}}(G(x^k) + \Delta^k). \end{aligned} \tag{Lin-P(k)}$$

[1] *Personal note:* This is not exactly how they were originally presented in the NLP world, but it is how we reinvented them in order to extend them to NCP.

satisfies strong duality. Roughly, this is a consequence of the genericity of Nondegeneracy[2]. This "asymptotic" kind of strong duality is AKKT, which has been initially proposed by Andreani et al. [10] for NLP problems and since then has gained many variants and extensions to several other contexts – see the references and the Introduction of [9] for more details. The appealing feature of AKKT is the fact that the sequences that certify its fulfilment can be computed in practice; and not only that, they can be computed along the execution of some algorithms that were originally designed to solve (NCP); there is no need to develop new algorithms for that, although one is encouraged to design new algorithms with them in mind. In other words, the output sequences of such algorithms can be used as optimality certificates for their limit points. Let us illustrate this with an *external penalty method*.

Theorem 3.1.1 (Extracted from Theorem 2 of [9]). *Let $\overline{x} \in \Omega$ be a local minimizer of (NCP) 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:*

$$F_k(x) \doteq f(x) + \frac{\rho_k}{2}\|\Pi_{\mathcal{K}^\circ}(G(x))\|^2 + \frac{1}{2}\|x - \overline{x}\|^2.$$

Proof. Let $\delta > 0$ be the radius of optimality of $\overline{x}$ and consider the regularized penalized subproblem of (NCP):

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & F_k(x) \\ \text{subject to} \quad & \|x - \overline{x}\| \leqslant \delta. \end{aligned} \tag{Reg-P(k)}$$

Let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence of local minimizers of (Reg-P(k)), which is of course bounded, and let $\overline{w}$ be an arbitrary limit point of it. Let $I \subseteq_\infty \mathbb{N}$ be such that $\lim_{k\in I} x^k = \overline{w}$. Then, by the optimality of x^k, we have $F_k(x^k) \leqslant F_k(\overline{x})$ for every $k \in \mathbb{N}$, which implies

$$\frac{\rho_k}{2}\|\Pi_{\mathcal{K}^\circ}(G(\overline{x}))\|^2 \geqslant f(x^k) - f(\overline{x}) + \frac{\rho_k}{2}\|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 + \frac{1}{2}\|x^k - \overline{x}\|^2.$$

Dividing everything by $\frac{\rho_k}{2}$, we obtain

$$\|\Pi_{\mathcal{K}^\circ}(G(\overline{x}))\|^2 \geqslant \frac{2(f(x^k) - f(\overline{x}))}{\rho_k} + \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 + \frac{\|x^k - \overline{x}\|^2}{\rho_k} \geqslant \frac{2(f(x^k) - f(\overline{x}))}{\rho_k}$$

for every $k \in \mathbb{N}$. Knowing that $\Pi_{\mathcal{K}^\circ}(G(\overline{x})) = 0$, $\{x^k\}_{k\in\mathbb{N}}$ is bounded, and $\rho_k \to \infty$, we obtain

$$\lim_{k\in I} \|\Pi_{\mathcal{K}^\circ}(G(x^k))\| = 0,$$

which means $G(\overline{w}) \in \mathcal{K}$. Moreover, note that for any $z \in \Omega$ such that $\|z - \overline{x}\| \leqslant \delta$ we have

$$f(x^k) + \frac{1}{2}\|x^k - \overline{x}\|^2 \leqslant F_k(x^k) \leqslant F_k(z) = f(z) + \frac{1}{2}\|z - \overline{x}\|^2.$$

Taking limits in I, we obtain $f(\overline{w}) + \frac{1}{2}\|\overline{w} - \overline{x}\|^2 \leqslant f(z) + \frac{1}{2}\|z - \overline{x}\|^2$. Hence, $\overline{w}$ is a global

[2] *Future work proposal:* I would love to have a proof of existence of AKKT sequences via "genericity of Nondegeneracy" without resorting to an external penalty method.

minimizer of the following localized problem:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & f(x) + \frac{1}{2}\|x-\overline{x}\|^2 \\ \text{subject to} & G(x) \in \mathcal{K} \\ & \|x-\overline{x}\| \leqslant \delta, \end{array} \tag{Loc-P}$$

but the unique global minimizer of (Loc-P) is $\overline{x}$, meaning that $\overline{w} = \overline{x}$. For $k \in I$ sufficiently large, we have $\|x^k - \overline{x}\| < \delta$ and thus, for every such k, the point x^k is a local minimizer of F_k, unconstrained. □

Now, let $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ be the sequence predicted by Theorem 3.1.1 and by the stationarity of each x^k with respect to F_k we obtain

$$\nabla f(x^k) + (x^k - \overline{x}) + DG(x^k)^*[\rho_k \Pi_{\mathcal{K}^\circ}(G(x^k))] = 0 \tag{3.1}$$

for every $k \in \mathbb{N}$. Take $\delta^k \doteq x^k - \overline{x}$ and $\Delta^k \doteq -\Pi_{\mathcal{K}^\circ}(G(x^k))$ for every $k \in \mathbb{N}$, and note that $\mu^k \doteq \rho_k \Pi_{\mathcal{K}^\circ}(G(x^k))$ is a feasible point of the dual linearized problem:

$$\begin{array}{ll} \underset{\mu\in N_{\mathcal{K}}(G(x^k)+\Delta^k)}{\text{Maximize}} & 0 \\ \text{subject to} & \nabla f(x^k) + \delta^k + DG(x^k)^*[\mu] = 0. \end{array} \tag{Lin-D(k)}$$

for every $k \in \mathbb{N}$ due to Moreau's decomposition (Theorem 2.1.14). Thus, strong duality holds between (Lin-P(k)) and (Lin-D(k)) for every $k \in \mathbb{N}$. Of course, when the sequence $\{\mu^k\}_{k\in\mathbb{N}}$ has a bounded subsequence, then all of its limit points are Lagrange multipliers associated with $\overline{x}$ with respect to (NCP). The sequences we have just built are called *AKKT sequences* and the existence of an AKKT sequence convergent to a given point characterizes what we call an *AKKT point*. To define it in a proper mathematical environment with standard notation, recall the *Lagrangian function* of (NCP), $L\colon \mathbb{R}^n \times \mathbb{Y} \to \mathbb{R}$, which is defined as

$$L(x,\mu) \doteq f(x) + \langle G(x), \mu\rangle$$

for every $x \in \mathbb{R}^n$ and $\mu \in \mathbb{Y}$. For the record, notice that the derivative of L with respect to the first-variable can be computed as follows:

$$\nabla_x L(x,\mu) = \nabla f(x) + DG(x)^*[\mu]$$

at any $x \in \mathbb{R}^n$ and $\mu \in \mathbb{Y}$. Then:

Definition 3.1.1 (Definition 4 of [9]). Let $\overline{x} \in \Omega$. We say that $\overline{x}$ is an *AKKT point* (alternatively, that $\overline{x}$ satisfies the *AKKT condition*) if there exists a sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, an approximate Lagrange multiplier sequence $\{\mu^k\}_{k\in\mathbb{N}} \subseteq \mathbb{Y}$, and some perturbation sequences $\{\delta^k\}_{k\in\mathbb{N}} \to 0$ and $\{\Delta^k\}_{k\in\mathbb{N}} \to 0$ such that, for every $k \in \mathbb{N}$:

- $\nabla_x L(x^k, \mu^k) + \delta^k = 0$;
- $\mu^k \in N_{\mathcal{K}}(G(x^k) + \Delta^k)$.

Observe that $N_{\mathcal{K}}(G(x^k) + \Delta^k) \neq \emptyset$ also implies that $G(x^k) + \Delta^k \in \mathcal{K}$, for each $k \in \mathbb{N}$.

The price we pay for such a convenient form of certifying optimality is that in certain situations AKKT may be too generic; for instance, there may be infinitely many non-locally optimal points that happen to be AKKT (even the whole feasible set), as we show in the next example, borrowed from [10].

Example 3.1.1 (Counterexample 3.1 of [10]). *Let* $\mathcal{K} \doteq \{0\} \times \mathbb{R}_+ \subseteq \mathbb{R}^2$ *and consider the problem:*

$$\begin{aligned} \underset{x\in\mathbb{R}^2}{\text{Minimize}} \quad & -x_2, \\ \text{subject to} \quad & (x_1x_2,\ x_1) \in \mathcal{K}, \end{aligned}$$

which is unbounded since $(0,\alpha)$ *is feasible for every* $\alpha \geqslant 0$. *Moreover, each point in the form* $\overline{x} \doteq (0,\alpha)$ *with* $\alpha \geqslant 0$ *is also AKKT with respect to the sequences* $\{x^k\}_{k\in\mathbb{N}}$, $\{\mu^k\}_{k\in\mathbb{N}}$, $\{\delta^k\}_{k\in\mathbb{N}}$, *and* $\{\Delta^k\}_{k\in\mathbb{N}}$ *given by:*

$$x^k \doteq \left(\frac{1}{k},\ \alpha\right), \quad \mu^k \doteq (k, -k\alpha), \quad \delta^k \doteq 0, \quad \Delta^k \doteq -G(x^k) = -\left(\frac{\alpha}{k},\ \frac{1}{k}\right)$$

because in this case

$$\nabla f(x^k) + DG(x^k)^*[\mu^k] = \begin{bmatrix} 0 \\ -1 \end{bmatrix} + \begin{bmatrix} \alpha & 1 \\ 1/k & 0 \end{bmatrix} \begin{bmatrix} k \\ -k\alpha \end{bmatrix} = 0$$

and $G(x^k) + \Delta^k = 0 \in \mathcal{K}$ *and* $\mu^k \in N_{\mathcal{K}}(0) = \mathcal{K}^\circ = \mathbb{R} \times \mathbb{R}_-$ *for every* $k \in \mathbb{N}$. *In fact, even if* $\alpha < 0$ *the point* $\overline{x}$ *happens to be AKKT with* $\{x^k\}_{k\in\mathbb{N}}$ *and* $\{\mu^k\}_{k\in\mathbb{N}}$ *defined by*

$$x^k \doteq \left(-\frac{1}{k},\ \alpha\right), \quad \mu^k \doteq (-k, k\alpha)$$

for every $k \in \mathbb{N}$, *and all the other sequences as above. Moreover, it is easy to see that all points in the form* $\overline{x} \doteq (\beta, 0)$ *with* $\beta > 0$ *are also AKKT (in fact, they are local minimizers) with*

$$x^k \doteq \left(\beta,\ \frac{1}{k}\right), \quad \mu^k \doteq \left(\frac{1}{\beta},\ 0\right), \quad \delta^k \doteq \left(-\frac{1}{\beta k},\ 0\right), \quad \Delta^k \doteq \left(-\frac{\beta}{k},\ 0\right).$$

Thus, all feasible points of this problem are AKKT.

We will discuss some ways of improving AKKT later on – that is, strengthening it so that it becomes less generic – but for now, let us focus on how they can be used. Although we could have started with such stronger sequential optimality conditions from the very beginning (which, by the way, is exactly how [9] is presented), we decided to start with AKKT because we believe its simplicity greatly helps understanding the essence of sequential conditions.

3.1.1 Global convergence of an Augmented Lagrangian method via AKKT

Perhaps one of the most recurrent applications of AKKT concerns the convergence theory of the Augmented Lagrangian method, which can be proved in a very simple way. The variant of Augmented Lagrangian we use is the *Powell-Hestenes-Rockafellar* method employing a *safeguarding* technique and some step control, which is the direct generalization of the one studied in [32], called ALGENCAN. In order to present it, let $\rho > 0$ be an arbitrary *penalty parameter* and let $\tilde{\mu} \in \mathcal{K}^\circ$ be a *safeguarded approximate multiplier* that is part of the safeguarding technique (it can also be chosen arbitrarily in practice, or even neglected). Then, define $L_{\rho,\tilde{\mu}} : \mathbb{R}^n \to \mathbb{R}$ as the *Augmented Lagrangian function* of (NCP), given by the following expression:

$$L_{\rho,\tilde{\mu}}(x) \doteq f(x) + \frac{\rho}{2} \left\| \Pi_{\mathcal{K}^\circ} \left(G(x) + \frac{\tilde{\mu}}{\rho} \right) \right\|^2 - \frac{1}{2} \left\| \tilde{\mu} \right\|^2.$$

Since it will be useful in the convergence proof, we compute the gradient of $L_{\rho,\tilde{\mu}}$ at a given point $x \in \mathbb{R}^n$ below:

$$\nabla L_{\rho,\tilde{\mu}}(x) = \nabla f(x) + DG(x)^* \left[\rho \Pi_{\mathcal{K}^\circ}\left(G(x) + \frac{\tilde{\mu}}{\rho}\right)\right]. \tag{3.2}$$

The algorithm is as follows:

Algorithm 1 Safeguarded Augmented Lagrangian method

Input: A sequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$ of positive scalars such that $\varepsilon_k \to 0$; a nonempty convex compact set $\mathcal{B} \subset \mathcal{K}^\circ$; real parameters $\tau > 1$, $\sigma \in (0,1)$, and $\rho_1 > 0$; and initial points $(x^0, \tilde{\mu}^1) \in \mathbb{R}^n \times \mathcal{B}$. Also, define $\|V^0\| = \infty$.

Initialize $k \leftarrow 1$. Then:

Step 1: Compute x^k such that $\|\nabla L_{\rho_k,\tilde{\mu}^k}(x^k)\| \leq \varepsilon_k$;

Step 2: Calculate

$$V^k \doteq \frac{\tilde{\mu}^k}{\rho_k} - \Pi_{\mathcal{K}^\circ}\left(G(x^k) + \frac{\tilde{\mu}^k}{\rho_k}\right); \tag{3.3}$$

Then,

a. If $k = 1$ or $\|V^k\| \leq \tau \|V^{k-1}\|$, set $\rho_{k+1} \doteq \rho_k$;

b. Otherwise, take ρ_{k+1} such that $\rho_{k+1} \geq \gamma \rho_k$.

Step 3: Choose any $\tilde{\mu}^{k+1} \in \mathcal{B}$, set $k \leftarrow k+1$ and go to Step 1.

When $\tilde{\mu}^k$ is set as zero for every $k \in \mathbb{N}$, Algorithm 1 reduces to the standard external penalty method.

It follows from Moreau's decomposition (Theorem 2.1.14) that $\tilde{\mu}^k = \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \tilde{\mu}^k)$ if, and only if, $\tilde{\mu}^k \in \mathcal{K}^\circ$, $G(x^k) \in \mathcal{K}$, and $\langle \tilde{\mu}^k, G(x^k) \rangle = 0$, which means that $V^k = 0$ if, and only if, the pair $(x^k, \tilde{\mu}^k)$ is primal-dual feasible and complementary. Moreover, note that Algorithm 1 does not necessarily keep record of the approximate multiplier sequence associated with $\{x^k\}_{k\in\mathbb{N}}$, which is given by:

$$\mu^k \doteq \rho_k \Pi_{\mathcal{K}^\circ}\left(G(x^k) + \frac{\tilde{\mu}^k}{\rho_k}\right). \tag{3.4}$$

These are usually computed, however, in several practical implementations of it, where $\tilde{\mu}^{k+1}$ is chosen as the projection of μ^k onto $\mathcal{B}$. Also, with these multipliers at hand, it is very easy to prove that any feasible limit point $\overline{x}$ of $\{x^k\}_{k\in\mathbb{N}}$ must satisfy AKKT.

Theorem 3.1.2. *Fix any choice of parameters in Algorithm 1 and let $\{x^k\}_{k\in\mathbb{N}}$ be the output sequence generated by it. If $\{x^k\}_{k\in\mathbb{N}}$ has a convergent subsequence $\{x^k\}_{k\in I} \to \overline{x}$, then:*

1. *The point $\overline{x}$ is stationary for the problem of minimizing $\frac{1}{2}\|\Pi_{\mathcal{K}}(-G(x))\|^2$;*
2. *If $\overline{x}$ is feasible, then $\overline{x}$ satisfies AKKT.*

Proof. Let $\{\varepsilon_k\}_{k\in\mathbb{N}} \to 0$, $\{\tilde{\mu}^k\}_{k\in\mathbb{N}} \subset \mathcal{B} \subset \mathcal{K}^\circ$, $\tau > 1$, $\sigma \in (0,1)$, and $\rho_1 > 0$ be input parameters for Algorithm 1, and let $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{V^k\}_{k\in\mathbb{N}}$ be as in Step 2. The proof of item 1 follows the same lines as its NLP version; see, for instance, [9, Proposition 1]. For item 2, let $\overline{x} \in \Omega$ be a limit point of the output sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by Algorithm 1 under the above settings and let $I \subseteq_\infty \mathbb{N}$ be such that $\{x^k\}_{k\in I} \to \overline{x}$.

Define $\{\mu^k\}_{k\in\mathbb{N}}$ as in (3.4) and take $\delta^k \doteq -\nabla_x L(x^k, \mu^k)$ and $\Delta^k \doteq V^k$ for all $k \in I$, where V^k is as given in (3.3). It follows from Step 1 that $\nabla_x L(x^k, \mu^k) = \nabla L_{\rho_k, \tilde{\mu}^k}(x^k) \to 0$, so $\lim_{k\in I} \nabla_x L(x^k, \mu^k) = 0$ also, and $\delta^k \to 0$. We also have

$$G(x^k) + \Delta^k = \Pi_{\mathcal{K}}\left(G(x^k) + \frac{\tilde{\mu}^k}{\rho_k}\right)$$

which yields $\langle \mu^k, G(x^k) + \Delta^k \rangle = 0$ for every $k \in I$, and this implies that $\mu^k \in N_{\mathcal{K}}(G(x^k) + \Delta^k)$ for all such k. If $\rho_k \to \infty$, then $V^k \to \Pi_{\mathcal{K}^\circ}(G(\overline{x}))$ which is in turn equal to zero since $\overline{x}$ is assumed to be feasible; on the other hand, if ρ_k remains bounded, then $V^k \to 0$ as well due to Step 2-a. Therefore, $\Delta^k \to 0$ and $\overline{x}$ satisfies AKKT. □

Loosely speaking, item 1 of Theorem 3.1.2 states that Algorithm 1 has a "tendency" to find feasible points, because should they not be feasible they are at least stationary points of a feasibility problem, and whenever it succeeds at finding a feasible point, its AKKT-type optimality is certified by the sequences produced during the execution of the method. This type of convergence theory is not limited to the Augmented Lagrangian, as we shall see next.

3.1.2 Global convergence of a Sequential Quadratic Programming method via AKKT

Here we build the global convergence theory of a direct generalization of the *Sequential Quadratic Programming* (SQP) method proposed by Correa and Ramírez's [43] for NSDP, and also of the general paradigm studied by Qi and Wei [80] for NLP.

Algorithm 2 General SQP method

Input: A real parameter $\tau > 1$, a pair of initial points $(x^1, \mu^1) \in \mathbb{R}^n \times \mathcal{K}$, and a positive definite approximation of $\nabla_x^2 L(x^1, \mu^1)$ denoted by H^1.

Initialize $k \leftarrow 1$. Then:

Step 1: Compute a solution d^k of the problem

$$\begin{aligned} \underset{d\in\mathbb{R}^n}{\text{Minimize}} \quad & \frac{1}{2} d^\top H^k d + \nabla f(x^k)^\top d, \\ \text{subject to} \quad & G(x^k) + DG(x^k)[d] \in \mathcal{K}, \end{aligned} \qquad \text{(Lin-QP}(k))$$

together with its Lagrange multiplier μ^{k+1}. If $d^k = 0$, ***stop*** and ***return*** x^k;

Step 2: Perform line search to find a steplength $\alpha^k \in (0, 1)$ satisfying *Armijo's rule*:

$$f(x^k + \alpha^k d^k) - f(x^k) \leq \tau \alpha^k \nabla f(x^k)^\top d^k. \qquad (3.5)$$

Step 3: Set $x^{k+1} \leftarrow x^k + \alpha^k d^k$, compute a positive definite approximation H^{k+1} of $\nabla_x^2 L(x^{k+1}, \mu^{k+1})$, set $k \leftarrow k + 1$, and go to Step 1.

The idea of the algorithm is very simple: it approximates the constraint function up to first-order and the objective function up to second-order; this induces a quadratic approximation of (NCP), which is successively solved in order to find directions of decrease until a stopping criterion is satisfied.

It is very easy to unify and extend the global convergence results for the SQP method obtained in [43, 80] using AKKT. Let's see it:

Proposition 3.1.3. *Assume that Step 1 of Algorithm 2 is always well defined. If there is an infinite subset* $I \subseteq_\infty \mathbb{N}$ *such that* $\lim_{k\in I} d^k = 0$ *and* $\{\|H^k\|\}_{k\in I}$ *is bounded, then any limit point* $\overline{x}$ *of* $\{x^k\}_{k\in I}$ *satisfies AKKT.*

Proof. By the KKT conditions for (Lin-QP(k)), for each $k \in I$ there exists some $\mu^k \in \mathcal{K}^\circ$ such that

$$\nabla f(x^k) + H^k d^k + DG(x^k)^*[\mu^k] = 0$$
$$\langle G(x^k) + DG(x^k)[d^k], \mu^k\rangle = 0.$$

Set $\Delta^k \doteq DG(x^k)[d^k]$ and $\delta^k \doteq H^k d^k$ for every $k \in I$ and since $\lim_{k\in I} d^k = 0$, we obtain that $\lim_{k\in I} \delta^k = 0$ and $\lim_{k\in I} \Delta^k = 0$ as well. Moreover, it follows from the feasibility of d^k that $G(x^k) + \Delta^k \in \mathcal{K}$ for every $k \in I$. Thus, $\overline{x}$ satisfies AKKT. □

The hypothesis on the convergence of a subsequence of $\{d^k\}_{k\in\mathbb{N}}$ to zero, directly or indirectly, and the well-definiteness of Step 1 are somewhat common regarding some types of SQP methods, as well as the boundedness of H^k – see, for instance, [18, 43, 80].

3.1.3 How to measure the strength of AKKT

Until this point we saw that:

- The quality of being an AKKT point is, on its own, an optimality condition;
- The Augmented Lagrangian method converges to AKKT points;
- The SQP method converges to AKKT points;

but what does this mean? The classical convergence theory for these methods is usually built in the form: "*if* $\overline{x}$ *is a feasible limit point of a sequence generated by Algorithms 1 and 2, and* $\overline{x}$ *satisfies a given Constraint Qualification, then* $\overline{x}$ *satisfies the KKT conditions*" [56]. This is a reasonable way to show some sort of reliability of the method, by describing a generic situation where it converges to stationary points. But it is not immediately clear how this is related to the theory presented in the previous two sections. As a starting point to clarify this relation, we prove that under Robinson's CQ (Definition 2.3.3, which is the standard choice for this purpose) AKKT and KKT are equivalent.

Theorem 3.1.4 (Proposition 6 of [9]). *Let* $\overline{x} \in \Omega$ *satisfy Robinson's CQ. Then,* $\overline{x}$ *satisfies AKKT if, and only if, it satisfies KKT.*

Proof. Suppose that $\overline{x}$ is AKKT certified by the sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, $\{\mu^k\}_{k\in\mathbb{N}} \subseteq \mathcal{K}^\circ$, $\{\delta^k\}_{k\in\mathbb{N}} \to 0$, and $\{\Delta^k\}_{k\in\mathbb{N}} \to 0$. To prove that $\overline{x}$ satisfies KKT, it suffices to show that $\{\mu^k\}_{k\in\mathbb{N}}$ has a bounded subsequence[3], which would guarantee that it has at least one limit point.

[3]Note that an analogous reasoning to this proof may be used to show that actually the whole sequence $\{\mu^k\}_{k\in\mathbb{N}}$ must be bounded under Robinson's CQ.

Then, it is straightforward to see that such a limit point must be a Lagrange multiplier for $\overline{x}$. Let us assume for a moment that $\{\mu^k\}_{k\in\mathbb{N}}$ has no bounded subsequence, that is, $\|\mu^k\| \to \infty$. This implies that the sequence $\{\mu^k/\|\mu^k\|\}_{k\in\mathbb{N}}$ is bounded, so it must have a nonzero limit point, say $\widehat{\mu} \in \mathcal{K}^\circ$, indexed by a subset $I \subseteq_\infty \mathbb{N}$. Then, observe that

$$\lim_{k\in I} \frac{\nabla_x L(x^k,\mu^k)}{\|\mu^k\|} = \lim_{k\in I} DG(x^k)^* \left[\frac{\mu^k}{\|\mu^k\|}\right] = DG(\overline{x})^*[\widehat{\mu}].$$

and that

$$\langle G(\overline{x}), \widehat{\mu}\rangle = \lim_{k\in I} \left\langle G(x^k) + \Delta^k,\ \frac{\mu^k}{\|\mu^k\|}\right\rangle = 0$$

so we conclude that $0 \neq \widehat{\mu} \in \text{Ker}(DG(\overline{x})^*) \cap N_{\mathcal{K}}(G(\overline{x}))$, which explicitly contradicts Robinson's CQ as in (2.26). Conversely, if $\overline{x}$ satisfies KKT with a Lagrange multiplier $\overline{\mu}$, simply define $x^k \doteq \overline{x}$, $\mu^k \doteq \overline{\mu}$, $\delta^k = 0$, and $\Delta^k = 0$, for every $k \in \mathbb{N}$ to see that $\overline{x}$ satisfies AKKT. □

Therefore, we see that the type of convergence theory of Sections 3.1.1 and 3.1.2 contains the classical one. But Robinson's CQ may not be the weakest Constraint Qualification that makes the conclusion of Theorem 3.1.4 true [20], which means that the AKKT-based convergence theory is actually sharper than the classical one. To phrase this conclusion properly, we are led to define the weakest Constraint Qualification that makes AKKT and KKT equivalent, which will be called the *AKKT-Regularity* condition. In NLP, it is also called the *cone continuity property* (CCP) [20].

Definition 3.1.2 (Definition 5 of [9]). Let $\overline{x} \in \Omega$. We say that $\overline{x}$ satisfies the *AKKT-Regularity* condition when the set-valued mapping

$$(x,\Delta) \mapsto \mathcal{H}(x,\Delta) \doteq DG(x)^*[N_{\mathcal{K}}(G(x)+\Delta)]$$

is upper semicontinuous at $(\overline{x},0)$; that is, when $\limsup_{(x,\Delta)\to(\overline{x},0)} \mathcal{H}(x,\Delta) = H(\overline{x})$ – see (2.24).

Recall that Guignard's CQ for linear Conic Programming problems consists simply of the closedness of $H(\overline{x})$. AKKT-Regularity, on the other hand, roughly consists on the "continuity of a closed-valued perturbation of H" at $\overline{x}$, which implies that $H(\overline{x})$ is itself closed since upper limits of set-valued mappings are always closed sets (this follows directly from the definition). What we want the reader to visualize, at least intuitively, is that KKT is related to Guignard's CQ just as AKKT is related to AKKT-Regularity. Now, to prove it once and for all, we proceed to the last theorem of this section.

Theorem 3.1.5 (Extracted from Theorem 5 of [9]). *A point $\overline{x} \in \Omega$ satisfies AKKT-Regularity if, and only if, for every objective function AKKT is equivalent to KKT.*

Proof. As we could see in the proof of Theorem 3.1.4, KKT always implies AKKT, so we are only concerned about the reverse implication. Suppose that $\overline{x}$ satisfies AKKT-Regularity and let f be any objective function such that $\overline{x}$ is an AKKT point certified by the sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, $\{\mu^k\}_{k\in\mathbb{N}} \subseteq \mathcal{K}^\circ$, $\{\delta^k\}_{k\in\mathbb{N}} \to 0$, and $\{\Delta^k\}_{k\in\mathbb{N}} \to 0$. Observe that $\nabla_x L(x^k,\mu^k) + \delta^k = 0$ implies that

$$-\nabla f(x^k) - \delta^k \in \mathcal{H}(x^k,\Delta^k)$$

for every $k \in \mathbb{N}$, so

$$-\nabla f(\overline{x}) = \lim_{k\to\infty} -\nabla f(x^k) - \delta^k \in \limsup_{(x,\Delta)\to(\overline{x},0)} \mathcal{H}(x,\Delta) = H(\overline{x})$$

due to AKKT-Regularity. Thus, the KKT conditions hold at $\overline{x}$. Conversely, suppose that AKKT and KKT are equivalent for every objective function f, and observe that the inclusion

$$H(\overline{x}) \subseteq \limsup_{(x,\Delta)\to(\overline{x},0)} \mathcal{H}(x,\Delta)$$

always holds and that $H(\overline{x}) \neq \emptyset$ because $G(\overline{x}) \in \mathcal{K}$, so $0 \in H(\overline{x})$. To prove the other inclusion, let $\overline{w} \in \limsup_{(x,\Delta)\to(\overline{x},0)} \mathcal{H}(x,\Delta)$ and there exist sequences $\{w^k\}_{k\in\mathbb{N}} \to \overline{w}$, $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, and $\{\Delta^k\}_{k\in\mathbb{N}} \to 0$, such that $w^k \in \mathcal{H}(x^k, \Delta^k)$ for every $k \in \mathbb{N}$. Since there is some $\mu^k \in N_{\mathcal{K}}(G(x^k) + \Delta^k)$ such that $w^k = DG(x^k)^*[\mu^k]$ for each $k \in \mathbb{N}$, define $f(x) \doteq -\langle \overline{w}, x\rangle$ and observe that $\overline{x}$ is an AKKT point with respect to f, where $\delta^k \doteq 0$ for every $k \in \mathbb{N}$. Therefore, $\overline{x}$ also satisfies the KKT conditions with respect to f, meaning that $\overline{w} \in H(\overline{x})$. Since $\overline{w}$ was chosen arbitrarily, we conclude that

$$H(\overline{x}) \supseteq \limsup_{(x,\Delta)\to(\overline{x},0)} \mathcal{H}(x,\Delta).$$

□

We should remark that Definition 3.1.2 is slightly different from the AKKT-regularity written in [9, Definition 5], but Theorem 3.1.5 shows that they are equivalent.

At this point, an immediate question arises: what happens when we consider other types of perturbations of (Lin-P) or even (Lin-D)? For instance, what can be obtained by introducing a perturbation to the normal cone in (Lin-D) instead of a perturbation to $G(\overline{x})$ in (Lin-P)? Does this result in AKKT again or is it something else? This is the main topic of next couple of sections.

3.2 Approximate Gradient Projection

Consider the following relaxation of the normal cone to $\mathcal{K}$ at a given $z \in \mathcal{K}$, with respect to some $\varepsilon > 0$:

$$N^{\varepsilon}_{\mathcal{K}}(z) \doteq \{w \in \mathcal{K}^{\circ} \colon |\langle z, w\rangle| \leqslant \varepsilon\} \tag{3.6}$$

and observe that $N^0_{\mathcal{K}}(z) = N_{\mathcal{K}}(z)$. Such an approximation can be used to define a sequential optimality condition in the same lines as it was done with AKKT, leading to what was called in [9] the *Approximate Gradient Projection* (AGP) condition:

Definition 3.2.1 (Definition 2 of [9]). Let $\overline{x} \in \Omega$. We say that $\overline{x}$ is an *AGP point* (or that $\overline{x}$ satisfies the *AGP condition*) if there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and $\{\mu^k\}_{k\in\mathbb{N}} \subseteq \mathbb{Y}$, as well as some perturbation sequences $\{\delta^k\}_{k\in\mathbb{N}} \to 0$ and $\{\varepsilon^k\}_{k\in\mathbb{N}} \to 0$ such that, for every $k \in \mathbb{N}$:

- $\nabla_x L(x^k, \mu^k) + \delta^k = 0$;
- $\mu^k \in N^{\varepsilon^k}_{\mathcal{K}}(\Pi_{\mathcal{K}}(G(x^k)))$.

Equivalently, it is straightforward to see that AGP holds at $\overline{x}$ if, and only if, there are sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and $\{\mu^k\}_{k\in\mathbb{N}} \subseteq \mathcal{K}^{\circ}$ such that

$$\nabla_x L(x^k, \mu^k) \to 0 \quad \text{and} \quad \langle \mu^k, \Pi_{\mathcal{K}}(G(x^k))\rangle \to 0,$$

which is how it was presented in [9]. The reason why this condition is called "Approximate Gradient Projection" is that it happens to generalize the sequential condition of Martínez and Svaiter [67] from NLP, called AGP – see [9, Theorem 1] and also [25, Theorem 2.7].

Every AGP point must also be an AKKT point with the same associate primal sequence and a translated dual sequence, as stated next:

Proposition 3.2.1 (Proposition 2 of [9]). *Let $\overline{x} \in \Omega$ be an AGP point certified by the sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, $\{\mu^k\}_{k\in\mathbb{N}} \subseteq \mathbb{Y}$ and the perturbations $\{\delta^k\}_{k\in\mathbb{N}} \to 0$ and $\{\varepsilon^k\}_{k\in\mathbb{N}} \to 0$. Then, $\overline{x}$ is also an AKKT point certified by the same $\{x^k\}_{k\in\mathbb{N}}$ together with $\{\widehat{\mu}^k \doteq \mu^k + \Pi_{\mathcal{K}}(G(x^k))\}_{k\in\mathbb{N}}$ and the perturbations $\{\delta^k\}_{k\in\mathbb{N}}$ and $\{\Delta^k \doteq -\Pi_{\mathcal{K}^\circ}(G(x^k))\}_{k\in\mathbb{N}} \to 0$.*

Even so, every local minimizer of (NCP) still satisfies AGP but, quite surprisingly, AGP is different from AKKT. Indeed, the implication of Proposition 3.2.1 is strict, and to see this we invite the reader back to Example 3.1.1 where all feasible points were AKKT points. Recall that Example 3.1.1 consists of minimizing $-x_2$ subject to $G(x_1, x_2) \doteq (x_1x_2,\ x_1) \in \{0\} \times \mathbb{R}_+ \subseteq \mathbb{R}^2$. Take $\overline{x} \doteq (0, \alpha)$ for any $\alpha > 0$. If there were sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and $\{\mu^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}\times\mathbb{R}_-$ certifying that $\overline{x}$ is AGP, then we would have:

- $\mu_1^k x_2^k + \mu_2^k \to 0$;
- $-1 + \mu_1^k x_1^k \to 0$;
- $\mu_2^k \max\{0, x_1^k\} \to 0$.

The first and last statements above, together with $(x_1^k, x_2^k) \to (0, \alpha)$ and $\alpha > 0$, imply that

$$0 = \lim_{k\to\infty} \mu_1^k x_2^k \max\{0, x_1^k\} + \mu_2^k \max\{0, x_1^k\} = \lim_{k\to\infty} \mu_1^k \max\{0, x_1^k\},$$

which further implies that

$$1 = \lim_{k\to\infty} \mu_1^k x_1^k = \lim_{k\to\infty} \mu_1^k \min\{0, x_1^k\} + \mu_1^k \max\{0, x_1^k\} = \lim_{k\to\infty} \mu_1^k \min\{0, x_1^k\}.$$

Then, for every k large enough, we must have $x_1^k < 0$ and $\mu_1^k < 0$, but since $\lim_{k\to\infty} \mu_1^k x_2^k + \mu_2^k \to 0$ and since $x_2^k \to \alpha > 0$ we must also have $\mu_2^k > 0$ for every large k, which contradicts $\mu_2^k \in \mathbb{R}_-$. Thus, $\overline{x}$ is not an AGP point for any $\alpha > 0$.

Example 3.1.1 tells us that there may be infinitely many points that fulfil AKKT and do not fulfil AGP, meaning that AGP is a much sharper necessary optimality condition than AKKT. Even so, the Augmented Lagrangian method is still guaranteed to converge to AGP points, as we state next:

Theorem 3.2.2 (Theorem 3 of [9]). *Let $\{x^k\}_{k\in\mathbb{N}}$ be any output sequence generated by Algorithm 1. Every feasible limit point of $\{x^k\}_{k\in\mathbb{N}}$ satisfies AGP.*

This fortunate surprise was new even for NLP problems at the time [9] was submitted, and it reveals that the Augmented Lagrangian method is actually more reliable than the literature at the time knew it was. That said, it should be remarked that the proof of Theorem 3.2.2 is not as straightforward as the proof of Theorem 3.1.2. Furthermore, a sequential-type Constraint Qualification in the same style as Definition 3.1.2 using, of course, the approximation of $H(\overline{x})$ given by the mapping

$$(x, \varepsilon) \mapsto \mathcal{H}_{\mathsf{AGP}}(x, \varepsilon) \doteq DG(x)^*[N_{\mathcal{K}}^{\varepsilon}(\Pi_{\mathcal{K}}(G(x)))].$$

When $\mathcal{H}(x, \varepsilon)$ is upper semicontinuous at $(\overline{x}, 0)$, we say that *AGP-Regularity* holds at $\overline{x}$ – see [9, Definition 2] also. Not surprisingly, AGP-Regularity is the weakest Constraint Qualification that makes AGP and KKT equivalent notions [9, Theorem 5]. To perceive its relation with the other Constraint Qualifications, notice that since AGP is strictly stronger than AKKT, then AGP-Regularity is strictly weaker than AGP-Regularity, which is in turn strictly weaker than Robinson's CQ.

3.3 Complementary AKKT

In general, although AGP is stronger than AKKT, any convergence theory build in terms of AGP can still be further improved at the cost of some generality loss. The tool we present for doing so is motivated by a second look at Moreau's decomposition Theorem 2.1.11. For any convergent sequence of iterates of a given algorithm, say $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, recall that

$$G(x^k) = \Pi_{\mathcal{K}}(G(x^k)) + \Pi_{\mathcal{K}^\circ}(G(x^k))$$

for every $k \in \mathbb{N}$, and that AGP only describes the convergence of its associate approximate Lagrange multipliers, say $\{\mu^k\}_{k\in\mathbb{N}}$, with respect to the first term of the decomposition. Imposing a more strict control by means of the condition

$$\langle \mu^k, \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle \to 0 \tag{3.7}$$

yields the so-called *Complementary AKKT* (CAKKT) sequential optimality condition. Let us properly define it:

Definition 3.3.1 (Definition 3 of [9]). Let $\overline{x} \in \Omega$. We say that $\overline{x}$ is a *CAKKT point* (or that $\overline{x}$ satisfies the *CAKKT condition*) if there are sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and $\{\mu^k\}_{k\in\mathbb{N}} \subseteq \mathbb{Y}$ certifying that $\overline{x}$ is AGP and, in addition, (3.7) holds.

This condition is strictly stronger than AGP, which is no surprise by construction, and since it generalizes a condition with the same name from NLP [24] – see also [9, Proposition 3] – we point the reader to [24] for a counterexample. Alternatively, [9, Example 2] serves for the same purpose. It is still true that every local minimizer satisfies CAKKT, which is a direct consequence of Theorem 3.1.1. However, it is not true that the Augmented Lagrangian method (Algorithm 1) always converges to CAKKT points; see the counterexample after [24, Theorem 5.1] for details. To solve this issue, we introduce a minor extension of a notion that is known in NLP as the *generalized Łojasiewicz property*.

3.3.1 The generalized Łojasiewicz property

Let $\Psi\colon \mathbb{R}^n \to \mathbb{R}$ be any differentiable function. The *generalized Łojasiewicz property* is defined at any point $\overline{x} \in \mathbb{R}^n$ by the existence of a neighborhood $\mathcal{V}$ of $\overline{x}$ plus a continuous function $\psi(x) : B(\overline{x}, \delta) \to \mathbb{R}$ such that $\lim_{x\to\overline{x}} \psi(x) = 0$ and

$$\forall x \in \mathcal{V}, \quad |\Psi(x) - \Psi(\overline{x})| \leqslant \psi(x)\|D\Psi(x)\|. \tag{3.8}$$

The reader may be familiar with this type of property, since it has been extensively used in the study of optimization methods due to Polyak [60]. Because all analytic functions satisfy (3.8), it is also known as the *Łojasiewicz inequality* [34] for analytic functions. Among its many applications, there is a result about the convergence of Algorithm 1 to CAKKT points.

Theorem 3.3.1. *Let $\overline{x}$ be a feasible limit point of a sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by Algorithm 1. If the function $\Psi(x) \doteq \|\Pi_{\mathcal{K}^\circ}(G(x))\|^2$ satisfies the generalized Łojasiewicz property at $\overline{x}$, then $\overline{x}$ satisfies CAKKT.*

We will not exhibit the proof of Theorem 3.3.1 because it is mostly technical – we refer to [9, Theorem 4] instead. However, it may be worth mentioning at least what is the role of the generalized Łojasiewicz inequality in it, which is fairly simple: due to Theorem 3.2.2 it all amounts to showing that $\langle \mu^k, \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle \to 0$ and one way to achieve this is to prove first that, if $\rho_k \to \infty$, feasibility is gained faster than ρ_k grows. That is, to show that $\rho_k \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \to 0$. But observe that the existence of a function ψ such that

$$\|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 = |\Psi(x^k)| \leqslant \psi(x^k)\|D\Psi(x^k)\| = 2\psi(x^k)\|DG(x^k)^*[\Pi_{\mathcal{K}^\circ}(G(x^k))]\|$$

for every large k leads us to the following:

$$\begin{aligned}\rho_k \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 &\leqslant 2\psi(x^k)\|\rho_k DG(x^k)^* \Pi_{\mathcal{K}^\circ}(G(x^k))\| \\ &\leqslant 2\psi(x^k)[\|DG(x^k)^*\|\|\tilde{\mu}^k\| + \|DG(x^k)^*[\mu^k]\|]\end{aligned}$$

and then sought relation follows from the continuity of DG together with the fact $\tilde{\mu}^k$ is bounded, $DG(x^k)^*[\mu^k] \to -\nabla f(\overline{x})$, and $\psi(x^k) \to 0$.

Thus, every limit point $\overline{x}$ of Algorithm 1 satisfies AGP without demanding any Constraint Qualification but if, in addition, the generalized Łojasiewicz inequality holds for the feasibility measure $\|\Pi_{\mathcal{K}^\circ}(G(x))\|^2$ at $\overline{x}$, there is an additional control over the approximate Lagrange multiplier sequence computed by the method. We should mention that the weakest Constraint Qualification that ensures CAKKT and KKT are equivalent can also be computed in a similar fashion as AGP- and AKKT-Regularity. Details will be omitted, since this is the main focus of our paper [9] and all computations that lead to *CAKKT-Regularity* are presented explicitly therein. It is also worth remarking that this enhanced convergence result to CAKKT under the Łojasiewicz property was already known in NLP, but it is new to Conic Programming, including even conic problems that have received some study on sequential optimality conditions prior to our work, such as NSDP and NSOCP.

3.3.2 Some other methods that converge to feasible AKKT, AGP, and CAKKT points

A natural question that the reader may have at this point is: *what's the point of defining increasingly stronger sequential optimality conditions given that our focus is always on Augmented Lagrangian methods? Wouldn't it be better to study the behavior of the output sequences of Algorithm 1 directly instead?* Although a bit bold, this question is very pertinent and it is one of the main motivations for the content of the next section. The answer, though, is that sequential optimality conditions are not designed with a specific algorithm in mind, as we could see in the previous sections. Instead, they are meant to unify and simplify the convergence theory of several different methods at once. We used the Augmented Lagrangian as an example to illustrate how sequential optimality conditions are meant to be used – and our results prove that they are indeed effective and very powerful tools. However, we acknowledge that claiming that our analyses can be carried over to other methods is too vague without showing examples. This is why we have prepared a list of some numerical methods (other than Algorithm 1) from the literature that were proven to converge to AKKT, AGP, or CAKKT points.

- The **SQP methods** for NLP of Qi and Wei [80] and Andreani et al. [18], which converge to AKKT points; the stabilized SQP method of Gill, Kungurtsev, and Robinson [49] also for NLP that converges to CAKKT points; and the SQP method of Yamakawa and Okuno [88] for NSDP that converges to AKKT points as well;

- Some **Interior-Point methods** such as the classical NLP method proposed by Chen and Goldfarb [41] that converges to AKKT points (see also [18]); and the Primal-Dual Interior-Point method for NSDP of Yamashita, Yabe, Harada [89] that converges to CAKKT points (see also [8]);

- A mixed **Shifted Primal-Dual Penalty-Barrier method** for NLP by Gill, Kungurtsev, and Robinson [50] that had its convergence theory built in terms of CAKKT;
- General **Inexact Restoration methods** such as the algorithm of Birgin, Bueno, and Martínez [33] for NLP, that is traditionally proven to converge to AGP points.

However, because our work is still recent, we cannot provide any example in a conic context that is more general than NSOCP, NSDP, and NLP. That said, as a final remark based on the degree of simplicity of sequential conditions, we strongly believe that, should any of the methods we just mentioned be extended to a more general conic context, it is likely to admit a simplified convergence theory in terms of our extensions of AKKT, AGP, or CAKKT.

3.4 Complete representation of Augmented Lagrangian methods

Although there are many methods that admit a sequential optimality condition-based convergence theory, we have given much more emphasis on the Augmented Lagrangian method in the previous sections. Now, at a complete loss of generality we are led to study the strongest sequential optimality condition that may be used to support the convergence theory of Algorithm 1, specifically. Defining such a condition is more or less straightforward:

Definition 3.4.1 (Extension of Definition 1 of [16])**.** Let $\overline{x} \in \Omega$. We say that $\overline{x}$ is an *AL-AKKT point* (or that $\overline{x}$ satisfies the *AL-AKKT condition*) if there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, $\{\rho_k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}_{++}$, and a bounded sequence $\{\tilde{\mu}^k\}_{k\in\mathbb{N}} \subseteq \mathcal{K}^\circ$ such that:

- $\nabla_x L(x^k, \mu^k) \to 0$;
- $V^k \to 0$;

where V^k is defined as in (3.3) and μ^k is defined as in (3.4).

Observe that $\{\rho_k\}_{k\in\mathbb{N}}$ does not necessarily diverge. The main property of AL-AKKT, which is expected but not trivial whatsoever, is that Algorithm 1 and its variants are the only methods that converges to AL-AKKT points. In fact, this is the main result of our paper [16], which we state next:

Theorem 3.4.1 (Theorem 1 of [16])**.** *Every feasible limit point of every output sequence generated by Algorithm 1 satisfies AL-AKKT. Conversely, for every AL-AKKT point $\overline{x}$ there exist some input parameters for Algorithm 1 such that its execution produces an output sequence that has $\overline{x}$ among its limit points.*

From this point onwards, we shall express a scenario where one chooses a set of input parameters for Algorithm 1 and runs it afterwards as an *instance of the method*. Then, the second part of Theorem 3.4.1 states that every AL-AKKT point must be a possible output of some *instance* of Algorithm 1. Although Theorem 3.4.1 was originally proved in [16] only for NLP, the same proof presented in [16] can be used to prove the conic version of Theorem 3.4.1 without any actual modification (besides notation, of course). Hence, it will be omitted.

The main reason why Theorem 3.4.1 is interesting is that it allows us to study some convergence properties of Algorithm 1 in a simplified language. A somewhat surprising fact that arises when we do this kind of analysis is that merely being an accumulation point of the Augmented Lagrangian method is itself a necessary optimality condition. To prove this properly, it suffices to adapt the proof of Theorem 3.1.1:

Theorem 3.4.2 (Extension of Theorem 2 of [16]). *Let $\overline{x} \in \Omega$ be a local minimizer of (NCP) and make any choice of parameters $\{\rho_k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}_{++}$ and $\{\tilde{\mu}^k\}_{k\in\mathbb{N}} \subseteq \mathcal{B} \subseteq \mathcal{K}^\circ$ such that $V^k \to 0$, where $\mathcal{B}$ is a compact set. 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:*

$$\mathcal{F}_k(x) \doteq f(x) + \frac{\rho_k}{2}\left\|\Pi_{\mathcal{K}^\circ}\left(G(x) + \frac{\tilde{\mu}^k}{\rho_k}\right)\right\|^2 + \frac{1}{2}\|x - \overline{x}\|^2.$$

Proof. If $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$, then we can follow the same steps of the proof of Theorem 3.1.1 with $\mathcal{F}_k$ instead of F_k to reach the conclusion of this theorem. Thus, the only case we need to elaborate is when $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded. In this case, we will assume it converges to some $\overline{\rho} > 0$ and, for simplicity, we will also assume that $\{\tilde{\mu}\}_{k\in\mathbb{N}}$ converges to some $\tilde{\mu}$ (otherwise, a subsequence must be taken). Anyway, this proof resembles Theorem 3.1.1 in some aspects. For example, we again let $\delta > 0$ be the radius of optimality of $\overline{x}$ and we study:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & \mathcal{F}_k(x) \\ \text{subject to} & \|x - \overline{x}\| \leqslant \delta. \end{array} \qquad \text{(LReg-P}(k))$$

Let $\{x^k\}_{k\in\mathbb{N}}$ be such that each x^k is a local minimizer of (LReg-P(k)) and let $\overline{w}$ be one of its accumulation points indexed by $I \subseteq_\infty \mathbb{N}$. Then, since $V^k \to 0$ it follows that $\tilde{\mu} = \Pi_{\mathcal{K}^\circ}(\overline{\rho}G(\overline{w}) + \tilde{\mu})$ and Moreau's decomposition (Theorem 2.1.14) tells us that

$$\tilde{\mu} \in \mathcal{K}^\circ, \; G(\overline{w}) \in \mathcal{K}, \text{ and } \langle \tilde{\mu}, G(\overline{w})\rangle = 0.$$

Now, for any $z \in \Omega$ such that $\|z - \overline{x}\| \leqslant \delta$ we have

$$f(x^k) + \frac{\rho_k}{2}\left\|\Pi_{\mathcal{K}^\circ}\left(G(x^k) + \frac{\tilde{\mu}^k}{\rho_k}\right)\right\|^2 + \frac{1}{2}\|x^k - \overline{x}\|^2 \leqslant f(z) + \frac{\rho_k}{2}\left\|\Pi_{\mathcal{K}^\circ}\left(G(z) + \frac{\tilde{\mu}^k}{\rho_k}\right)\right\|^2 + \frac{1}{2}\|z - \overline{x}\|^2 \tag{3.9}$$

and taking limits in I, the above equation (3.9) implies

$$\begin{aligned} 2f(\overline{w}) + \|\tilde{\mu}\|^2 + \|\overline{w} - \overline{x}\|^2 &\leqslant 2f(z) + \|\Pi_{\mathcal{K}^\circ}(\overline{\rho}G(z) + \tilde{\mu})\|^2 + \|z - \overline{x}\|^2 \\ &= 2f(z) + \|\Pi_{\mathcal{K}^\circ}(\overline{\rho}G(z) + \tilde{\mu}) - \Pi_{\mathcal{K}^\circ}(\overline{\rho}G(z))\|^2 + \|z - \overline{x}\|^2 \\ &\leqslant 2f(z) + \|\overline{\rho}G(z) + \tilde{\mu} - \overline{\rho}G(z)\|^2 + \|z - \overline{x}\|^2 \\ &= 2f(z) + \|\tilde{\mu}\|^2 + \|z - \overline{x}\|^2 \end{aligned}$$

because $\Pi_{\mathcal{K}^\circ}(\overline{\rho}G(z)) = 0$. Consequently, we obtain $f(\overline{w}) + \frac{1}{2}\|\overline{w} - \overline{x}\|^2 \leqslant f(z) + \frac{1}{2}\|z - \overline{x}\|^2$. Now we recall the strategy of the proof Theorem 3.1.1 and we observe that $\overline{w}$ is a global minimizer of (Loc-P), implying that $\overline{w} = \overline{x}$. Moreover, for all $k \in I$ sufficiently large the point x^k is also an unconstrained local minimizer of $\mathcal{F}_k$. □

Corollary 3.4.3. *Every local minimizer of (NCP) satisfies AL-AKKT, regardless of the choices of $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{\tilde{\mu}^k\}_{k\in\mathbb{N}}$.*

It is worth remarking that Theorem 3.4.2 and Corollary 3.4.3 compose a new extension of [16, Theorem 2], which originally concerned NLP problems, to the context of (NCP). Besides revealing that Algorithm 1 carries a global convergence proof within its definition, Theorem 3.4.2 and Corollary 3.4.3 can be used to study some implementation issues around Algorithm 1. However, we must mention that Algorithm 1 has not yet received any competitive implementation in nonlinear conic contexts other than NLP.

3.4.1 Some implementation choices and their impact over the convergence of the Augmented Lagrangian method

Quite often, the global convergence theory of a method is constructed over a simplified version of it, slightly different from the one that was actually implemented and tested in practice. The latter usually contains a few technical adaptations for better numerical performance. But as far as we know, a proper clarification of the effect of such "technical adaptations" over the convergence of the method is rarely addressed, if ever.

We collected some of these "details" and we used AL-AKKT to study some of their actual effects. These (mostly technical) findings can be found in detail in [16] – some of which were expected, some were not – and a brief summary is presented next:

1. **About the penalty parameter control:** In [16, Proposition 1] we prove[4] that a feasible point $\bar{x}$ is AL-AKKT if, and only if, there are sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$, $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$, and $\{\tilde{\mu}^k\}_{k\in\mathbb{N}} \subseteq \mathcal{K}^\circ$ such that $\nabla_x L(x^k, \mu^k) \to 0$, where μ^k is defined as in (3.4). To Algorithm 1 this means that removing Step 2, which consists of a growth control over $\{\rho_k\}_{k\in\mathbb{N}}$, does not affect the range of convergence of the method. On the other hand, this practice provides a computational gain of stability since it allows ρ_k to remain bounded, and our result states that this stability comes, in a sense, *for free*;

2. **About multifold problems:** Assuming that

$$\mathcal{K} \doteq \mathcal{K}_1 \times \ldots \times \mathcal{K}_q \quad \text{and} \quad G(x) \doteq (G_1(x), \ldots, G_q(x)),$$

Problem (NCP) is equivalent to the problem of minimizing $f(x)$ subject to $G_j(x) \in \mathcal{K}_j$ for every $j \in \{1, \ldots, q\}$; that is, a *multifold problem*. This opens the possibility for one to use different penalty parameters, one of each constraint, and construct a variant of Algorithm 1 based on the following penalized function:

$$L_{\rho_1,\ldots,\rho_q,\tilde{\mu}_1,\ldots,\tilde{\mu}_q}(x) \doteq f(x) + \sum_{j=1}^{q} \frac{\rho_j}{2} \left\| \Pi_{\mathcal{K}^\circ}\left(G_j(x) + \frac{\tilde{\mu}_j}{\rho_j} \right) \right\|^2 - \frac{1}{2\rho_j} \|\tilde{\mu}_j\|^2 .$$

However, although this seems a good practice for allowing one to distribute preferences among constraints based on how fast they are being fulfilled, this excludes some stationary points (and local minimizers) from the range of convergence of the original Algorithm 1 (see [16, Section 3.4]). Besides, some numerical experiments (regarding NLP) by Birgin et al. [3] also suggest that using different penalty parameters is worse than using the same parameter for all constraints;

3. **About the safeguarding technique:** In a recent paper by Kanzow and Steck [59] there is a simple example, with a single minimizer that was also a KKT point, showing that Algorithm 1 generates sequences whose feasible limits points are exactly the problem minimizer, whereas the method without safeguarding (i.e., without Step 3 and using $\tilde{\mu}^k \doteq \mu^{k-1}$ for every $k \in \mathbb{N}$) was proven to be unable to converge to it. This means that the safeguarded method is *more reliable* in a certain sense. We contribute to this discussion by showing [16, Theorem 5] that as long as $\|\tilde{\mu}^k\| = o(\rho_k)$ the safeguards can be removed without affecting the quality of convergence of Algorithm 1, and we also show that this requirement cannot be relaxed, for instance, to $\|\tilde{\mu}^k\| = O(\rho_k)$ without losing reliability;

[4] Again, this result was originally proven to NLP, but its extension to NCP is straightforward.

4. **About forcing stationarity or feasibility:** With a couple of examples [16, Examples 1 and 2] we show that forcing exact stationarity over the subproblems by means of setting $\varepsilon_k \doteq 0$ for every $k \in \mathbb{N}$ may rule out some local minimizers from the range of convergence of the method (and this is why $\{\varepsilon_k\}_{k\in\mathbb{N}}$ is expected in Algorithm 1 to be *a sequence of positive scalars*). Moreover, forcing feasibility by means of demanding $V^k = 0$ for every $k \in \mathbb{N}$ may cause the method to diverge.

Studying several variants of Algorithm 1 and comparing it with the original method is the simplest application of its characterization via AL-AKKT. A more ambitious idea is to apply the strategy we used in [16] to different methods; that is, characterizing their feasible limit points and output sequences in the sequential optimality condition style. Then, hopefully, we should also be able to compare different methods by means of comparing their associate sequential conditions.

3.4.2 On the best achievable quality of limit points of Augmented Lagrangian schemes

Perhaps the most meaningful feature of AL-AKKT is the possibility of characterizing the weakest Constraint Qualification that makes AL-AKKT equivalent to KKT, which will be called *AL-Regularity* to match the other "□-Regularity" conditions. In particular, it is the weakest Constraint Qualification that ensures convergence of Algorithm 1 to KKT points, and thus, it provides the tightest possible convergence theory for this method. In order to define it properly, consider the function:

$$(x, \rho) \mapsto \mathcal{H}_{\mathsf{AL}}(x, \rho) \doteq DG(x)^*[\rho \Pi_{\mathcal{K}^\circ}(G(x))].$$

Then, following the pattern of the previous sections, we say that *AL-Regularity* holds at a feasible point $\overline{x}$ when

$$H(\overline{x}) = \limsup_{x \to \overline{x},\ \rho \to \infty} \mathcal{H}_{\mathsf{AL}}(x, \rho). \tag{3.10}$$

But one inclusion is always true, as we shall see in a direct extension of [16, Theorem 4].

Proposition 3.4.4 (Direct extension of Theorem 4 of [16]). *Let $\overline{x} \in \Omega$ and let $\mathcal{H}_{\mathsf{AL}}$ be defined as above. Then,*

$$H(\overline{x}) \subseteq \limsup_{x \to \overline{x},\ \rho \to \infty} \mathcal{H}_{\mathsf{AL}}(x, \rho).$$

Proof. Let $y \in H(\overline{x})$, so there exists some $\mu_y \in N_{\mathcal{K}}(G(\overline{x}))$ such that $y = DG(\overline{x})^*[\mu_y]$. Borrowing an idea of [19, Lemma 4.3] we invoke [82, Theorem 6.11], which states that since $y \in H(\overline{x}) \subseteq T_\Omega(\overline{x})^\circ$ (that is, since y is a *regular normal* vector to Ω at $\overline{x}$) then there exists a continuously differentiable (in fact, smooth) function F such that $\overline{x}$ is a strict global maximizer of F over Ω and $\nabla F(\overline{x}) = y$. Applying the very same reasoning of the proof of Theorem 3.1.1 to the problem of minimizing $-F(x)$ over Ω is enough to obtain a sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ given any $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$, such that

$$-\nabla F(x^k) + DG(x^k)^*[\rho_k \Pi_{\mathcal{K}^\circ}(G(x^k))] \to 0$$

for every $k \in \mathbb{N}$. Then, observe that $y^k \doteq DG(x^k)^*[\rho_k \Pi_{\mathcal{K}^\circ}(G(x^k))] = \mathcal{H}_{\mathsf{AL}}(x^k, \rho_k)$ for every $k \in \mathbb{N}$ and that $\{y^k\}_{k\in\mathbb{N}} \to y$. Thus, $y \in \limsup_{x \to \overline{x},\ \rho \to \infty} \mathcal{H}_{\mathsf{AL}}(x, \rho)$. □

In particular, we obtain as a corollary of Proposition 3.4.4 that every KKT point is also AL-AKKT, meaning that all KKT points (even those who are not local minimizers) are in the range of convergence of Algorithm 1, and so are their associate Lagrange multipliers. To close this section, we also extend from [16] the proof for the fact AL-Regularity is the minimal regularity condition associated with AL-AKKT.

Theorem 3.4.5. *A point $\overline{x} \in \Omega$ satisfies AL-Regularity if, and only if, for every objective function AL-AKKT is equivalent to KKT.*

Proof. Let $\overline{x}$ be an AL-AKKT point associated with the sequences $\{x^k\}_{k\in\mathbb{N}}$, $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{\tilde{\mu}^k\}_{k\in\mathbb{N}}$. If $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded, then $\overline{x}$ must clearly satisfy the KKT conditions regardless of the fulfilment of AL-regularity due to Moreau's decomposition (Theorem 2.1.14) and the definition of V^k (3.3). Thus, from now on suppose that $\rho_k \to \infty$ and that $\overline{x}$ satisfies AL-Regularity. Observe that

$$DG(x^k)^* \left[\Pi_{\mathcal{K}^\circ}\left(\rho_k G(x^k) + \tilde{\mu}^k\right)\right] = \underbrace{DG(x^k)^* \left[\Pi_{\mathcal{K}^\circ}\left(\rho_k G(x^k)\right)\right]}_{=\mathcal{H}_{\mathsf{AL}}(x^k)} + DG(x^k)^* \left[\widehat{\mu}^k\right] \tag{3.11}$$

where

$$\widehat{\mu}^k \doteq \Pi_{\mathcal{K}^\circ}\left(\rho_k G(x^k) + \tilde{\mu}^k\right) - \Pi_{\mathcal{K}^\circ}\left(\rho_k G(x^k)\right).$$

Moreover, Proposition 2.1.12 (non-expansiveness of the projection) tells us that $\|\widehat{\mu}^k\| \leqslant \|\tilde{\mu}^k\|$ for every $k \in \mathbb{N}$, and since $\{\tilde{\mu}^k\}_{k\in\mathbb{N}}$ is bounded, so is $\{\widehat{\mu}^k\}_{k\in\mathbb{N}}$. We assume, for simplicity, that $\{\widehat{\mu}^k\}_{k\in\mathbb{N}}$ converges to some $\widehat{\mu} \in \mathcal{K}^\circ$; otherwise a subsequence would have to be taken. Furthermore, observe that the left-hand side of (3.11) converges to $-\nabla f(\overline{x})$. By AL-regularity at $\overline{x}$, we conclude that

$$-\nabla f(\overline{x}) \in \left[\limsup_{x\to\overline{x},\ \rho\to\infty} \mathcal{H}_{\mathsf{AL}}(x,\rho)\right] + H(\overline{x}) \subset H(\overline{x}) + H(\overline{x}) = H(\overline{x}),$$

that is, $\overline{x}$ is a KKT point.

Conversely, let $\overline{w} \in \limsup_{x\to\overline{x},\ \rho\to\infty} \mathcal{H}_{\mathsf{AL}}(x,\rho)$. Then, there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$ such that

$$DG(x^k)^* \left[\Pi_{\mathcal{K}^\circ}\left(\rho_k G(x^k)\right)\right] \to \overline{w}.$$

Considering the objective function $f(x) \doteq -\overline{w}^\top x$, the above expression implies

$$\nabla f(x^k) + DG(x^k)^* \left[\Pi_{\mathcal{K}^\circ}\left(\rho_k G(x^k)\right)\right] \to 0.$$

Thus, we conclude that $\overline{x}$ is an AL-AKKT point by taking $\tilde{\mu}^k \doteq 0$ for all $k \in \mathbb{N}$. Since we are assuming that AL-AKKT and KKT are equivalent, then $\overline{x}$ is also KKT, which means $\overline{w} = -\nabla f(\overline{x}) \in H(\overline{x})$ and, consequently, it follows that AL-regularity holds at $\overline{x}$. □

3.5 Wrap up

In this chapter we presented extensions of the most consolidated sequential optimality conditions from Nonlinear Programming (NLP) to Nonlinear Conic Programming (NCP). This was done by revisiting them from a geometric point of view. The so-called Approximate Karush-Kuhn-Tucker (AKKT) condition is the simplest of all, as it can be seen as a perturbed form of the classical KKT conditions, and in this summary of our paper [9] it served as a comprehensible example of how to use such sequential conditions as theoretical tools for building algorithmic convergence results. Our main advance in [9] is obtained by the Approximate Gradient Projection (AGP, Definition 3.2.1) condition, which is strictly stronger than AKKT but still keeps most of its known properties; moreover, by equipping AGP with a certain control on the approximate Lagrange multiplier sequence we obtain an improved convergence theory of some algorithms (see Definition 3.3.1 and Section 3.3.2) with special emphasis on the Augmented Lagrangian method. In fact, it is worth repeating that Theorem 3.2.2 was new even in the context of NLP and Theorem 3.3.1 was new in Nonlinear Semidefinite and Second-Order Cone Programming at the time [9] was submitted for publication.

Chapter 4

The many faces of constant rank conditions

Perhaps one of the most significant Constraint Qualifications in optimization is the so-called *Constant Rank Constraint Qualification* (CRCQ). To introduce it properly, let us recall its birthplace, the standard *Nonlinear Programming* (NLP) problem with separate equality constraints:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & g(x) \geqslant 0, \\ & h(x) = 0, \end{aligned} \tag{NLP}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$, $g\colon \mathbb{R}^n \to \mathbb{R}^m$, $h\colon \mathbb{R}^n \to \mathbb{R}^p$ are at least twice continuously differentiable. This classical problem can be seen as a particular case of (NCP) by setting

$$G(x) \doteq (g(x), h(x)) \quad \text{and} \quad \mathcal{K} \doteq \mathbb{R}^m_+ \times \{0\}^p,$$

and as usual in NLP (see also Example 2.3.1), given a feasible point $\overline{x}$ of (NLP), we will denote the set of active inequality constraints at $\overline{x}$ by $\mathcal{A}(\overline{x}) \doteq \{j \in \{1, \dots, m\} \colon g_j(\overline{x}) = 0\}$. According to Janin [57], the initial purpose of CRCQ was to study differentiability properties of the *value function* of a *parametric version of* (NLP) which is roughly a generalization of the function $\mathcal{P}$ for (Par-LP). Prior to Janin, several authors have found upper and lower bounds for the directional derivatives of the value function along specific directions under Mangasarian-Fromovitz' CQ (which in this case is equivalent to Robinson's CQ), and an actual formula was only obtainable under the stronger condition:

- **Linear Independence CQ:** We say that $\overline{x} \in \Omega$ satisfies the Linear Independence CQ (LICQ) when the family of gradients of active constraints at $\overline{x}$:

$$\{\nabla g_j(\overline{x})\}_{j\in\mathcal{A}(\overline{x})} \bigcup \{\nabla h_i(\overline{x})\}_{i\in\{1,\dots,p\}} \tag{4.1}$$

  has full rank, i.e., it is linearly independent;

which coincides with Nondegeneracy after being specialized to NLP [57]. Janin managed to improve these results by considering more general perturbations and obtaining such a formula under a relaxed notion of regularity, which allows the rank of (4.1) to be incomplete as long as it remains the same, together with all of its subfamilies, in a neighborhood of $\overline{x}$:

Definition 4.0.1 (Definition 1 of [57]). Let $\overline{x}$ be a feasible point of (NLP). We say that the *Constant Rank Constraint Qualification for NLP* (CRCQ) holds at $\overline{x}$ if there exists a neighborhood

$\mathcal{V}$ of $\overline{x}$ such that: for every $I \subseteq \{1, \ldots, p\}$ and every $J \subseteq \mathcal{A}(\overline{x})$ the rank of the family

$$\{\nabla g_j(x)\}_{j \in J} \bigcup \{\nabla h_i(x)\}_{i \in I}$$

remains constant for all $x \in \mathcal{V}$.

The reason why all subsets of $\mathcal{A}(\overline{x})$ and $\{1, \ldots, p\}$ must be considered appears to be mostly technical: to prove that CRCQ guarantees existence of Lagrange multipliers, Janin sought to show that it implies $L_\Omega(\overline{x}) \subseteq T_\Omega(\overline{x})$, which would in turn imply the NLP version of Abadie's CQ. But after picking an arbitrary direction

$$d \in L_\Omega(\overline{x}) = \left\{ d \in \mathbb{R}^n \colon \begin{array}{l} \langle \nabla g_j(\overline{x}), d \rangle \geq 0, \ j \in \mathcal{A}(\overline{x}), \\ \langle \nabla h_j(\overline{x}), d \rangle = 0, \ j \in \{1, \ldots, p\} \end{array} \right\},$$

although it was sufficient to have the constant rank assumption only for the indices $j \in \mathcal{A}(\overline{x})$ such that $\langle \nabla g_j(\overline{x}), d \rangle = 0$ for proving that $d \in T_\Omega(\overline{x})$, those indices were not determined *a priori*. Then, he considered all possibilities. Several years later, Minchenko and Stakhovski [68] noticed that taking subsets of the equality constraints was completely superfluous, which led to an enhanced definition of CRCQ:

Definition 4.0.2 (RCRCQ [68])**.** Let $\overline{x}$ be a feasible point of (NLP). We say that *Relaxed Constant Rank Constraint Qualification for NLP* (RCRCQ) holds at $\overline{x}$ if there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that, for every subset $J \subseteq \mathcal{A}(\overline{x})$, the rank of the family

$$\{\nabla g_j(x)\}_{j \in J} \bigcup \{\nabla h_i(x)\}_{i \in \{1, \ldots, p\}}$$

remains constant for all $x \in \mathcal{V}$.

With this new condition at hand, Minchenko and Stakhovski were able to improve all of Janin's work [68, 69]. In fact, a closer look at Janin's paper [57] reveals that some of his proofs could also be completed with RCRCQ, which suggests that RCRCQ is actually a corrected version of CRCQ instead of just an improvement. Nevertheless, even RCRCQ still seems too technical without a geometric way of visualizing it; and perhaps this is the main reason why it has never been extended to a non-polyhedral conic context before. In fact, some geometric insight seems fundamental in the study of regularity for (NCP) since almost all Constraint Qualifications available for it are roughly variants of Nondegeneracy or Robinson's CQ – both of which have high geometric appeal – and the first attempt [91] of extending (R)CRCQ to a non-polyhedral problem (for the record, NSOCP) without following a geometric approach was incorrect [7].

4.1 Revisiting (R)CRCQ from a geometric point of view

The main difficulty in extending (R)CRCQ to the conic setting is that it relies too much on the notions of "active constraints" and "subsets of the indices of active constraints," which seem somehow intrinsic to NLP. But with the proper background (Chapter 2) generalizing these concepts becomes simply a matter of reinterpretation. For instance, recall from Example 2.3.1 that isolating active constraints is simply one type of *reduction*, in the sense of Section 2.3.3. Indeed, for any given $\overline{x} \in \Omega$, the cone $\mathbb{R}^m_+ \times \{0\}^p$ is reducible at $G(\overline{x})$ to

$$\mathcal{C} \doteq \mathbb{R}^{|\mathcal{A}(\overline{x})|}_+ \times \{0\}^p$$

in a neighborhood $\mathcal{N}$ of $G(\overline{x})$ by the mapping $\Xi\colon \mathcal{N} \to \mathbb{R}^{|\mathcal{A}(\overline{x})|+p}$ such that

$$\Xi(y,z) \doteq \left[[y_j]_{j\in\mathcal{A}(\overline{x})},\ z\right]$$

for every $(y,z) \in \mathcal{N}$, and in this case the reduced constraint function of (NLP) at $\overline{x}$ takes the form

$$\mathcal{G}(x) \doteq \Xi(G(x)) = \left[[g_j(x)]_{j\in\mathcal{A}(\overline{x})},\ h(x)\right]. \tag{4.2}$$

The notion of "subsets of the indices of active constraints," on the other hand, can be interpreted in terms of faces – see Section 2.1.3. It is easy to see that every face of $\mathbb{R}_+^{|\mathcal{A}(\overline{x})|}$ can be written in terms of a unique subset of the canonical vectors of $\mathbb{R}^{|\mathcal{A}(\overline{x})|}$, which we will denote by $c_1,\ldots,c_{|\mathcal{A}(\overline{x})|}$. That is, $R \trianglelefteq \mathbb{R}_+^{|\mathcal{A}(\overline{x})|}$ if, and only if, there exists some $J \subseteq \{1,\ldots,|\mathcal{A}(\overline{x})|\}$ such that

$$R = \mathbb{R}_+^{|\mathcal{A}(\overline{x})|} \bigcap_{j\in J} \{c_j\}^{\perp}, \tag{4.3}$$

where R and J are clearly in a one-to-one correspondence. Each face F of $\mathcal{C}$ has, in turn, the form

$$F = R \times \{0\}^p$$

for some $R \trianglelefteq \mathbb{R}_+^{|\mathcal{A}(\overline{x})|}$ (see Lemma 2.1.9). With this in mind, notice that

$$\begin{aligned}\text{rank}\left(\{\nabla g_j(x)\}_{j\in J}\bigcup\{\nabla h_i(x)\}_{i\in\{1,\ldots,p\}}\right) &= \dim\left(D\mathcal{G}(x)^\top[\text{span}\left(\{c_j\}_{j\in J}\right)\times\mathbb{R}^p]\right)\\ &= \dim\left(D\mathcal{G}(x)^\top[F^\perp]\right)\end{aligned}$$

for every $x \in \mathbb{R}^n$. And this leads to a natural characterization of RCRCQ in terms of the faces of $\mathcal{C}$, as follows:

Proposition 4.1.1 (Proposition 3.1 of [14]). *Let $\overline{x}$ be a feasible point of* (NLP)*. Then, RCRCQ holds at $\overline{x}$ if, and only if, there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that, for each $F \trianglelefteq \mathbb{R}_+^{|\mathcal{A}(\overline{x})|}\times\{0\}^p$, the dimension of*

$$D\mathcal{G}(x)^\top[F^\perp]$$

remains constant for every $x \in \mathcal{V}$, where $\mathcal{G}$ is as defined in (4.2)*.*

This equivalent form of RCRCQ is what allows us to visualize what it actually describes. Indeed, for any fixed $x \in \mathbb{R}^n$, note that for every $y \in D\mathcal{G}(x)^{-1}(\text{span}(F))$ and every $w \in F^\perp$ we have

$$\langle D\mathcal{G}(x)[y], w\rangle = \langle y, D\mathcal{G}(x)^\top[w]\rangle = 0$$

which implies that[1]

$$(D\mathcal{G}(x)^\top[F^\perp])^\perp = D\mathcal{G}(x)^{-1}(\text{span}(F)).$$

Then, RCRCQ holds at $\overline{x}$ if, and only if, the dimension of $D\mathcal{G}(x)^{-1}(\text{span}(F))$ remains constant in a neighborhood of $\overline{x}$, for each $F \trianglelefteq \mathcal{C}$. Now, intuitively speaking, $\text{span}(F)$ can be regarded as a linear approximation of $\mathcal{C}$, while $D\mathcal{G}(x)$ is a linear approximation of $\mathcal{G}(x)$ near $\overline{x}$ because $\mathcal{G}(\overline{x}) = 0$; hence, $D\mathcal{G}(x)^{-1}(\text{span}(F))$ is essentially a "linear approximation" of the feasible set $\Omega = \mathcal{G}^{-1}(\mathcal{C})$ around $\overline{x}$. In these terms,

RCRCQ holds at $\overline{x}$ when every linear approximation of Ω at $\overline{x}$ is invariant to sufficiently small perturbations, up to isomorphism.

[1] In fact, for any linear operator U and any linear space S, it holds that $U^*[S^\perp]^\perp = U^{-1}(S)$. A particular case of this result is the well-known equation: $\text{Im}(U^*)^\perp = \text{Ker}(U)$, which is obtained by setting $S = \{0\}$.

Well, since $\mathcal{G}$ is differentiable, its best linear approximation at $\overline{x}$ is uniquely determined; but $\mathcal{C}$ may have many different approximations, one for each face, and it is not clear a priori which one is the "best" one, so we must consider all of them. This is why all subsets of active constraints are taken into account by RCRCQ.

To obtain a characterization of the original CRCQ in the same style of Proposition 4.1.1, one should first reformulate the equality constraint $h(x) = 0$ as a pair of inequality constraints $h(x) \geqslant 0$ and $-h(x) \geqslant 0$. In other words, take

$$\mathcal{K} \doteq \mathbb{R}^m_+ \times \mathbb{R}^p_+ \times \mathbb{R}^p_+ \quad \text{and} \quad G(x) \doteq (g(x), h(x), -h(x)),$$

repeat the reasoning that precedes Proposition 4.1.1, and a characterization of CRCQ appears. After realizing how unnatural this sounds, we are led to conclude once again that RCRCQ is indeed a corrected version of CRCQ instead of a mere improvement. For this reason, from this point onwards, we will always refer to both RCRCQ and CRCQ as simply "CRCQ".

4.2 Facial Constant Rank for reducible cones

"I never bothered about whether what would come out would be suitable for this or that, but just tried to understand – and it always turned out that understanding was all that mattered."

Alexander Grothendieck

Geometric or not, CRCQ is already very well known – perhaps even classical – in NLP, and in this context Proposition 4.1.1 is nothing but a lovely, though apparently useless, observation. However, contrary to the original definition of CRCQ, our characterization is almost trivially extensible to the general class of optimization problems with *reducible* cone-constraints. That is:

$$\begin{aligned} &\underset{x \in \mathbb{R}^n}{\text{Minimize}} && f(x), \\ &\text{subject to} && G(x) \in \mathcal{K}, \end{aligned} \tag{NCP}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $G\colon \mathbb{R}^n \to \mathbb{Y}$ are assumed to be at least twice continuously differentiable, and $\mathcal{K}$ is reducible in $\operatorname{Im}(G)$, in the sense of Section 2.3.3. That is, for each $\overline{x} \in \Omega$ there is a neighborhood $\mathcal{N}$ of $G(\overline{x})$, a reduction mapping $\Xi\colon \mathcal{N} \to \mathbb{F}$, and a pointed closed convex cone $\mathcal{C} \subseteq \mathbb{F}$ such that $\mathcal{G}(\overline{x}) \doteq \Xi(G(\overline{x})) = 0$ and

$$\mathcal{N} \cap \Omega = \mathcal{N} \cap \mathcal{G}^{-1}(\mathcal{C}).$$

In this setting $\Xi \doteq \Xi_{\overline{x}}$ and $\mathcal{G} \doteq \mathcal{G}_{\overline{x}}$ both depend on $\overline{x}$, but this fact will be implicit in our notation, unless the object of such dependency is not clear by the context. Observe that a direct generalization of CRCQ from Proposition 4.1.1 does not define a Constraint Qualification, since it would always hold when G (and, by extension, $\mathcal{G}$) is affine, but this quality alone is not enough to guarantee existence of Lagrange multipliers unless the set

$$\begin{aligned} H(\overline{x}) &\doteq DG(\overline{x})^*[N_{\mathcal{K}}(G(\overline{x}))] \\ &= D\mathcal{G}(\overline{x})^*[\mathcal{C}^\circ] \end{aligned}$$

is closed; see also (2.24) and (2.34). Assembling all of this together, we arrive at the following extension of CRCQ:

Definition 4.2.1 (CRCQ). Let $\overline{x} \in \Omega$.

1. The *Facial Constant Rank* (FCR) property holds at $\overline{x}$ with respect to a given reduction mapping Ξ if there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that, for every $F \trianglelefteq \mathcal{C}$, the dimension of $D\mathcal{G}(x)^*[F^\perp]$ remains constant for every $x \in \mathcal{V}$;

2. The *Constant Rank Constraint Qualification* (CRCQ) holds at $\overline{x}$ if it satisfies the Facial Constant Rank property with respect to some Ξ and the set $H(\overline{x}) \doteq D\mathcal{G}(\overline{x})^*[\mathcal{C}^\circ]$ is closed.

In order to prove that CRCQ is a Constraint Qualification for (NCP) we first need to extend a result of Andreani et al. [6, Proposition 3.1], which is in turn a simplified version of the result originally employed by Janin [57, Proposition 2.2].

Proposition 4.2.1. *(Proposition 3.1 of [6]) Let $\{\zeta_i\}_{i\in\mathcal{I}}$ be a finite family of differentiable functions $\zeta_i\colon \mathbb{R}^n \to \mathbb{R}$, $i \in \mathcal{I}$, such that the family of its gradients $\{\nabla\zeta_i(x)\}_{i\in\mathcal{I}}$ remains with constant rank for every x in a neighborhood of $\overline{x}$, and consider the linear subspace*

$$\mathcal{S} \doteq \{y \in \mathbb{R}^n \colon \forall i \in \mathcal{I},\ \langle \nabla\zeta_i(x), y\rangle = 0\}.$$

Then, there exists some neighborhoods $\mathcal{V}_1$ and $\mathcal{V}_2$ of $\overline{x}$, and a diffeomorphism $\psi : \mathcal{V}_1 \to \mathcal{V}_2$, such that $\psi(\overline{x}) = \overline{x}$, $D\psi(\overline{x}) = \mathbb{I}_n$, and

$$\zeta_i(\psi^{-1}(\overline{x} + y)) = \zeta_i(\psi^{-1}(\overline{x}))$$

for every $y \in \mathcal{S} \cap (\mathcal{V}_2 - \overline{x})$ and every $i \in \mathcal{I}$.

To make this result suitable for the conic environment, we will extend it as follows:

Lemma 4.2.2 (Curve builder). *Let $\mathcal{G}\colon \mathbb{R}^n \to \mathbb{F}$ be twice differentiable, and let $\mathcal{W} \subseteq \mathbb{F}$ be any subspace with dimension N. Also, let $\overline{x}, d \in \mathbb{R}^n$ be such that $D\mathcal{G}(\overline{x})[d] \in \mathcal{W}^\perp$. If there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that $D\mathcal{G}(x)^*[\mathcal{W}]$ remains with constant dimension for every $x \in \mathcal{V}$, then there exists some $\varepsilon > 0$ and a twice differentiable curve $\xi\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ such that*

$$\langle \mathcal{G}(\xi(t)), w\rangle = \langle \mathcal{G}(\overline{x}), w\rangle$$

for every $t \in (-\varepsilon, \varepsilon)$ and every $w \in \mathcal{W}$; moreover, $\xi(0) = \overline{x}$ and $\xi'(0) = d$.

Proof. Let $\eta_1, \ldots, \eta_N$ be a basis of $\mathcal{W}$, and note that

$$D\mathcal{G}(x)^*[\mathcal{W}] = \text{span}\left(\{D\mathcal{G}(x)^*[\eta_i]\}_{i\in\{1,\ldots,N\}}\right). \tag{4.4}$$

Therefore, the hypothesis on the constant dimension of $D\mathcal{G}(x)^*[\mathcal{W}]$ can be equivalently stated as the constant rank of the family

$$\{D\mathcal{G}(x)^*[\eta_i]\}_{i\in\{1,\ldots,N\}}$$

in a neighborhood of $\overline{x}$. Furthermore, let

$$\zeta_i(x) \doteq \langle \mathcal{G}(x), \eta_i\rangle$$

and note that

$$\nabla\zeta_i(x) = D\mathcal{G}(x)^*[\eta_i]$$

for every $i \in \{1, \ldots, N\}$.

Then, by Proposition 4.2.1, there exist neighborhoods $\mathcal{V}_1$ and $\mathcal{V}_2$ of $\overline{x}$, and a diffeomorphism $\psi \colon \mathcal{V}_1 \to \mathcal{V}_2$ such that:

- $\psi(\overline{x}) = \overline{x}$;
- $D\psi(\overline{x}) = \mathbb{I}_n$;
- $\zeta_i(\psi^{-1}(\overline{x} + y)) = \zeta_i(\overline{x})$ for every $i \in \{1, \ldots, N\}$ and every $y \in \mathcal{S}$;

where

$$\mathcal{S} \doteq \{y \in \mathbb{R}^n \colon \forall i \in \{1, \ldots, N\},\ \langle \nabla \zeta_i(\overline{x}), y \rangle = 0\}.$$

Since $D\mathcal{G}(\overline{x})[d] \in \mathcal{W}^\perp$, we see that

$$\langle d, D\mathcal{G}(\overline{x})^*[\eta_i] \rangle = \langle D\mathcal{G}(\overline{x})[d], \eta_i \rangle = 0$$

for every $i \in \{1, \ldots, N\}$, so $d \in \mathcal{S}$. Then, let $\varepsilon > 0$ be such that $\overline{x} + td \in \mathcal{V}_2$ for every $t \in (-\varepsilon, \varepsilon)$ and define $\xi(t) \doteq \psi^{-1}(\overline{x} + td)$ for every such t. Note that $\xi'(t) = d$ and $\xi(0) = \overline{x}$. Moreover, for every $t \in (-\varepsilon, \varepsilon)$, we have $\langle \mathcal{G}(\xi(t)), \eta_i \rangle = \langle \mathcal{G}(\overline{x}), \eta_i \rangle$ for every $i \in \{1, \ldots, N\}$. Finally, it is essential to recall that the degree of differentiability of ξ is the same as the one of $\mathcal{G}$, which is a fact that follows from [69, Page 328]. □

We are now ready to prove that CRCQ[2] as in Definition 4.2.1 is indeed a Constraint Qualification for (NCP).

Theorem 4.2.3. *If $\overline{x} \in \Omega$ satisfies the FCR property, then $T_\Omega(\overline{x}) = L_\Omega(\overline{x})$.*

Proof. Because $T_\Omega(\overline{x}) \subseteq L_\Omega(\overline{x})$ is always true, we will only prove that $L_\Omega(\overline{x}) \subseteq T_\Omega(\overline{x})$, so let $d \in L_\Omega(\overline{x})$. Due to (2.33) we know that $L_\Omega(\overline{x}) = D\mathcal{G}(\overline{x})^{-1}(\mathcal{C})$ and $D\mathcal{G}(\overline{x})[d] \in \mathcal{C}$. Now take

$$F \doteq F_{\min}(D\mathcal{G}(\overline{x})[d]) \trianglelefteq \mathcal{C}$$

and the FCR property ensures that the dimension of $D\mathcal{G}(x)^*[F^\perp]$ remains constant in a neighborhood of $\overline{x}$. Hence, the curve builder (Lemma 4.2.2) may be applied with $\mathcal{W} \doteq F^\perp$ to construct a twice differentiable diffeomorphism $\xi \colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ such that $\xi(0) = \overline{x}$, $\xi'(0) = d$, and $\langle \mathcal{G}(\xi(t)), w \rangle = \langle \mathcal{G}(\overline{x}), w \rangle$ for every $w \in F^\perp$ and every $t \in (-\varepsilon, \varepsilon)$. This implies that

$$\mathcal{G}(\xi(t)) \in (F^\perp)^\perp = \mathrm{span}(F)$$

for every $t \in (-\varepsilon, \varepsilon)$. Moreover, taking the Taylor expansion of $\mathcal{G}(\xi(t))$ around $t = 0$, we see that

$$\mathcal{G}(\xi(t)) = tD\mathcal{G}(\overline{x})[d] + o(t),$$

but since $D\mathcal{G}(\overline{x})[d] \in \mathrm{ri}(F)$ it follows that $\mathcal{G}(\xi(t)) \in \mathrm{ri}(F) \subseteq \mathcal{C}$ as well, for every $t \in [0, \varepsilon)$ shrinking ε if necessary. Hence, $d \in T_\Omega(\overline{x})$, but since d is arbitrary, $L_\Omega(\overline{x}) \subseteq T_\Omega(\overline{x})$. □

Corollary 4.2.4. *If $\overline{x} \in \Omega$ satisfies CRCQ, then it also satisfies Abadie's CQ. By consequence, $\Lambda(\overline{x}) \neq \emptyset$.*

Theorem 4.2.3 and Corollary 4.2.4 extend and grotesquely simplify both specialized results presented in [14, Theorems 4.1 and 5.2] for NSOCP and NSDP. Even so, we recommend reading [14] to see how concrete examples of reductions are built, but the similarities with it end here.

[2]It is worth emphasizing that every occurrence of the initials "CRCQ" from this point onwards will always refer to Definition 4.2.1 instead of the original NLP condition, unless specified otherwise.

4.2.1 Constant Rank of the Subspace Component

At the beginning of this section, we stated that CRCQ describes a scenario where all "linear approximations" of Ω are invariant to small perturbations, up to isomorphism. But why do we need *all* linear approximations? Isn't the *best* linear approximation enough? Well, the answer is *yes*, but we must know which face of $\mathcal{C}$ defines the best approximation of Ω.

To get some intuition on which face is the good one, let us drive our attention to the particular case of (NLP) once more. For the sake of simplicity, we will omit equality constraints in the next paragraphs, since they can be easily carried over in all of the following computations, so we deal with $\mathcal{G}(x) \doteq [g_j(x)]_{j\in\mathcal{A}(\overline{x})}$ and $\mathcal{C} \doteq \mathbb{R}_+^{|\mathcal{A}(\overline{x})|}$. In Andreani et al. [18] the authors presented a relaxation of CRCQ that was consequence of realizing that the index set

$$J_- \doteq \{j \in \mathcal{A}(\overline{x})\colon \nabla g_j(\overline{x}) \in L_\Omega(\overline{x})^\circ\}$$

was more special than the others because, in NLP,

$$\text{span}\left(\{\nabla g_j(\overline{x})\}_{j\in J_-}\right) = \text{lin}(L_\Omega(\overline{x})^\circ)$$

which is the *subspace component* of $L_\Omega(\overline{x})^\circ$. With this in mind, they defined the so-called *Constant Rank of the Subspace Component* (CRSC) condition [18, Definition 1.3], which holds at $\overline{x}$ when the family of vectors

$$\{\nabla g_j(x)\}_{j\in J_-}$$

remains with constant rank for every x in a neighborhood of $\overline{x}$. The CRSC condition was then studied by Kruger et al. [63], who gave remarkable insights on it. For starters, they noticed that the set J_- can be equivalently written as

$$J_- = \{j \in \mathcal{A}(\overline{x})\colon \langle \nabla g_j(\overline{x}), d\rangle = 0,\ \forall d \in L_\Omega(\overline{x})\}. \tag{4.5}$$

Then, for every $d \in L_\Omega(\overline{x})$ and every $j \in J_-$, we obtain from (4.5) that

$$\langle D\mathcal{G}(\overline{x})[d], e_j\rangle = \langle \nabla g_j(\overline{x}), d\rangle = 0. \tag{4.6}$$

To adapt these facts to our "facial" language, we consider the face of $\mathcal{C}$ associated with J_-, given by

$$F_{J_-} \doteq \mathbb{R}^m_+ \bigcap_{j\in J_-} \{e_j\}^\perp$$

and then observe that (4.6) implies

$$D\mathcal{G}(\overline{x})[L_\Omega(\overline{x})] \subseteq F_{J_-}.$$

To proceed further, we can still rely on the discussion in Kruger et al. [63, below Definition 6], which guides us to see that for each $j \notin J_-$ there is some $d_j \in L_\Omega(\overline{x})$ such that $\langle D\mathcal{G}(\overline{x})[d_j], e_j\rangle > 0$ so $d_{J_-} \doteq \sum_{j\in J_-} d_j \in L_\Omega(\overline{x})$ satisfies $\langle D\mathcal{G}(\overline{x})[d_{J_-}], e_j\rangle > 0$ for every $j \notin J_-$ and, consequently,

$$D\mathcal{G}(\overline{x})[d_{J_-}] \in \text{ri}(F_{J_-}) \cap D\mathcal{G}(\overline{x})[L_\Omega(\overline{x})].$$

Thus, since $\text{ri}(F_{J_-}) \cap DG(\overline{x})[L_\Omega(\overline{x})] \neq \emptyset$, we conclude by Proposition 2.1.10 (item 2) that

$$F_{J_-} = F_{\min}\left(D\mathcal{G}(\overline{x})[L_\Omega(\overline{x})]\right). \tag{4.7}$$

We have just found our *special face*. Back to the general problem (NCP) and inspired by the discussion above, for any given $\overline{x} \in \Omega$ consider the subset $\Gamma_{\mathcal{C}}(\overline{x})$ of $\mathcal{C}$ given by

$$\Gamma_{\mathcal{C}}(\overline{x}) \doteq D\mathcal{G}(\overline{x})[L_{\Omega}(\overline{x})] \tag{4.8}$$

and we can extend CRSC to the context of NCP as well.

Definition 4.2.2 (CRSC). Let $\overline{x} \in \Omega$ and define $F_{J_-} \doteq F_{\min}(\Gamma_{\mathcal{C}}(\overline{x})) \trianglelefteq \mathcal{C}$. We say that the *Constant Rank of the Subspace Component* (CRSC) condition holds at $\overline{x}$ if $H(\overline{x})$ is closed and there exists some reduction mapping Ξ and a neighborhood $\mathcal{V}$ of $\overline{x}$ such that the dimension of $D\mathcal{G}(x)^*[F_{J_-}^{\perp}]$ remains constant for every $x \in \mathcal{V}$.

From our previous discussion, it follows immediately that Definition 4.2.2 fully recovers the CRSC condition from NLP when it is seen as a particular case of (NCP), since $H(\overline{x})$ is always closed in this case. Moreover, CRSC[3] as in Definition 4.2.2 is clearly implied by CRCQ but the following example, extracted from [18, Page 1113], shows that this implication is strict.

Example 4.2.1. *Consider the constraint*

$$G(x) \doteq (x, -x, -x^2) \in \mathbb{R}^3_+$$

along with its unique feasible point $\overline{x} = 0$. *In this case, the reduction mapping is simply the identity function, so* $\mathcal{G} = G$, *and* $\mathcal{C} = \mathcal{K}$. *Then, observe that*

$$\mathcal{L}_{\Omega}(\overline{x}) = \{0\} \subseteq \mathbb{R}, \quad \text{and} \quad \Gamma_{\mathcal{C}}(\overline{x}) = F_{J_-} = \{0\} \subseteq \mathbb{R}^3$$

and that

$$\dim(D\mathcal{G}(x)^*[F_{J_-}^{\perp}]) = \mathsf{rank}(D\mathcal{G}(x)) = 1, \ \forall x \in \mathbb{R}$$

so CRSC holds at $\overline{x}$, *but taking the face* $F = \mathsf{span}(\{(1,0,0),(0,1,0)\})$, *we see that*

$$\dim(D\mathcal{G}(x)^*[F^{\perp}]) = \dim(\mathsf{span}(\{2x\})) = \begin{cases} 1, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

meaning CRCQ fails at $\overline{x}$.

Now, to prove that CRSC still ensures that $\Lambda(\overline{x}) \neq \emptyset$ similarly to Theorem 4.2.3 and Corollary 4.2.4, we need a few extra lemmas, some of which were already presented in Chapter 2, and a new one:

Lemma 4.2.5. *Let* $\overline{x} \in \Omega$. *Then:*

1. *There exists some* $d \in D\mathcal{G}(\overline{x})^{-1}(\mathcal{C})$ *such that* $D\mathcal{G}(\overline{x})[d] \in \mathsf{ri}(F_{J_-})$;
2. $\mathsf{ri}(L_{\Omega}(\overline{x})) = D\mathcal{G}(\overline{x})^{-1}(\mathsf{ri}(F_{J_-}))$.

Proof. For item 1, recall from the definition that F_{J_-} is the smallest face of $\mathcal{C}$ that contains $\Gamma_{\mathcal{C}}(\overline{x})$, which means that $\mathsf{ri}(\Gamma_{\mathcal{C}}(\overline{x})) \cap \mathsf{ri}(F_{J_-}) \neq \emptyset$ (see Proposition 2.1.10, item 1). Therefore, there exists some $d \in D\mathcal{G}(\overline{x})^{-1}(\mathcal{C})$ such that $D\mathcal{G}(\overline{x})[d] \in \mathsf{ri}(F_{J_-})$. For item 2, note that $D\mathcal{G}(\overline{x})^{-1}(\mathsf{ri}(F_{J_-})) \neq \emptyset$ thanks to item 1, and Theorem 2.1.3 tells us that

$$\mathsf{ri}(L_{\Omega}(\overline{x})) = \mathsf{ri}(D\mathcal{G}(\overline{x})^{-1}(F_{J_-})) = D\mathcal{G}(\overline{x})^{-1}(\mathsf{ri}(F_{J_-}))$$

which is the desired result. ☐

[3] Similarly to CRCQ, every occurrence of "CRSC" from now on refers to Definition 4.2.2, not the NLP condition with the same name, unless specified otherwise.

These simple lemmas are not just technical and we shall elaborate on their meaning in a later section. For now, notice that item 1 implies that Slater's CQ always holds for the linearized constraint $D\mathcal{G}(\overline{x})[d] \in F_{J_-}$, regardless of its fulfilment for the original linearized constraint $D\mathcal{G}(\overline{x})[d] \in \mathcal{C}$, and item 2 gives a practical characterization of $\text{ri}(L_\Omega(\overline{x}))$ that enables us to borrow a clever argument from Minchenko and Stakhovski [68] in the following proof:

Theorem 4.2.6. *If $\overline{x} \in \Omega$ satisfies CRSC, then it also satisfies Abadie's CQ.*

Proof. Because $T_\Omega(\overline{x})$ is closed, it suffices to prove that $\text{ri}(L_\Omega(\overline{x})) \subseteq T_\Omega(\overline{x})$ to conclude that $L_\Omega(\overline{x}) \subseteq T_\Omega(\overline{x})$, and consequently, that $L_\Omega(\overline{x}) = T_\Omega(\overline{x})$ since the reverse inclusion always holds. So let $d \in \text{ri}(L_\Omega(\overline{x}))$ and we have $D\mathcal{G}(\overline{x})[d] \in \text{ri}(F_{J_-})$ due to Lemma 4.2.5, item 2. Just as in the proof of Theorem 4.2.3, take $\mathcal{W} = F_{J_-}^\perp$ and invoke the curve builder (Lemma 4.2.2) to obtain some $\varepsilon > 0$ and a curve $\xi\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ such that $\xi(0) = \overline{x}$, $\xi'(0) = d$, and $\mathcal{G}(\xi(t)) \in \text{ri}(F_{J_-})$ for every $t \in [0, \varepsilon)$. This implies that $d \in T_\Omega(\overline{x})$; hence, $L_\Omega(\overline{x}) \subseteq T_\Omega(\overline{x})$, and since $H(\overline{x})$ is closed by hypothesis, the proof is over. □

A couple of counterexamples presented by Janin [57, Examples 2.1 and 2.2] prove that the NLP version of CRCQ (and, by extension, its NCP version) is independent of Robinson's CQ. On the other hand, Robinson's CQ implies CRSC.

Proposition 4.2.7. *If $\overline{x} \in \Omega$ satisfies Robinson's CQ, then it also satisfies CRSC.*

Proof. Note that Robinson's CQ holds for the constraint $G(x) \in \mathcal{K}$ if, and only if, it holds for the reduced constraint $\mathcal{G}(x) \in \mathcal{C}$ regardless of the reduction mapping. On the other hand, note that we can assume without loss of generality that $\text{int}(\mathcal{C}) \neq \emptyset$ because otherwise $\mathbb{F}$ could be replaced with $\mathbb{F} \cap \text{span}(\mathcal{C})$ without any additional change. With this in mind, Robinson's CQ implies that there exists some $d \in \mathbb{R}^n$ such that $D\mathcal{G}(\overline{x})[d] \in \text{int}(\mathcal{C})$, which further implies that

$$D\mathcal{G}(\overline{x})[D\mathcal{G}(\overline{x})^{-1}(\text{int}(\mathcal{C}))] \cap \text{int}(\mathcal{C}) \neq \emptyset,$$

whence follows that $D\mathcal{G}(\overline{x})[L_\Omega(\overline{x})] \cap \text{int}(\mathcal{C}) \neq \emptyset$ and, consequently, $F_{J_-} = \mathcal{C}$ and $D\mathcal{G}(x)^*[F_{J_-}^\perp] = 0$. Since the latter has constant rank for every $x \in \mathbb{R}^n$, CRSC is satisfied. □

In fact, observe that if $\text{int}(\mathcal{C}) \neq \emptyset$ (which can always be assumed without loss of generality), then Robinson's CQ holds at $\overline{x}$ if, and only if, $F_{J_-} = \mathcal{C}$.

4.2.2 Facial reduction in Nonlinear Conic Programming

We start this section by calling the reader's attention to the fact

$$D\mathcal{G}(\overline{x})^{-1}(\mathcal{C}) = D\mathcal{G}(\overline{x})^{-1}(F_{J_-}) \tag{4.9}$$

which is true because if $d \in D\mathcal{G}(\overline{x})^{-1}(\mathcal{C})$, then $D\mathcal{G}(\overline{x})[d] \in \Gamma_\mathcal{C}(\overline{x}) \subseteq F_{J_-}$ and as a consequence $d \in D\mathcal{G}(\overline{x})^{-1}(F_{J_-})$; the other inclusion is trivial. In other words, the linearized constraint $D\mathcal{G}(\overline{x})[d] \in \mathcal{C}$ of the reduced problem (Red-NCP) is equivalent to its "facially reduced" version $D\mathcal{G}(\overline{x})[d] \in F_{J_-}$. Also, Lemma 4.2.5, item 1, tells us Slater's CQ always holds for the latter. This is not a coincidence, for F_{J_-} is essentially the output of a *facial reduction* algorithm applied to the linear approximation of (NCP) at $\overline{x}$. Naturally, one should question: is (NCP) equivalent to the "facially reduced" problem

$$\begin{array}{ll} \text{Minimize} & f(x), \\ \text{s.t.} & \mathcal{G}(x) \in F_{J_-}. \end{array} \tag{FRed-NCP}$$

at least in a neighborhood of $\overline{x}$, under some condition?

A positive answer has already been given under the NLP version of CRSC by Andreani et al. [18] in the context of NLP, which was obtained as a consequence of a result by Lu [66] regarding CRCQ. However, it appears that they were not aware of this interpretation of their results at the time, most likely due to the lack of Proposition 4.1.1. Thus, we will rely on everything we developed until this point to revisit their work. Let us start by recalling Lu's result, as stated in a paper by Andreani et al. [18]:

Proposition 4.2.8. *Let $\zeta_i\colon \mathbb{R}^n \to \mathbb{R}$, $i \in \{1,\ldots,s\}$ be continuously differentiable functions, let $\overline{x}$ be such that $\zeta_i(\overline{x}) = 0$ for every i, and let $J \subseteq \{1,\ldots,s\}$ be such that the family $\{\nabla\zeta_i(x)\}_{i\in J}$ remains with constant rank in a neighborhood $\mathcal{V}$ of $\overline{x}$. If there exists a $|J|$-dimensional vector $\gamma > 0$ (indexed by J) such that*

$$\sum_{i\in J} \nabla\zeta_i(\overline{x})\gamma_i = 0,$$

then there exists a neighborhood $\mathcal{U}$ of $\overline{x}$ such that for every $i \in J$ and every $x \in \mathcal{U}$, we have that $\zeta_i(x) = 0$ if, and only if, $\zeta_i(x) \geqslant 0$.

Now, let us adapt Proposition 4.2.8 to the Conic Programming context:

Proposition 4.2.9. *Let $\overline{x} \in \Omega$ and let $F \trianglelefteq \mathcal{C}$ be such that:*

1. *There exists some $0 \neq \eta \in \mathrm{ri}(F^{\triangle})$ such that $D\mathcal{G}(\overline{x})^*[\eta] = 0$;*
2. *The dimension of $D\mathcal{G}(x)^*[F^{\triangle}]$ remains constant in a neighborhood $\mathcal{V}$ of $\overline{x}$.*
3. *$F^{\triangle\perp} \cap \mathcal{C} = F$.*

Then, there exists a neighborhood $\mathcal{U}$ of $\overline{x}$ such that $\mathcal{G}^{-1}(\mathcal{C}) \cap \mathcal{U} = \mathcal{G}^{-1}(F) \cap \mathcal{U}$.

Proof. Let $s \doteq \dim(\mathrm{span}(F^{\triangle}))$ and let $0 \neq \eta \in \mathrm{ri}(F^{\triangle})$ be such that $D\mathcal{G}(\overline{x})^*[\eta] = 0$; this implies that $s \geqslant 1$. Also, let $\eta_1,\ldots,\eta_s \in F^{\triangle}$ be the vectors described in Lemma 2.1.6, which form a basis of $\mathrm{span}(F^{\triangle})$ and

$$\eta = \sum_{i=1}^{s} \alpha_i\eta_i,$$

for some vector of positive scalars $\alpha \doteq (\alpha_1,\ldots,\alpha_s)$. Now, consider the functions

$$\zeta_i(x) \doteq \langle \mathcal{G}(x), \eta_i\rangle, \quad i \in \{1,\ldots,s\}$$

along with the vector function $\zeta(x) \doteq (\zeta_1(x),\ldots,\zeta_s(x))$, and note that $\nabla\zeta_i(x) = D\mathcal{G}(x)^*[\eta_i]$ for every $i \in \{1,\ldots,s\}$. In particular, this implies that

$$\mathrm{rank}\left(\{\nabla\zeta_i(x)\}_{i\in\{1,\ldots,s\}}\right) = \dim\left(D\mathcal{G}(x)^*[\mathrm{span}(F^{\triangle})]\right)$$

for every $x \in \mathbb{R}^n$ and, moreover,

$$D\zeta(\overline{x})^\top\alpha = D\mathcal{G}(\overline{x})^*[\eta] = 0. \tag{4.10}$$

Applying Proposition 4.2.8, we obtain a neighborhood $\mathcal{U}$ of $\overline{x}$ such that $\zeta(x) = 0$ for every $x \in \mathcal{U}$ such that $\zeta(x) \leqslant 0$. However, note that $\eta_i \in \mathcal{C}^\circ$ for every $i \in \{1,\ldots,s\}$, so in particular this holds for every $x \in \Omega \cap \mathcal{U}$ because

$$\begin{aligned}\Omega &= \{x \in \mathbb{R}^n \colon \langle \mathcal{G}(x), \vartheta\rangle \leqslant 0,\ \forall\vartheta \in \mathcal{C}^\circ\}\\ &\subseteq \{x \in \mathbb{R}^n \colon \zeta(x) \leqslant 0\}\end{aligned}$$

Furthermore, since $\zeta(x) = 0$ is equivalent to $\mathcal{G}(x) \in F^{\triangle\perp}$ we conclude that

$$\mathcal{G}^{-1}(\mathcal{C}) \cap \mathcal{U} = \mathcal{G}^{-1}(F^{\triangle\perp} \cap \mathcal{C}) \cap \mathcal{U} = \mathcal{G}^{-1}(F) \cap \mathcal{U},$$

because $F^{\triangle\perp} \cap \mathcal{C} = F$. □

Proposition 4.2.9 can be considered an extension of the facial reduction lemma for (NCP), Now, let us recall a result by Pataki [75] to see how CRSC contributes to it.

Theorem 4.2.10 (Theorem 1 of [75]). *Let C be a nice closed convex cone, let M be a linear operator, and $F \doteq F_{\min}(\mathrm{Im}(M) \cap C)$. The following statements are equivalent:*

- $M^*[C^\circ]$ *is closed;*
- $\mathrm{ri}(F^\triangle) \cap \mathrm{Ker}(M^*) \neq \emptyset$ *and* $\mathrm{Im}(M) \cap F^{\triangle\perp} = \mathrm{Im}(M) \cap \mathrm{span}(F)$*;*
- $M^*[F^\triangle] = M^*[F^\perp]$.

This leads to the following generalization of the result by Andreani et al. [18], which can be interpreted as a nonlinear version of the facial reduction theorem:

Theorem 4.2.11. *Let $\overline{x} \in \Omega$ satisfy CRSC, suppose that $\mathcal{C}$ is nice, and let*

$$F \doteq F_{\min}(\Gamma_{\mathcal{C}}(\overline{x})).$$

Further, assume that there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that:

A1. $H(x)$ *is closed for every* $x \in \mathcal{V}$*;*

A2. $F = F_{\min}(\Gamma_{\mathcal{C}}(x))$, *for every* $x \in \mathcal{V}$*;*

A3. $F^{\triangle\perp} \cap \mathcal{C} = F$.

Then, there exists a neighborhood $\mathcal{U}$ of $\overline{x}$ such that

$$\mathcal{G}^{-1}(\mathcal{C}) \cap \mathcal{U} = \mathcal{G}^{-1}(F_{J_-}) \cap \mathcal{U}.$$

Proof. In the above theorem, take $M \doteq D\mathcal{G}(\overline{x})$, $C \doteq \mathcal{C}$, $F \doteq F_{\min}(\Gamma_{\mathcal{C}}(\overline{x}))$, and let $\eta \in \mathrm{ri}(\Gamma_{\mathcal{C}}(\overline{x}))$. Note also that

$$\Gamma_{\mathcal{C}}(\overline{x}) = D\mathcal{G}(\overline{x})[D\mathcal{G}(\overline{x})^{-1}(\mathcal{C})] = \mathrm{Im}(D\mathcal{G}(\overline{x})) \cap \mathcal{C}$$

and that $M^*[C^\circ] = H(\overline{x})$ in this case. By CRSC, $H(\overline{x})$ is closed, and because $\mathcal{C}$ is nice,

$$\mathrm{ri}(F_{J_-}^\triangle) \cap \mathrm{Ker}(D\mathcal{G}(\overline{x})^*) \neq \emptyset.$$

That is, there exists some $0 \neq \eta \in \mathrm{ri}(F_{J_-}^\triangle)$ such that $D\mathcal{G}(\overline{x})^*[\eta] = 0$. By A1 and A2, $D\mathcal{G}(x)^*[\mathcal{C}^\circ]$ is closed for every such $x \in \mathcal{V}$ and F is the minimal face of $\Gamma_{\mathcal{C}}(x)$, for every $x \in \mathcal{V}$, which implies that $D\mathcal{G}(x)^*[F^\triangle] = D\mathcal{G}(x)^*[F^\perp]$ for every such x. Since by CRSC the dimension of $D\mathcal{G}(x)^*[F^\perp]$ remains constant near x, say in $\mathcal{V}$, so does the dimension of $D\mathcal{G}(x)^*[F^\triangle]$. By Proposition 4.2.9, there is a neighborhood $\mathcal{U}$ of $\overline{x}$ such that $\mathcal{G}^{-1}(\mathcal{C}) \cap \mathcal{U} = \mathcal{G}^{-1}(F) \cap \mathcal{U}$. □

It is worth noticing that in NLP assumptions A1 and A2 are not required due to the polyhedricity of $\mathcal{C}$. Also, A3 is always true in NLP, NSOCP, and NSDP. Whether they can be removed from Theorem 4.2.11 or not in the general case is an open problem. We conjecture it is true:

Conjecture 4.2.12. *Let $\overline{x} \in \Omega$ satisfy CRSC, let $F \doteq F_{\min}(\Gamma_{\mathcal{C}}(\overline{x}))$, and suppose that $\mathcal{C}$ is nice. Then, there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that A1, A2, and A3 hold.*

4.3 The Strong Second-Order Optimality Condition

In the specialized work [14] for NSDP and NSOCP, we used CRCQ to obtain the Strong Second-Order Optimality Condition for (NCP) depending on any single given Lagrange multiplier. This result is, in particular, stronger than the classical second-order condition of Bonnans, Cominetti, and Shapiro [36]. However, since Robinson's CQ is not enough for obtaining a similar result, the same holds for CRSC. We can, though, define a new Constraint Qualification for (NCP) that is in-between CRCQ and CRSC, which can be used to obtain the same strong second-order condition of [14]. Formally:

Definition 4.3.1 (Strong-CRSC)**.** Let $\overline{x} \in \Omega$ and define $F_{J_-} \doteq F_{\min}(D\mathcal{G}(\overline{x})[L_\Omega(\overline{x})]) \trianglelefteq \mathcal{C}$. We say that *Strong-CRSC* holds at $\overline{x}$ when $H(\overline{x})$ is closed and there exists a reduction mapping Ξ and a neighborhood $\mathcal{V}$ of $\overline{x}$ such that for every $F \trianglelefteq F_{J_-}$, the dimension of $D\mathcal{G}(x)^*[F^\perp]$ remains constant for every $x \in \mathcal{V}$.

Observe that Strong-CRSC implies CRSC by taking $F = F_{J_-}$, and note also that CRCQ implies Strong-CRSC since all faces of F_{J_-} are, in particular, faces of $\mathcal{C}$. On the other hand, Example 4.2.1 proves that Strong-CRSC is still strictly weaker than CRCQ since in that example $F_{J_-} = \{0\}$ and thus Strong-CRSC coincides with CRSC there. Finally, it is worth mentioning that Strong-CRSC is new even in NLP, and any second-order result involving it is also an improvement of the existing NLP results (which are based on CRCQ). In fact, observe that Strong-CRSC is essentially an application of CRCQ to the "facially reduced" constraint $\mathcal{G}(x) \in F_{J_-}$.

To proceed, let us prove the strong second-order condition (SSOC, Definition 2.3.5) under Strong-CRSC:

Theorem 4.3.1. *Let $\overline{x} \in \Omega$ be a local minimizer of (NCP) that satisfies Strong-CRSC. Then, for every $\overline{\mu} \in \Lambda(\overline{x})$ and every $d \in C_T(\overline{x}) \doteq L_\Omega(\overline{x}) \cap \{\nabla f(\overline{x})\}^\perp$, the following inequality is satisfied:*

$$d^\top \nabla^2 f(\overline{x}) d + \left\langle D^2 G(\overline{x})[d,d],\ \overline{\mu} \right\rangle \geqslant \sigma(\overline{\mu},\ T^2_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d])). \tag{4.11}$$

where $T^2_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d])$ is the second-order tangent set[4] to $\mathcal{K}$ at $G(\overline{x})$ along $DG(\overline{x})[d]$ – see Definition 2.1.4.

Proof. Let $\overline{\mu} \in \Lambda(\overline{x})$ and $d \in C_T(\overline{x})$ be arbitrary; so $\nabla f(\overline{x})^\top d = 0$ and by [37, Equation 3.271] we see that $C_T(\overline{x}) = L_\Omega(\overline{x}) \cap \{\nabla f(\overline{x})\}^\perp$ and therefore $D\mathcal{G}(\overline{x})[d] \in F_{J_-}$ because of (4.9). Moreover, let F be the smallest face of F_{J_-} that contains $D\mathcal{G}(\overline{x})[d]$ in its relative interior[5]. Then, by Strong-CRSC, similarly to the proof of Theorem 4.2.6, there exists some $\varepsilon > 0$ and a twice continuously differentiable curve $\xi\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ such that $\xi(0) = \overline{x}$, $\xi'(0) = d$, and

$$\mathcal{G}(\xi(t)) \in \mathrm{ri}(F)$$

for all $t \in [0, \varepsilon)$. Since $\overline{x}$ is a local minimizer of (NCP) and $\xi(t)$ is, in particular, feasible for every small t, then $t = 0$ is a local minimizer of the function $\phi(t) \doteq f(\xi(t))$ subject to the constraint $t \geq 0$. Consequently,

$$\phi''(0) = d^\top \nabla^2 f(\overline{x}) d + \nabla f(\overline{x})^\top \xi''(0) \geqslant 0. \tag{4.12}$$

The rest of the proof consists of computing the term $\nabla f(\overline{x})^\top \xi''(0)$. To do so, let $\overline{\eta}$ be such that $\overline{\mu} = D\Xi(G(\overline{x}))^*[\overline{\eta}]$, which is uniquely determined since $D\Xi(G(\overline{x}))^*$ is injective. By the KKT

[4] Recall that reducibility implies second-order regularity (Proposition 2.3.4) which in turn implies that the inner and outer second-order tangent sets coincide (Proposition 2.1.17).

[5] Noteworthy, if $d \in \mathrm{ri}(L_\Omega(\overline{x}))$, this face is F_{J_-} itself.

conditions, we have that

$$\langle d, \nabla f(\overline{x})\rangle = -\langle d,\ DG(\overline{x})^*[\overline{\mu}]\rangle = -\langle DG(\overline{x})[d],\ \overline{\mu}\rangle = 0.$$

Therefore,

$$\begin{aligned} 0 = \langle DG(\overline{x})[d],\ \overline{\mu}\rangle &= \langle DG(\overline{x})[d],\ D\Xi(G(\overline{x}))^*[\overline{\eta}]\rangle \\ &= \langle D\Xi(G(\overline{x}))DG(\overline{x})[d],\ \overline{\eta}\rangle \\ &= \langle D\mathcal{G}(\overline{x})[d],\ \overline{\eta}\rangle \end{aligned}$$

so $D\mathcal{G}(\overline{x})[d] \in F \cap \{\overline{\eta}\}^\perp$ which is also a face of $\mathcal{C}$ since $\overline{\eta} \in \mathcal{C}^\circ$, but since F is by construction the minimal face containing $D\mathcal{G}(\overline{x})[d]$ we must have $F = F \cap \{\overline{\eta}\}^\perp$, hence $F \subseteq \{\overline{\eta}\}^\perp$ and, consequently, $\overline{\eta} \in F^\perp$. Then, consider the parametric complementarity function

$$R(t) \doteq \langle \mathcal{G}(\xi(t)),\ \overline{\eta}\rangle$$

and our previous reasoning implies that $R(t) = 0$ for every small $t \geq 0$. Differentiating R, we obtain:

$$R'(t) = \langle D\Xi(G(\xi(t)))DG(\xi(t))\xi'(t),\ \overline{\eta}\rangle .$$

Differentiating it once more, and taking the limit $t \to 0^+$, we obtain:

$$\begin{aligned} R''(0) &= \frac{d}{dt}\langle D\Xi(G(\xi(t)))DG(\xi(t))\xi'(t), \overline{\eta}\rangle\,|_{t=0} \\ &= \big\langle D^2\Xi(G(\overline{x}))[DG(\overline{x})[d], DG(\overline{x})[d]] + D\Xi(G(\overline{x}))D^2G(\overline{x})[d,d] + \\ &\qquad\qquad + D\Xi(G(\overline{x}))DG(\overline{x})\xi''(0),\ \overline{\eta}\big\rangle \\ &= 0. \end{aligned}$$

Using the fact

$$\langle D\Xi(G(\overline{x}))DG(\overline{x})\xi''(0),\ \overline{\eta}\rangle = \langle \xi''(0),\ DG(\overline{x})^*[\overline{\mu}]\rangle = -\nabla f(\overline{x})^\top \xi''(0),$$

it follows that

$$\nabla f(\overline{x})^\top \xi''(0) = \big\langle D^2\Xi(G(\overline{x}))[DG(\overline{x})[d], DG(\overline{x})[d]],\ \overline{\eta}\big\rangle + \big\langle D^2G(\overline{x})[d,d],\ \overline{\mu}\big\rangle . \tag{4.13}$$

Moreover, we obtain from Lemma 2.3.5 that

$$\sigma(\overline{\mu},\ T^2_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d])) = -\big\langle D^2\Xi(G(\overline{x}))[DG(\overline{x})[d], DG(\overline{x})[d]],\ \overline{\eta}\big\rangle .$$

Thus, substituting the above expressions in (4.12), we conclude that

$$d^\top \nabla^2 f(\overline{x})d + \big\langle D^2G(\overline{x})[d,d],\ \overline{\mu}\big\rangle \geqslant \sigma(\overline{\mu},\ T^2_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d])).$$

Since d and $\overline{\mu}$ were chosen arbitrarily, this proof is complete. □

Using only CRSC, we can obtain a weaker result with an analogous proof (see Footnote 5); however, we do not know whether this implies the second-order condition of Theorem 4.3.1 or not. What follows is a formal statement of such result:

Corollary 4.3.2. *Let $\overline{x} \in \Omega$ be a local minimizer of (NCP) that satisfies CRSC. Then, (4.11) holds for every $\overline{\mu} \in \Lambda(\overline{x})$ and every $d \in \mathrm{ri}(L_\Omega(\overline{x})) \cap \{\nabla f(\overline{x})\}^\perp \subseteq C_T(\overline{x})$.*

Another interesting result that can be extracted from the proof of the previous theorem is the following:

Proposition 4.3.3. *Let $\overline{x} \in \Omega$ be a local minimizer of (NCP) that satisfies Strong-CRSC. Then, for all Lagrange multipliers $\overline{\mu}_1, \overline{\mu}_2 \in \Lambda(\overline{x})$ and all $d \in C_T(\overline{x})$, we have that*

$$\left\langle D^2 G(\overline{x})[d,d],\ \overline{\mu}_1 - \overline{\mu}_2 \right\rangle - \sigma\left(\overline{\mu}_1 - \overline{\mu}_2,\ T^2_{\mathcal{K}}(G(\overline{x}), DG(\overline{x})[d])\right) = 0. \tag{4.14}$$

Proof. The conclusion follows directly from (4.13) and Lemma 2.3.5. □

As a corollary, it follows that under Strong-CRSC (and by extension CRCQ) the quadratic form associated with the generalized Hessian of (NCP) is invariant to the Lagrange multiplier over the directions in $C_T(\overline{x})$. A further implication is that if we assume that the dimension of $D\mathcal{G}(x)^*[F^\perp]$ remains constant for all x in a neighborhood of $\overline{x}$, only for the face $F = \{0\} \in \mathcal{C}$, then (4.14) holds true at least for every $d \in \operatorname{lin}(C_T(\overline{x})) = \{d \in \mathbb{R}^n \colon D\mathcal{G}(\overline{x})[d] = 0\}$, which generalizes a result obtained in [29, Theorem 3.3].

As a final comment, recall the discussion after Proposition 4.2.7 and observe that in the presence of Robinson's CQ we have $F_{J_-} = \mathcal{C}$ so Strong-CRSC coincides with CRCQ. Then, we recall two NLP examples of Janin [57] to show that Strong-CRSC is still independent of Robinson's CQ. First, we show that Robinson's CQ does not necessarily imply Strong-CRSC:

Example 4.3.1 (Example 2.1 of [57]). *Consider the constraint*

$$G(x) \doteq \begin{bmatrix} x_1 + x_2^2 \\ x_1 \end{bmatrix} \in \mathbb{R}^2_+$$

at the point $\overline{x} = 0$. Note that Robinson's CQ holds at $\overline{x}$ because for $d = (1,0)$ we have

$$G(\overline{x}) + DG(\overline{x})[d] = \begin{bmatrix} \langle (1, 2\overline{x}_2), (1,0) \rangle \\ \langle (1,0), (1,0) \rangle \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \operatorname{int}(\mathbb{R}^2_+).$$

Moreover, noticing that $F_{J_-} = \mathcal{C} = \mathbb{R}^2_+$ for any reduction mapping Ξ due to Robinson's CQ, it suffices to pick the face $F \doteq \{0\}$ to see that Strong-CRSC does not hold at $\overline{x}$ for

$$\dim(D\mathcal{G}(x)^*[F^\perp]) = \dim(D\mathcal{G}(x)^*[D\Xi(G(x))^*[F^\perp]]) = \dim\left(\operatorname{span}\left(\left\{ \begin{bmatrix} 1 \\ 2x_1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}\right)\right)$$

because $D\Xi(G(\overline{x}))^*$ is injective, so $\dim(D\mathcal{G}(x)^*[F^\perp])$ is equal to 2 for every x such that $x_1 \neq 0$, and equal to 1 at $\overline{x}$.

For the converse we refer directly to [57, Example 2.2], which analyses the constraint $G(x) \doteq [x, -x] \in \mathbb{R}^2_+$ at $\overline{x} = 0$, since it is very easy to see that Robinson's CQ does not hold at $\overline{x}$ whilst CRCQ does, implying also that Strong-CRSC holds at $\overline{x}$. Thus, Strong-CRSC is strictly weaker than CRCQ, stronger than CRSC, and independent of Robinson's CQ.

4.4 Wrap up

In this chapter, we introduced a new extension of the celebrated Constant Rank Constraint Qualification (CRCQ, Definition 4.2.1) for Nonlinear Conic Programming (NCP) problems with cone reducible constraints, that is defined in terms of the faces of the reduced cone $\mathcal{C}$. Because it recovers Janin's original definition when (NCP) is reduced to a standard Nonlinear Programming (NLP) problem, this condition is independent of Robinson's CQ and strictly implied by Nondegeneracy. A specialized version of this condition for Nonlinear Semidefinite and Second-Order Cone Programming has first appeared in our paper [14], where we presented an analogue of Theorem 4.3.1 as its main application. But this chapter contains advances of that paper in all directions. We presented a relaxation of the Constant Rank of the Subspace Component (CRSC, Definition 4.2.2) condition which also extends a concept with the same name from NLP [18] and, therefore, shares the property of being strictly weaker than both Robinson's CQ and CRCQ. This CRSC condition ensures existence of Lagrange multipliers, partially enables a certain type of nonlinear facial reduction result, and a slightly stronger version of it (Strong-CRSC, Definition 4.3.1) which is in-between CRCQ and CRSC and new even in NLP, can be used to derive the same result obtained in [14] regarding a stronger second-order optimality condition that strictly improves the classical one, but in a much more general context.

Chapter 5

Constant rank by paths and global convergence of algorithms

Despite the good theoretical properties of CRCQ as in Definition 4.2.1, it is still lacking practical applications. In this section, we study an alternative path for extending the Nonlinear Programming version of CRCQ (together with one of its most relevant variants) to the Conic Programming world. However, a general nonlinear problem such as (NCP) provides too few information and the faces of $\mathcal{K}$ are rarely used in algorithms that are designed to solve it. Thus, in this section we focus on its most popular instances: Nonlinear Semidefinite Programming (NSDP) and Nonlinear Second-Order Cone Programming (NSOCP). But since they are similar in several aspects, we will only present our results in the easier setting[1]:

$$\begin{array}{ll} \text{Minimize} & f(x), \\ \text{s.t.} & G(x) \succeq 0, \end{array} \tag{NSDP}$$

which is a particular case of (NCP) with $\mathcal{K} \doteq \mathbb{S}^m_+$ being the semidefinite cone. Also, we recall that the notation $G(x) \succeq 0$ stands for $G(x) \in \mathbb{S}^m_+$.

Throughout this whole chapter, we denote the eigenvalues of a given matrix $M \in \mathbb{S}^m$ arranged in non-decreasing order by

$$\lambda_1(M) \leqslant \lambda_2(M) \leqslant \ldots \leqslant \lambda_m(M)$$

and their associate eigenvectors by

$$u_1(M),\ u_2(M),\ \ldots,\ u_m(M),$$

respectively[2]. Moreover, assume that $M \succeq 0$, let r be the rank of M, and suppose that $r < m$. Then, there exists a neighborhood $\mathcal{V}$ of M such that every $N \in \mathcal{V}$ satisfies $\lambda_{m-r+1}(N) > \lambda_{m-r}(N)$. For all such N define the set

$$\mathcal{E}_r(N) \doteq \left\{ E \in \mathbb{R}^{m \times m-r} \colon \begin{array}{l} NE = E\mathrm{Diag}(\lambda_1(N), \ldots, \lambda_{m-r}(N)) \\ E^\top E = \mathbb{I}_{m-r} \end{array} \right\}, \tag{5.1}$$

that consists of all matrices whose columns are orthonormal eigenvectors associated with the $m-r$ smallest eigenvalues of N. In particular, observe that the elements of $\mathcal{E}_r(M)$ are precisely the matrices whose columns form an orthonormal basis of Ker(M). By convention, we define $\mathcal{E}_r(N) \doteq \emptyset$ for all $N \in \mathcal{V}$ when $r = m$.

[1] For a detailed exposition of the NSOCP case, we refer to our paper [15].

[2] For non-simple eigenvalues, any choice of eigenvectors will do, as long as it remains fixed.

5.1 Revisiting Nondegeneracy and Robinson's CQ

Inspired by the way Janin first derived CRCQ for NLP by relaxing the LICQ condition, we are led to investigate whether the Nondegeneracy condition of Shapiro and Fan [85] could also induce an extension of CRCQ to NSDP, given that it is the natural extension of LICQ. While the definition of Nondegeneracy for the general case (Definition 2.3.4) may discourage this kind of approach, the following characterization by Shapiro [83] seems to fit our purposes.

Theorem 5.1.1 (Proposition 6 of [83]). *Let $\overline{x} \in \Omega$ and $r \doteq \operatorname{rank}(G(\overline{x}))$. Also, let $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ be arbitrary and denote its columns by $\overline{e}_1, \ldots, \overline{e}_{m-r}$. Then $\overline{x}$ satisfies Nondegeneracy if, and only if, the vectors*

$$v_{ij}(\overline{x}, \overline{E}) \doteq \begin{bmatrix} \overline{e}_i^\top D_{x_1} G(\overline{x}) \overline{e}_j \\ \vdots \\ \overline{e}_i^\top D_{x_m} G(\overline{x}) \overline{e}_j \end{bmatrix} = DG(\overline{x})^* \left[\frac{\overline{e}_i \overline{e}_j^\top + \overline{e}_j \overline{e}_i^\top}{2} \right]$$

with i, j such that $1 \leqslant i \leqslant j \leqslant m - r$, are linearly independent[3].

Now, one could easily be fooled by the convenient phrasing of Theorem 5.1.1 and quickly imagine that an extension of CRCQ would easily come by replacing the "linear independence" requirement with a "constant rank in a neighborhood for every subfamily" one, following Janin's original approach. This would lead to the following condition:

- **Impostor CRCQ:** In the same setup of Theorem 5.1.1, the *impostor CRCQ* holds at $\overline{x}$ if, for every subset $J \subseteq \{(i, j) \colon 1 \leqslant i \leqslant j \leqslant m - r\}$, the family of vectors

$$\{v_{i,j}(x, \overline{E})\}_{(i,j) \in J}$$

  remains with constant rank for every x in a neighborhood of $\overline{x}$.

However, this is not a valid Constraint Qualification. In fact, let us examine the problem of minimizing $-x$ in $\mathbb{R}$ subject to

$$G(x) \doteq \begin{bmatrix} x & x + x^2 \\ x + x^2 & x \end{bmatrix} \succeq 0,$$

which has $\overline{x} \doteq 0$ as its unique global minimizer. Observe that for $\overline{E} \doteq \mathbb{I}_2 \in \mathcal{E}_r(G(\overline{x}))$ we have

$$v_{11}(x, \overline{E}) = v_{22}(x, \overline{E}) = 1 \quad \text{and} \quad v_{12}(x, \overline{E}) = 1 + 2x$$

for every $x \in \mathbb{R}$. So, clearly, the rank of every subfamily of $\{v_{ij}(x, \overline{E}) \colon 1 \leqslant i \leqslant j \leqslant 2\}$ remains constant for every x near $\overline{x}$, but $\overline{x}$ does not satisfy the KKT conditions – see Example 3.1 of [12]. So what is the problem?

On the one hand, the vectors $v_{i,j}(x, \overline{E})$ can be somehow seen as the gradients of the entries of the "active block" of $G(x)$ at $\overline{x}$, which is $\overline{E}^\top G(x) \overline{E}$, because

$$v_{ij}(x, \overline{E}) = \nabla_x \left(\overline{e}_i^\top G(x) \overline{e}_j \right)$$

for every i and j. But on the other hand, the off-diagonal entries of $\overline{E}^\top G(x) \overline{E}$, that is, $\overline{e}_i^\top G(x) \overline{e}_j$ for $i \neq j$, have no meaning on their own. Our point is that the impostor CRCQ would only make sense if the constraint $G(x) \succeq 0$ was somehow locally equivalent to the constraint $\overline{e}_i G(x) \overline{e}_j \geqslant 0$

[3] Observe that this holds for any $\overline{E}$, and that the linear independence statement does not depend on $\overline{E}$.

for all $i, j \in \{1, \ldots, m-r\}$ such that $i \leqslant j$, which is not the case. What is true, however, is that

$$\begin{aligned} G(x) \succeq 0 \quad &\Leftrightarrow \quad \forall i \in \{1, \ldots, m-r\},\ \lambda_i(G(x)) = u_i(G(x))^\top G(x) u_i(G(x)) \geqslant 0, \\ &\Leftrightarrow \quad \forall \overline{E} \in \mathcal{E}_r(G(\overline{x})),\ \forall E \doteq [e_1, \ldots, e_{m-r}] \approx \overline{E} \text{ s.t. } E^\top E = \mathbb{I}_{m-r}, \\ &\qquad\ \forall i \in \{1, \ldots, m-r\},\ e_i^\top G(x) e_i \geqslant 0. \end{aligned} \tag{5.2}$$

for every x close enough to $\overline{x}$, because although $u_i(G(x))$ is not a continuous function of x, every accumulation point $\tilde{E}$ of $[u_1(G(x)), \ldots, u_{m-r}(G(x))]$ as $x \to \overline{x}$ must belong to $\mathcal{E}_r(G(\overline{x}))$. Intuitively speaking, (5.2) can be seen as a *semi-infinite* model of Ω in a neighborhood of $\overline{x}$. This motivates the following characterization of Nondegeneracy:

Proposition 5.1.2 (Proposition 3.2 of [13]). *Let $\overline{x} \in \Omega$ and denote $r \doteq \operatorname{rank}(G(\overline{x}))$. Then, $\overline{x}$ satisfies Nondegeneracy if, and only if, for every $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ the family of vectors*

$$\{v_{ii}(\overline{x}, \overline{E})\}_{i \in \{1, \ldots, m-r\}}$$

is linearly independent.

Proof. Let us assume that $r < m$ since the result follows trivially otherwise. If there exists some $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ such that $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in \{1, \ldots, m-r\}}$ is linearly dependent, then the larger family $\left\{v_{ij}(\overline{x}, \overline{E}) \colon 1 \leqslant i \leqslant j \leqslant m-r\right\}$ is also linearly dependent, whence follows that $\overline{x}$ violates Nondegeneracy.

Conversely, let $B \in \mathcal{E}_r(G(\overline{x}))$ be arbitrary and consider the following linear operator:

$$\psi_{\overline{x}} \colon d \mapsto B^\top DG(\overline{x})[d] B.$$

It is straightforward (and it is a classical result, see [37, Section 3.4.4]) to verify that $\overline{x}$ satisfies Nondegeneracy if, and only if $\psi_{\overline{x}}$ is surjective or, equivalently, that its adjoint $\psi_{\overline{x}}^*$ given by

$$\psi_{\overline{x}}^* \colon \mu \mapsto \left[\left\langle B^\top D_{x_\ell} G(\overline{x}) B,\ \mu \right\rangle\right]_{\ell \in \{1, \ldots, n\}}$$

is injective. Then, let $\mu \in \mathbb{S}^{m-r}$ be such that $\psi_{\overline{x}}^*[\mu] = 0$, and let $C \in \mathbb{R}^{m-r \times m-r}$ be an orthogonal matrix that diagonalizes μ; that is, $C^\top \mu C = \operatorname{Diag}(y_1, \ldots, y_{m-r})$. Then, it follows that

$$\begin{aligned} 0 &= \psi_{\overline{x}}^* \left[C \operatorname{Diag}(y_1, \ldots, y_{m-r}) C^\top\right] \\ &= \left[\left\langle B^\top D_{x_\ell} G(\overline{x}) B,\ C \operatorname{Diag}(y_1, \ldots, y_{m-r}) C^\top \right\rangle\right]_{\ell \in \{1, \ldots, n\}} \\ &= \left[\left\langle (BC)^\top D_{x_\ell} G(\overline{x}) BC,\ \operatorname{Diag}(y_1, \ldots, y_{m-r}) \right\rangle\right]_{\ell \in \{1, \ldots, n\}}. \end{aligned} \tag{5.3}$$

Set $\overline{E} \doteq BC$, which belongs to $\mathcal{E}_r(G(\overline{x}))$ since C is nonsingular, and denote its i-th column by $\overline{e}_i$. From (5.3), we obtain that

$$0 = \sum_{i=1}^{m-r} y_i v_{ii}(\overline{x}, \overline{E}) = \left[\sum_{i=1}^{m} \left(\overline{e}_i^\top D_{x_\ell} G(\overline{x}) \overline{e}_i\right) y_i\right]_{\ell \in \{1, \ldots, n\}} \tag{5.4}$$

and by hypothesis it follows that $y_i = 0$ for all $i \in \{1, \ldots, m-r\}$; hence, $\mu = 0$. □

Naturally, an analogous characterization can be derived for Robinson's CQ as well, which ends up in a very similar format to how Mangasarian-Fromovitz' CQ is usually presented in the NLP literature.

Proposition 5.1.3 (Proposition 5.1 of [13]). *Let $\overline{x} \in \Omega$ and $r \doteq \text{rank}(G(\overline{x}))$. Then, $\overline{x}$ satisfies Robinson's CQ if, and only if, for every $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ the family*

$$\{v_{ii}(\overline{x}, \overline{E})\}_{i \in \{1,\dots,m-r\}}$$

is positively linearly independent.

Proof. If $r = m$ there is nothing to prove, so let us assume otherwise. For any fixed $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$, recall that $\left\{v_{ii}(\overline{x}, \overline{E})\right\}_{i \in \{1,\dots,m-r\}}$ is positively linearly independent if, and only if, the following holds: if the scalars $\alpha_1 \geqslant 0, \dots, \alpha_{m-r} \geqslant 0$ satisfy

$$\sum_{i=1}^{m-r} \alpha_i DG(\overline{x})^*[\overline{e}_i \overline{e}_i^\top] = \sum_{i=1}^{m-r} \alpha_i \left[\overline{e}_i^\top D_{x_j} G(\overline{x}) \overline{e}_i\right]_{j \in \{1,\dots,n\}} = 0, \tag{5.5}$$

then one must have $\alpha_1 = \ldots = \alpha_{m-r} = 0$. That is, the set of our interest is positively linearly independent if, and only if, for every matrix μ satisfying

$$-\mu \doteq \sum_{i=1}^{m-r} \alpha_i \overline{e}_i \overline{e}_i^\top = \overline{E} \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_{m-r} \end{bmatrix} \overline{E}^\top \tag{5.6}$$

with $\alpha_1 \geqslant 0, \dots, \alpha_{m-r} \geqslant 0$, we have that

$$DG(\overline{x})^*[\mu] = 0 \ \Rightarrow \ \mu = 0. \tag{5.7}$$

With this in mind, let us assume that $\left\{v_{ii}(\overline{x}, \overline{E})\right\}_{i \in \{1,\dots,m-r\}}$ is positively linearly independent for all $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$. Now let μ be any matrix such that $DG(\overline{x})^*[\mu] = 0$, $\langle G(\overline{x}), \mu \rangle = 0$ and $-\mu \succeq 0$. Because there always exists some matrix $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ such that $-\mu$ has the form (5.6) – for instance, any matrix that simultaneously diagonalizes $G(\overline{x})$ and μ – it follows from our hypothesis that $\mu = 0$ and since μ is arbitrary, Robinson's CQ holds. Conversely, assume that Robinson's CQ holds and let $\overline{E}$ and $\alpha_1 \geqslant 0, \dots, \alpha_{m-r} \geqslant 0$ be such that (5.5) holds. Then, define μ as in (5.6), which implies $-\mu \succeq 0$ and $\langle G(\overline{x}), \mu \rangle = 0$, so it follows from Robinson's CQ that $\mu = 0$ and $\alpha_1 = \ldots = \alpha_{m-r} = 0$. Thus, $\left\{v_{ii}(\overline{x}, \overline{E}) \colon i \in \{1, \dots, m-r\}\right\}$ is positively linearly independent, for every $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$. □

These simple characterizations compose the cornerstone of this chapter and the papers that inspired it [12, 13] together with their extension for NSOCP [15].

5.1.1 Sequential constant rank conditions for NSDP

Relaxing the characterization of Proposition 5.1.2 leads to the following:

Definition 5.1.1 (Definition 4.2 of [12]). Let $\overline{x} \in \Omega$ and $r \doteq \text{rank}(G(\overline{x}))$. We say that $\overline{x}$ satisfies the *Sequential CRCQ* (Seq-CRCQ) condition if for each matrix $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ there is a neighborhood $\mathcal{V}$ of $(\overline{x}, \overline{E})$ such that, for every $J \subseteq \{1, \dots, m-r\}$, the rank of $\{v_{ii}(x, E)\}_{i \in J}$ remains constant for every $(x, E) \in \mathcal{V}$ such that $E^\top E = \mathbb{I}_{m-r}$.

The relationship between Seq-CRCQ and the "facial" CRCQ of the previous chapter (Definition 4.2.1) is not yet fully understood. At best, we know that Definition 4.2.1 reduced to the context of NSDP with a particular choice of reduction mapping does not imply Seq-CRCQ [14, Example 5.1] but nothing else is known[4]. Anyway, it is important to mention that they have different purposes. While Definition 4.2.1 easily induces the Strong Second-Order Optimality Condition, we were not able to obtain such a result under Seq-CRCQ. Instead, Seq-CRCQ is meant to aid in the study of global convergence of algorithms, which in turn is not in the known range of applications of Definition 4.2.1 due to its intrinsically geometric nature. To prove our previous statements on Seq-CRCQ, we will first show two useful characterizations of it, which also explain the prefix "sequential".

Theorem 5.1.4 (Proposition 4.2 of [12]). *Let $\overline{x} \in \Omega$ and $r \doteq \operatorname{rank}(G(\overline{x}))$. The following statements are equivalent:*

1. *Seq-CRCQ holds at $\overline{x}$;*
2. *For each $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ there exists some neighborhood $\mathcal{V}$ of $(\overline{x}, \overline{E})$ such that, for every $J \subseteq \{1, \ldots, m-r\}$: if $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is linearly dependent, then $\{v_{ii}(x, E)\}_{i \in J}$ is also linearly dependent for every $(x, E) \in \mathcal{V}$ such that $E^\top E = \mathbb{I}_{m-r}$;*
3. *Either $r = m$ or for all sequences $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$ and $\{\Delta^k\}_{k \in \mathbb{N}} \subseteq \mathbb{S}^m$ with $\Delta^k \to 0$, there exists some $I \subseteq_\infty \mathbb{N}$ and a sequence $\{E^k\}_{k \in I} \to \overline{E}$, such that $E^k \in \mathcal{E}_r(G(x^k) + \Delta^k)$ for every $k \in I$ and, for every subset $J \subseteq \{1, \ldots, m-r\}$: if the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is linearly dependent, then $\{v_{ii}(x^k, E^k)\}_{i \in J}$ remains linearly dependent, for all $k \in I$ large enough.*

In Theorem 5.1.4, item 2 is based on a clever reasoning first presented by Qi and Wei [80] regarding the NLP version of CRCQ, and item 3 makes its algorithmic uses explicit. Moreover, in [80] the authors present a relaxation of CRCQ for NLP inspired by Robinson's CQ, which they called the *Constant Positive Linear Dependence* (CPLD) condition. At the time, however, they could not prove that CPLD was indeed a Constraint Qualification. This was only accomplished by Andreani et al. in [21]. Joining our characterization of Robinson's CQ (Proposition 5.1.3) and Theorem 5.1.4 together yields an extension of CPLD for NSDP:

Definition 5.1.2 (Definition 4.2 of [12]). Let $\overline{x} \in \Omega$ and $r \doteq \operatorname{rank}(G(\overline{x}))$. The point $\overline{x}$ satisfies the *Sequential CPLD* (Seq-CPLD) condition if for each matrix $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ there exists a neighborhood $\mathcal{V}$ of $(\overline{x}, \overline{E})$ such that, for every $J \subseteq \{1, \ldots, m-r\}$: if $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is positively linearly dependent, then $\{v_{ii}(x, E)\}_{i \in J}$ is linearly dependent for every $(x, E) \in \mathcal{V}$ such that $E^\top E = \mathbb{I}_{m-r}$.

This condition also admits a characterization similar to the one of Theorem 5.1.4, item 3, for Seq-CRCQ. It goes as follows:

Theorem 5.1.5 (Proposition 4.2 of [12]). *Let $\overline{x} \in \Omega$ and $r \doteq \operatorname{rank}(G(\overline{x}))$. The point $\overline{x}$ satisfies Seq-CPLD if, and only if, either $r = m$ or for all sequences $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$ and $\{\Delta^k\}_{k \in \mathbb{N}} \subseteq \mathbb{S}^m$ with $\Delta^k \to 0$, there exists some $I \subseteq_\infty \mathbb{N}$ and a sequence $\{E^k\}_{k \in I} \to \overline{E}$, such that $E^k \in \mathcal{E}_r(G(x^k) + \Delta^k)$ for every $k \in I$ and, for every subset $J \subseteq \{1, \ldots, m-r\}$: if the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is linearly dependent, then $\{v_{ii}(x^k, E^k)\}_{i \in J}$ remains linearly dependent, for all $k \in I$ large enough.*

Recall from Chapter 3 that each sequential optimality condition has an associate class of algorithms that are supported by them, being the largest one the AKKT class. Therefore, proving that Seq-CPLD (or Seq-CRCQ) is sufficient for AKKT and KKT to be equivalent, is the easiest path for proving global convergence of every algorithm of the AKKT class to stationary points, all at once. Moreover, because every local minimizer of (NSDP) is AKKT, this also proves that Seq-CPLD (and, by extension, Seq-CRCQ) is indeed a genuine Constraint Qualification condition.

[4] *Personal note:* I conjecture they are completely independent.

Theorem 5.1.6 (Theorem 4.2 of [12]). *Let $\overline{x} \in \Omega$ be an AKKT point that satisfies Seq-CPLD. Then, $\overline{x}$ is also a KKT point.*

Proof. Let $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, $\{\mu^k\}_{k\in\mathbb{N}} \subseteq \mathbb{S}^m_-$, and $\{\tilde{\Delta}^k\}_{k\in\mathbb{N}} \to 0$ be the AKKT sequences from Definition 3.1.1. Observe that $\lambda_i(G(x^k)) > 0$ for every $i \in \{m-r+1,\dots,m\}$ for every large k, where r is the rank of $G(\overline{x})$, then $\lambda_i(G(x^k)+\tilde{\Delta}^k) > 0$ and $\lambda_i(\mu^k) = 0$ for every such i and all k large enough because $\langle \mu^k, G(x^k)+\tilde{\Delta}^k\rangle = 0$. Hence, the spectral decomposition of μ^k can be represented in the format

$$\mu^k = \sum_{i=1}^{m-r} \lambda_i(\mu^k) u_i^k (u_i^k)^\top$$

where $u_i^k \doteq u_i(G(x^k)+\tilde{\Delta}^k) = u_i(\mu^k)$, $i \in \{1,\dots,m-r\}$, are shared orthonormal eigenvectors between μ^k and $G(x^k)+\tilde{\Delta}^k$, associated with the $m-r$ smallest eigenvalues of both μ^k and $G(x^k)+\tilde{\Delta}^k$, respectively. Defining $E^k \doteq [u_1^k,\dots,u_{m-r}^k]$ for every k, we obtain

$$\begin{aligned}
\nabla_x L(x^k,\mu^k) &= \nabla f(x^k) + \sum_{i=1}^{m-r} \lambda_i(\mu^k) DG(x^k)^* \left[u_i^k (u_i^k)^\top\right] \\
&= \nabla f(x^k) + \sum_{i=1}^{m-r} \lambda_i(\mu^k) v_{ii}(x^k, E^k) \\
&\to 0.
\end{aligned}$$

For each $k \in \mathbb{N}$, let $P^k \in \mathbb{R}^{m\times r}$ be a matrix whose columns are orthonormal eigenvectors associated with the r largest eigenvalues of $G(x^k)$, construct

$$D^k \doteq \left[\begin{array}{cccc|ccc}
\|x^k-\overline{x}\| & & & & & & \\
 & 2\|x^k-\overline{x}\| & & & & & \\
 & & \ddots & & & 0 & \\
 & & & (m-r)\|x^k-\overline{x}\| & & & \\
\hline
 & & & & \lambda_{m-r+1}(G(x^k)) & & \\
 & 0 & & & & \ddots & \\
 & & & & & & \lambda_m(G(x^k))
\end{array}\right],$$

and define $M^k \doteq U^k D^k (U^k)^\top$, where $U^k \doteq [E^k, P^k]$ for every $k \in \mathbb{N}$. Note that $M^k \to G(\overline{x})$ and that the $m-r$ smallest eigenvalues of M^k are simple, if $x^k \neq \overline{x}$, meaning their (normalized) associate eigenvectors are unique up to sign, when k is large enough. Consequently, $v_{ii}(x^k, E^k)$ is invariant to the choice of $E^k \in \mathcal{E}_r(M^k)$, for all such k, and every $i \in \{1,\dots,m-r\}$.

Using Carathéodory's Lemma 2.1.5 for the family $\{v_{ii}(x^k,E^k)\}_{i\in\{1,\dots,m-r\}}$, for each fixed $k \in I$, we obtain some $J^k \subseteq \{1,\dots,m-r\}$ such that $\{v_{ii}(x^k,E^k)\}_{i\in J^k}$ is linearly independent and

$$\nabla f(x^k) + \sum_{i=1}^{m-r} \lambda_i(\mu^k) v_{ii}(x^k,E^k) = \nabla f(x^k) + \sum_{i\in J^k} \tilde{\alpha}_i^k v_{ii}(x^k,E^k), \tag{5.8}$$

where $\tilde{\alpha}_i^k \leqslant 0$ for every $k \in I$ and every $i \in J^k$. By the infinite pigeonhole principle, we can assume J^k is the same, say equal to J, for all $k \in I$ large enough. We claim that the sequences $\{\tilde{\alpha}_i^k\}_{k\in I}$ are all bounded. In order to prove this, suppose that

$$m^k \doteq \max_{i\in J}\{\tilde{\alpha}_i^k\}$$

is unbounded with $k \in I$, divide (5.8) by m^k and note that if $m^k \to \infty$ on a subsequence, then the vectors $v_{ii}(\overline{x}, \overline{E})$, $i \in J$, would be positively linearly dependent. On the other hand, by construction the vectors $v_{ii}(x^k, E^k)$, with $i \in J$, are linearly independent for all large k, which contradicts Seq-CPLD. Finally, note that every collection of limit points $\{\overline{\alpha}_i \colon i \in J\}$ of their respective sequences $\{\tilde{\alpha}_i^k\}_{k\in\mathbb{N}}, i \in J$, generates a Lagrange multiplier associated with $\overline{x}$, which is $\overline{\mu} \doteq \sum_{i\in J} \overline{\alpha}_i u_i(G(\overline{x}))$. Thus, $\overline{x}$ is KKT. □

Now, to emphasize our previous comments, let us formalize them.

Corollary 5.1.7. *Every local minimizer of (NSDP) that satisfies Seq-CPLD (or Seq-CRCQ) must also satisfy the KKT conditions.*

Corollary 5.1.8. *Every feasible accumulation point of the Augmented Lagrangian algorithm (Algorithm 1) or the SQP method (Algorithm 2) that satisfies Seq-CPLD (or Seq-CRCQ) must satisfy KKT as well.*

The first corollary is also a consequence of Theorem 3.1.1 and the second one follows from Theorem 3.1.2 and Proposition 3.1.3. However, it is worth recalling that these results may be extended far beyond Algorithms 1 and 2 – see also Section 3.3.2. It is immediate from Theorem 5.1.1 that Nondegeneracy implies Seq-CRCQ, and [12, Example 4.2] shows that this implication is strict. Moreover, [12, Examples 3.2 and 4.3] prove that Seq-CRCQ and Robinson's CQ are not related with each other. It is also easy to see that Robinson's CQ implies Seq-CPLD, but [12, Example 4.3] shows that Seq-CPLD does not imply Robinson's CQ. At last, it is important to remark that Seq-CPLD is also strictly implied by Seq-CRCQ because the latter does not necessarily imply Robinson's CQ.

Remark 1. In contrast with Definition 4.2.1, neither Seq-CRCQ nor Seq-CPLD are defined in terms of reductions mappings.

5.2 Continuity of spectral decompositions and weak CQs

The reader may be aware of the fact that when a NLP problem with constraints

$$g_1(x) \geqslant 0, \dots, g_m(x) \geqslant 0 \tag{5.9}$$

is modelled in terms of (NSDP) by means of a structurally diagonal matrix constraint; that is, something in the form

$$G(x) \doteq \operatorname{Diag}(g_1(x), \dots, g_m(x)) \doteq \begin{bmatrix} g_1(x) & & \\ & \ddots & \\ & & g_m(x) \end{bmatrix} \succeq 0, \tag{5.10}$$

Nondegeneracy never holds at any point $\overline{x} \in \Omega$ should the dimension of $\operatorname{Ker}(G(\overline{x}))$ be greater than one, regardless of the fulfilment of LICQ for the NLP model (at $\overline{x}$). To illustrate this, suppose that $G(\overline{x}) = 0$ and $m > 1$, take $\overline{E} = \mathbb{I}_m$, and apply Theorem 5.1.1. Note that $v_{ij}(\overline{x}, \overline{E}) = 0$

whenever $i \neq j$ and thus Nondegeneracy fails. It is even possible to say that Nondegeneracy may become meaningless in the presence of "structural zeros", in general[5].

One could say this is not a surprise, for even the LICQ condition from NLP is ruined by identically zero constraints, say $x \mapsto 0$. However, while it is straightforward to deal with such constraints in NLP, this is not necessarily true in NSDP. To investigate this issue in detail, let us take a closer look at our characterization of Nondegeneracy (Proposition 5.1.2) and observe that if LICQ holds for (5.9) at some $\overline{x}$, then $\{v_{ii}(\overline{x}, \overline{E})\}_{i\in\{1,\ldots,m-r\}}$ defined for (5.10) with $\overline{E} \doteq [e_i]_{i\in\mathcal{A}(\overline{x})}$ is linearly independent, where e_i is the i-th vector of the canonical basis of $\mathbb{R}^m$ and $r \doteq \text{rank}(G(\overline{x})) = \mathcal{A}(\overline{x})$. With this in mind, we ask a very simple question: why *all* $\overline{E}$ instead of *some* $\overline{E}$? We answer this question by specifying *which* $\overline{E}$:

Definition 5.2.1 (Definition 3.2 of [13]). Let $\overline{x} \in \Omega$ and $r \doteq \text{rank}(G(\overline{x}))$. We say that $\overline{x}$ satisfies *Weak-Nondegeneracy* if either $r = m$ or, for each sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there exists some

$$\overline{E} \in \limsup_{k\in\mathbb{N}} \mathcal{E}_r(G(x^k))$$

such that the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i\in\{1,\ldots,m-r\}}$ is linearly independent.

This relaxed version of Nondegeneracy only takes into account certain matrices $\overline{E}$ whose columns are limits of eigenvector sequences. In very informal language, it only considers certain "points of continuity of $\mathcal{E}_r(G(x))$ by paths." As its definition suggests, Weak-Nondegeneracy has the following property:

Proposition 5.2.1 (Remark 2 of [13]). *Let $G\colon x \mapsto \text{Diag}(g_1(x), \ldots, g_m(x))$. Then, $\overline{x} \in \Omega$ satisfies Weak-Nondegeneracy if, and only if, the family $\{\nabla g_i(\overline{x})\}_{i\in\mathcal{A}(\overline{x})}$ is linearly independent.*

That is, when G is structurally diagonal, Weak-Nondegeneracy reduces to the standard LICQ condition from NLP. Consequently, any diagonal model of a NLP problem at a point satisfying LICQ with $r < m - 1$ suffices to show that Weak-Nondegeneracy does not imply Nondegeneracy (for instance, [13, Example 3.1]). But because $\limsup_{k\in\mathbb{N}} \mathcal{E}_r(G(x^k)) \subseteq \mathcal{E}_r(G(\overline{x}))$ for every $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, we see that Weak-Nondegeneracy is implied by Nondegeneracy.

Despite being significantly weaker than standard Nondegeneracy, the Weak-Nondegeneracy condition still guarantees existence of Lagrange multipliers at local minimizers of (NSDP) and, moreover, it supports the convergence theory of the classical external penalty method for NSDP, which is a particular instance of Algorithm 1 obtained by setting $\tilde{\mu}^k \doteq 0$ for every $k \in \mathbb{N}$ – see [13, Theorem 3.2].

Now, a natural question is: what happens if we restrict other Constraint Qualifications to the so-called "points of continuity of $\mathcal{E}_r(G(x))$" as in Weak-Nondegeneracy? This motivates us to take a second look at Robinson's CQ, Seq-CRCQ, and Seq-CPLD, resulting in the following:

Definition 5.2.2 (Definition 3.2 of [12]). Let $\overline{x} \in \Omega$ and let r be the rank of $G(\overline{x})$. We say that $\overline{x}$ satisfies the:

- *Weak-Robinson's CQ* for NSDP if for each $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ there exists some

$$\overline{E} \in \limsup_{k\in\mathbb{N}} \mathcal{E}_r(G(x^k))$$

  such that the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i\in\{1,\ldots,m-r\}}$ is positively linearly independent.

[5] This fact may be one the reasons why this approach is usually called in informal conversations as "the wrong way of modelling", and the "right" way consists of viewing NLP constraints as multiple unidimensional NSDP constraints, that is, $g_1(x) \succeq 0, \ldots, g_m(x) \succeq 0$.

- *Weak-CRCQ* condition for NSDP (respectively, Weak-CPLD) if either $r = m$ or, for each sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there exists some

$$\overline{E} \in \limsup_{k\in\mathbb{N}} \mathcal{E}_r(G(x^k))$$

  such that, for every subset $J \subseteq \{1, \ldots, m-r\}$: if the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i\in J}$ is (positively) linearly dependent, then $\{v_{ii}(x^k, E^k)\}_{i\in J}$ remains linearly dependent, for all $k \in I$ large enough. Here, $I \subseteq_\infty \mathbb{N}$, and $\{E^k\}_{k\in I}$ is a sequence converging to $\overline{E}$ and such that

$$\forall k \in I,\ E^k \in \mathcal{E}_r(G(x^k)),$$

  which is part of the Painlevé-Kuratowski outer limit definition.

It is clear that Weak-Nondegeneracy (respectively, Weak-Robinson's CQ) implies Weak-CRCQ (respectively, Weak-CPLD) and it is possible to construct an example [12, Example 3.3] that proves that the converse is not true. This example also shows, together with [12, Example 3.2], that Weak-CRCQ is independent of Robinson's CQ. Further, it is immediate to see that Weak-CRCQ (respectively, Weak-CPLD) is also strictly implied by Seq-CRCQ (respectively, Seq-CPLD) by taking $\Delta^k \doteq 0$ for every $k \in \mathbb{N}$ and checking [12, Example 4.1]. Nevertheless, Weak-CRCQ and Weak-CPLD still enjoy the defining property of Weak-Nondegeneracy, namely it recovers the original NLP version of CRCQ and CPLD, respectively, when G is structurally diagonal:

Proposition 5.2.2 (Proposition 3.1 of [12]). *Let $G\colon x \mapsto \mathrm{Diag}(g_1(x), \ldots, g_m(x))$. Then, Weak-CRCQ (respectively, Weak-CPLD) holds at $\overline{x} \in \Omega$ if, and only if, for every $J \subseteq \mathcal{A}(\overline{x})$: if $\{\nabla g_i(\overline{x})\}_{i\in J}$ is (positively) linearly dependent, then $\{\nabla g_i(x)\}_{i\in J}$ is also linearly dependent for every x close enough to $\overline{x}$.*

Moreover, Weak-CRCQ and Weak-CPLD can be used to build global convergence of a general external penalty method, which implies together with Theorem 3.1.1 that they ensure existence of Lagrange multipliers at all local minimizers of (NSDP).

Theorem 5.2.3 (Theorem 3.1 of [12]). *Let $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$ and $\{x^k\}_{k\in\mathbb{N}} \to \overline{x} \in \Omega$ be such that*

$$\nabla_x L\left(x^k,\ \rho_k \Pi_{\mathbb{S}^m_-}(G(x^k))\right) \to 0.$$

If $\overline{x}$ satisfies Weak-CRCQ, then $\overline{x}$ satisfies the KKT conditions. In particular, every local minimizer of (NSDP) that satisfies Weak-CPLD (or Weak-CRCQ) also satisfies KKT.

Proof. Analogous to the proof of Theorem 5.1.6. □

However, we still do not know whether the result above can be extended to methods other than the external penalty or not. In fact, we do not know which properties of Seq-CRCQ and Seq-CPLD are inherited by Weak-CRCQ and Weak-CPLD. The potential of such "weak" Constraint Qualifications is not well-understood yet and this is a challenging problem that we have left open. The relation between Weak-Robinson's CQ and Robinson's CQ is also partially unknown.

5.3 Dealing with a general sparsity structure

Weak-Nondegeneracy is specialized for dealing with problems that present some diagonal structured sparsity or even a block-diagonal structure [13, Remark 3.2]. But we also proposed in [13] a parallel approach for treating general sparsity that was based on a regularity condition introduced by Forsgren [47, Sect. 2.3]. Before presenting it, let us introduce some notation: given a point $\overline{x}$ and a matrix-valued function $\mathcal{F}\colon \mathbb{R}^n \to \mathbb{S}^\beta$, consider the set $\mathcal{S}(\mathcal{F},\overline{x})$ defined as follows:

$$\begin{aligned}\mathcal{S}(\mathcal{F},\overline{x}) &\doteq \left\{M \in \mathbb{S}^\beta \colon M_{ij} = 0 \text{ if } \mathcal{F}_{ij}(x) \text{ is structurally zero near } \overline{x}\right\}\\ &= \left\{M \in \mathbb{S}^\beta \colon M_{ij} = 0 \text{ if } \exists \varepsilon > 0 \text{ such that } \mathcal{F}_{ij}(x) = 0, \forall x \in B(\overline{x},\varepsilon)\right\}.\end{aligned}$$

For example, if $\beta = 3$ and for all x close to $\overline{x}$, we are able to identify non trivial mappings $\mathcal{F}_{ij}$ such that

$$\mathcal{F}(x) = \begin{bmatrix}\mathcal{F}_{11}(x) & 0 & \mathcal{F}_{13}(x)\\ 0 & \mathcal{F}_{22}(x) & 0\\ \mathcal{F}_{13}(x) & 0 & \mathcal{F}_{33}(x)\end{bmatrix}, \text{ then } M \in \mathcal{S}(\mathcal{F},\overline{x}) \Leftrightarrow M = \begin{bmatrix}M_{11} & 0 & M_{13}\\ 0 & M_{22} & 0\\ M_{13} & 0 & M_{33}\end{bmatrix},$$

where M_{11}, M_{13}, M_{22}, and M_{33} may or may not be zero. Also, we define

$$\mathcal{I}(\mathcal{F},\overline{x}) \doteq \{(i,j)\colon \forall \varepsilon > 0, \exists x \in B(\overline{x},\varepsilon) \text{ such that } \mathcal{F}_{ij}(x) \neq 0,\ 1 \leqslant i \leqslant j \leqslant \beta\}$$

as the set of indices that define the elements of $\mathcal{S}(\mathcal{F},\overline{x})$. For any matrix $\overline{E}$ that spans Ker$(G(\overline{x}))$, consider the function

$$\widehat{G}^{\overline{E}}(x) \doteq \overline{E}^\top G(x)\overline{E}$$

and note that $\nabla \widehat{G}^{\overline{E}}_{ij}(\overline{x}) = v_{ij}(\overline{x},\overline{E})$ for all $i,j \in \{1,\ldots,m-r\}$ with $i \leqslant j$. Our strategy consists of simply incorporating this notation into Nondegeneracy as characterized in Theorem 5.1.1 to introduce a new Constraint Qualification.

Definition 5.3.1 (Definition 4.2 of [13]). We say that *Sparse-Nondegeneracy* holds at $\overline{x} \in \Omega$ when either Ker$(G(\overline{x})) = \{0\}$ or there exists a matrix $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ such that:

1. The family $\left\{v_{ij}(\overline{x},\overline{E})\colon (i,j) \in \mathcal{I}(\widehat{G}^{\overline{E}},\overline{x}),\ 1 \leqslant i \leqslant j \leqslant m-r\right\}$ is linearly independent;
2. $(i,i) \in \mathcal{I}(\widehat{G}^{\overline{E}},\overline{x})$ for all $i \in \{1,\ldots,m-r\}$.

Should item 2 of Definition 5.3.1 be violated, there is always an equivalently reformulated problem that fulfils it.

Proposition 5.3.1 (Remark 4.2 of [13]). *Let $\overline{x} \in \Omega$ and $r \doteq$ rank$(G(\overline{x}))$. Also, let $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ be such that the set*

$$J(\overline{E}) \doteq \{i \in \{1,\ldots,m-r\}\colon (i,i) \notin \mathcal{I}(\widehat{G}^{\overline{E}},\overline{x})\}$$

is nonempty and maximal among all $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$. Then, there is a neighborhood $\mathcal{V}$ of $\overline{x}$ and an orthogonal matrix $U \doteq [U_1,U_2] \in \mathbb{R}^{m\times(m-\omega)+\omega}$, possibly depending on $\overline{E}$, such that

$$G(x) \in \mathbb{S}^m_+ \quad \Leftrightarrow \quad \begin{array}{l}U_1^\top G(x)U_1 \in \mathbb{S}^{m-\omega}_+\\ U_2^\top G(x) = 0\end{array}$$

for all $x \in \mathcal{V}$ and $(i,i) \in \mathcal{I}(U_1^\top G(x)U_1,\overline{x})$ for all $i \in \{1,\ldots,m-\omega\}$, where $\omega \doteq |J(\overline{E})|$.

Proof. Let $\overline{e}_1,\ldots,\overline{e}_{m-r}$ denote the columns of $\overline{E}$ and it follows directly from the definitions of $J(\overline{E})$ and $\mathcal{I}(\widehat{G}^{\overline{E}},\overline{x})$ that there exists some neighborhood $\mathcal{V}$ of $\overline{x}$ such that

$$G(x) \in \mathbb{S}^m_+ \quad \Leftrightarrow \quad G(x) \in \mathbb{S}^m_+ \bigcap_{i\in J(\overline{E})} \{\overline{e}_i\overline{e}_i^\top\}^\perp \doteq F \trianglelefteq \mathbb{S}^m_+$$

for every $x \in \mathcal{V}$. Because F is a face of $\mathbb{S}^m_+$, there is an orthogonal matrix $U \in \mathbb{R}^{m\times m}$ such that

$$U^\top F U = \left\{ \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix} : M \in \mathbb{S}^{m-\omega}_+ \right\},$$

where ω is the cardinality of $J(\overline{E})$ [74, Equation 2.3]. Let us partition $U \doteq [U_1, U_2]$ with $U_1 \in \mathbb{R}^{m\times m-\omega}$ and $U_2 \in \mathbb{R}^{m\times\omega}$ and note that $G(x) \in F$ if, and only if, $U_2^\top G(x) = 0$ and $U_1^\top G(x) U_1 \in \mathbb{S}^{m-\omega}_+$. Observe that $(i,i) \in \mathcal{I}(U_1^\top G(x)U_1, \overline{x})$ for every $i \in \{1,\ldots,m-\omega\}$; indeed, if there was any i stating otherwise, we could denote by u_i the i-th column of U_1 and note that $\tilde{E} \doteq [u_i, U_2] \in \mathcal{E}_r(G(\overline{x}))$ which would contradict the maximality of $J(\overline{E})$ since $|J(\tilde{E})| = \omega+1 > |J(\overline{E})|$. □

Proposition 5.3.1 is clearly another type of nonlinear facial reduction, similarly to Section 4.2.2. In fact, note that if $J(\overline{E})$ is not maximal, then the construction of Proposition 5.3.1 can be repeated iteratively until a minimal face is reached. Thus, every problem can be equivalently reformulated (reducing dimension if necessary), such that item 2 of Definition 5.3.1 always holds. In particular, when G is an affine function, then this procedure can be computed via standard facial reduction – see Section 2.2.2.

We now prove that Sparse-Nondegeneracy guarantees uniqueness of the Lagrange multiplier with respect to a fixed sparsity pattern, which is a property inherited from standard Nondegeneracy.

Proposition 5.3.2 (Lemma 4.2 and Proposition 4.4 of [13]). *Let $\overline{x} \in \Omega$ and let $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$, where $r \doteq$ rank$(G(\overline{x}))$. Then:*

1. *Item 1 of Definition 5.3.1 holds at $\overline{x}$ with respect to $\overline{E}$ if, and only if, there is no nonzero $\tilde{\mu} \in \mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})$ such that $DG(\overline{x})^*[\overline{E}\tilde{\mu}\overline{E}^\top] = 0$;*
2. *If $\overline{x}$ satisfies KKT and item 1 of Definition 5.3.1 holds at $\overline{x}$ with respect to $\overline{E}$, then $\Lambda(\overline{x}) \cap \overline{E}\mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})\overline{E}^\top$ is a singleton.*

Proof. The first item follows directly by noticing that

$$\sum_{(i,j)\in\mathcal{I}(\widehat{G}^{\overline{E}},\overline{x})} v_{ij}(\overline{x},\overline{E})\tilde{\mu}_{ij} = DG(\overline{x})^*[\overline{E}\tilde{\mu}\overline{E}^\top] \tag{5.11}$$

for every $\tilde{\mu} \in \mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})$. Now, for the second item, let $\mu_1, \mu_2 \in \Lambda(\overline{x}) \cap \overline{E}\mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})\overline{E}^\top$ be Lagrange multipliers associated with $\overline{x}$, define $\mu \doteq \mu_1 - \mu_2$, and by definition there exists some $\tilde{\mu} \in \mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})$ such that $\mu = \overline{E}\tilde{\mu}\overline{E}^\top$ and $DG(\overline{x})^*[\overline{E}\tilde{\mu}\overline{E}^\top] = 0$. By the previous item we must have $\tilde{\mu} = 0$ and, consequently, $\mu_1 = \mu_2$. □

Another important property of Sparse-Nondegeneracy is that the number of structural zeros of $\widehat{G}^{\overline{E}}$, at points that satisfy it, remains the same regardless of $\overline{E}$.

Proposition 5.3.3 (Proposition 4.5 of [13]). *Let $\overline{x} \in \Omega$ be such that* Ker$G(\overline{x}) \neq \{0\}$*, and let $\overline{E}, \overline{W} \in \mathcal{E}_r(G(\overline{x}))$ be such that item 1 of Definition 5.3.1 holds. Then,*

$$\#\mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x}) = \#\mathcal{I}(\widehat{G}^{\overline{W}}, \overline{x}).$$

Proposition 5.3.3 tells us that the strength of Sparse-Nondegeneracy is invariant with respect to $\overline{E}$. That is, if there are multiple matrices $\overline{E}$ certifying Sparse-Nondegeneracy at a point $\overline{x}$, then they all induce similar conditions.

Next, we prove that Sparse-Nondegeneracy implies Robinson's CQ, which also shows that it is indeed a Constraint Qualification.

Proposition 5.3.4. *If $\overline{x} \in \Omega$ fulfils Sparse-Nondegeneracy, then it fulfils Robinson's CQ.*

Proof. Let $r \doteq \text{rank}(G(\overline{x}))$ and note that the result follows trivially when $r = m$, so let us assume that $r < m$. Suppose that Sparse-Nondegeneracy holds at $\overline{x}$ and take any $\eta \succeq 0$ such that $\langle \eta, G(\overline{x}) \rangle = 0$ and $DG(\overline{x})^*[\eta] = 0$, then there exists some $\mu \in \mathbb{S}^{m-r}_+$ such that $\eta = \overline{E}\mu\overline{E}^\top$. Define the matrix $\tilde{\mu} \in \mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})$ whose (i,j)-th entry is given by

$$\tilde{\mu}_{ij} \doteq \begin{cases} \mu_{ij}, & \text{if } (i,j) \in \mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x}) \\ 0, & \text{otherwise,} \end{cases}$$

and note that

$$DG(\overline{x})^*[\eta] = DG(\overline{x})^*[\overline{E}\mu\overline{E}^\top] = DG(\overline{x})^*[\overline{E}\tilde{\mu}\overline{E}^\top] = 0, \tag{5.12}$$

so $\tilde{\mu} = 0$ due to Proposition 5.3.2, item 1. Moreover, from item 2 of Definition 5.3.1, $(i,i) \in \mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x})$ for all $i \in \{1, \ldots, m-r\}$, so the diagonal of μ must consist only of zeros, which implies that $\mu = 0$ and, consequently, $\eta = 0$. Since η is arbitrary, Robinson's CQ holds. □

As for the converse statement, it can be analysed by means of a result similar to Proposition 5.2.1 which states that when Sparse-Nondegeneracy is applied to a structurally diagonal problem (5.10), it becomes equivalent to LICQ for its NLP reformulation.

Proposition 5.3.5. *Let $G \colon x \mapsto \text{Diag}(g_1(x), \ldots, g_m(x))$. Then, $\overline{x} \in \Omega$ satisfies condition Sparse-Nondegeneracy if, and only if, the family $\{\nabla g_i(\overline{x})\}_{i \in \mathcal{A}(\overline{x})}$ is linearly independent.*

Consequently, it is easy to build a diagonal counterexample that proves that Sparse-Nondegeneracy is not implied by Robinson's CQ. For instance, take $m = 2$ and set $\overline{x} = 0$; then, define the constraint

$$G(x) \doteq \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix}, \tag{5.13}$$

and note that $v_{11}(\overline{x}, \overline{E}) = v_{22}(\overline{x}, \overline{E}) = 1$ for every matrix $\overline{E}$ that spans $\text{Ker}(G(\overline{x}))$. Hence, Sparse-Nondegeneracy does not hold, although Robinson's CQ does.

Finally, as it may be interesting for some potential applications, we will prove that Sparse-Nondegeneracy is stable (or robust), in the sense it is maintained in the neighborhood of points that satisfy it.

Theorem 5.3.6. *Let $\overline{x} \in \Omega$ satisfy Sparse-Nondegeneracy. Then, there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that every $x \in \mathcal{V}$ satisfies Sparse-Nondegeneracy.*

Proof. Suppose that the statement above is false. That is, suppose that there exists a feasible sequence $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$ such that Sparse-Nondegeneracy fails at each x^k, but it holds at $\overline{x}$. Our aim is to prove that this leads to an absurd. So let $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ and, for each $k \in \mathbb{N}$ let Π^k be the projection matrix onto the space spanned by the eigenvectors associated with the $m-r$ smallest eigenvalues of $G(x^k)$, which is well defined when k is sufficiently large. Define $\tilde{W}^k \doteq \Pi^k \overline{E}$, for all such $k \in \mathbb{N}$. It is well-known (see, for instance, [37, Example 3.98]) that the columns of $\tilde{W}^k$ are linearly independent, which allows us to apply the Gram-Schmidt orthonormalization process to them and arrange its output in the columns of a new matrix, which we will denote by W^k. It is also known that $W^k \to \overline{E}$ as $k \to \infty$.

Because Sparse-Nondegeneracy fails at x^k, we know that the rank r^k of $G(x^k)$ is smaller than m, and by the pigeonhole principle we can even assume that r^k is the same, say $\tilde{r}$, for every $k \in \mathbb{N}$. Also, note that $m - \tilde{r} \leqslant m - r$ and that, by construction, we can assume that the first $m - \tilde{r}$ columns of each W^k, which we will arrange in a matrix denoted by E^k, span

Ker$(G(x^k))$. Since Sparse-Nondegeneracy fails at x^k it holds that $\{v_{ij}(x^k, E^k)\}_{(i,j)\in\mathcal{I}(\widehat{G}^{E^k}, x^k)}$ is linearly dependent for each $k \in \mathbb{N}$. Observe that since $\lim_{k\to\infty} E^k$ is a submatrix of $\overline{E}$ we have that

$$\limsup_{k\to\infty} \mathcal{I}(\widehat{G}^{E^k}, x^k) \subseteq \mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x})$$

and

$$\limsup_{k\to\infty} \{v_{ij}(x^k, E^k)\}_{(i,j)\in\mathcal{I}(\widehat{G}^{E^k}, x^k)} \subseteq \{v_{ij}(\overline{x}, \overline{E})\}_{(i,j)\in\mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x})}.$$

The left side of the expression above is linearly dependent, which makes $\{v_{ij}(\overline{x}, \overline{E})\}_{(i,j)\in\mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x})}$ linearly dependent as well. Because $\overline{E}$ is arbitrary, it follows that Sparse-Nondegeneracy fails at $\overline{x}$, which is a contradiction. □

5.4 A weak second-order condition under a weak stability assumption

The *Strong Second-Order Optimality Condition* (SSOC, see Definition 2.3.5) has one disadvantage: it is very unlikely to be useful for the convergence analysis of numerical methods. Gould and Toint [52] presented a simple example, that consists of a NLP problem with a quadratic objective function and a box-constraint, for which a large class of barrier-type methods may produce an output sequence whose limit points fail to satisfy SSOC (and also BSOC as in Definition 2.3.5) even when every iterate of the sequence satisfies the Second-Order Sufficient Condition for its respective penalized problem. Later, Andreani and Secchin [26] made a small modification in Gould and Toint's counterexample to obtain the same conclusion for a class of Augmented Lagrangian-type algorithms. Following this line of research, Andreani et al. [17] attempted to characterize the weakest second-order Constraint Qualification that could guarantee the fulfilment of SSOC on the limit points of a larger class of penalization-type algorithms that encompasses, for instance, the two we just mentioned. However, such a Constraint Qualification was proven not to imply nor be implied by LICQ [17, Examples 4.5 and 4.6], and to be violated even for box-constraints. Moreover, it is known [72] that even checking whether SSOC holds or not at a given point with respect to a given Lagrange multiplier is an NP-hard problem. No surprise, since checking SSOC means evaluating the sign of a quadratic form over a cone.

The classical practical alternative to such cone-based second-order conditions is the WSOC condition (again, see Definition 2.3.5). In NLP, Augmented Lagrangian methods [4] and a Regularized SQP method [49] were proven to satisfy WSOC at their limit points under MFCQ and a *Weak Constant Rank* property; and a *Curvilinear Search Interior-Point method* [71] was also proven to converge to points satisfying WSOC under LICQ and strict complementarity. This suggests, of course, that the best approach to prove global convergence of algorithms to WSOC-type points is to rely on constant rank. But it seems we got ourselves a dilemma: we saw in Chapter 4 a constant rank condition (Definition 4.2.1) that is associated with SSOC but, as far as we know, it is not related to any algorithm. On the other hand, we saw in the previous sections of this chapter another constant rank condition (Definition 5.1.1) that supports global convergence of a large class of methods, but as far as we know, is not related to any second-order optimality condition. How to proceed, then?

5.4.1 The best of both worlds: A weak notion of constant rank with second-order properties and algorithmic implications

Let $M \succeq 0$, $r \doteq \text{rank}(M)$, and let $\overline{E} \in \mathcal{E}_r(M)$ be arbitrary. We will now recall the general form of the tool we employed to prove Theorem 5.3.6 which induces a classical reduction mapping that can be found, for instance, in Bonnans and Shapiro [37, Example 3.98]. In a sufficiently small neighborhood $\mathcal{N}$ of M, we consider the function $\mathcal{E}_{\overline{E}} \colon \mathcal{N} \to \mathbb{R}^{m\times m-r}$ given by

$$\mathcal{E}_{\overline{E}}(N) \doteq \text{gramschmidt}\left(\Pi(N)\overline{E}\right), \tag{5.14}$$

for every $N \in \mathcal{N}$, where $\Pi(N)$ denotes the orthogonal projection matrix onto the space spanned by $u_1(N), \dots, u_{m-r}(N)$ and $\text{gramschmidt}(\Pi(N)\overline{E})$ denotes the output of the *Gram-Schmidt orthonormalization* procedure after being applied to the columns of $\Pi(N)\overline{E}$.

Lemma 5.4.1. *For any given $M \succeq 0$ whose rank is denoted by r and any $\overline{E} \in \mathcal{E}_r(M)$ it holds that:*

1. *$\mathcal{E}_{\overline{E}}$ is well-defined and analytic provided $\mathcal{N}$ is small enough;*
2. *$\mathcal{E}_{\overline{E}}(N)^\top \mathcal{E}_{\overline{E}}(N) = \mathbb{I}_{m-r}$ and $\text{Im}(\mathcal{E}_{\overline{E}}(N)) = \text{span}(\{u_1(N), \dots, u_{m-r}(N)\})$, for all $N \in \mathcal{N}$;*
3. *$\mathcal{E}_{\overline{E}}(M) = \overline{E}$.*

Proof. For item 1, observe that $N \mapsto \Pi(N)$ is an analytic function of N in a sufficiently small neighborhood, say $\mathcal{N}$, of M (see, for example, [61, Theorem 1.8]), then $N \mapsto \Pi(N)\overline{E}$ is also analytic in $\mathcal{N}$ and, moreover, $\Pi(M)\overline{E} = \overline{E}$. Shrinking $\mathcal{N}$ if necessary, we have that for all $N \in \mathcal{N}$, the rank of $\Pi(N)\overline{E}$ is equal to the rank of $\Pi(M)\overline{E} = \overline{E}$, meaning that the $m-r$ columns of $\Pi(N)\overline{E}$ are linearly independent for every $N \in \mathcal{N}$; as a consequence, the function $N \mapsto \mathcal{E}_{\overline{E}}(N) \doteq \text{gramschmidt}\left(\Pi(N)\overline{E}\right)$ is well-defined and also analytic in (a possibly smaller) $\mathcal{N}$.

Regarding item 2, note that $\mathcal{E}_{\overline{E}}(N)^\top \mathcal{E}_{\overline{E}}(N) = \mathbb{I}_{m-r}$ due to the Gram-Schmidt procedure, and it follows from the linear independence of the columns of $\Pi(N)\overline{E}$ that $\text{Im}(\mathcal{E}_{\overline{E}}(N)) = \text{span}(\{u_1(N), \dots, u_{m-r}(N)\})$ whenever $N \in \mathcal{N}$. Finally, observe that $\mathcal{E}_{\overline{E}}(M) = \Pi(M)\overline{E} = \overline{E}$ which proves item 3. □

Notice, however, that the columns of $\mathcal{E}_{\overline{E}}(N)$ are not necessarily eigenvectors of $N \in \mathcal{N}$.

Remark 2. If $M = 0$, then for every orthogonal matrix $\overline{E} \in \mathbb{R}^{m\times m}$ it holds that $\mathcal{E}_{\overline{E}}(N) = \overline{E}$ for every $N \in \mathbb{S}^m$. Indeed, in this case we have $\text{span}(\{u_1(N), \dots, u_m(N)\}) = \mathbb{R}^m$ for every $N \in \mathbb{S}^m$, and since $\overline{E}$ is itself orthogonal, it follows that

$$\mathcal{E}_{\overline{E}}(N) = \text{gramschmidt}\left(\Pi(N)\overline{E}\right) = \text{gramschmidt}\left(\overline{E}\right) = \overline{E}$$

for every $N \in \mathbb{S}^m$.

Now, let $\overline{x} \in \Omega$, denote the rank of $G(\overline{x})$ by r, and let $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ be arbitrary. Let $\mathcal{E}_{\overline{E}}$ be constructed as in (5.14) around $M = G(\overline{x})$ and because G is continuous, there is a neighborhood $\mathcal{V}$ of $\overline{x}$ such that the function:

$$\mathcal{U}(x) \doteq \mathcal{E}_{\overline{E}}(G(x))$$

is well defined. For simplicity of notation we shall omit the subscript $\overline{E}$ from this point forth, unless $\overline{E}$ is not clear from the context. Then, consider the vector:

$$\overline{v}_{ij}(x) \doteq \begin{bmatrix} \overline{u}_i(x)^\top D_{x_1}G(x)\overline{u}_j(x) \\ \vdots \\ \overline{u}_i(x)^\top D_{x_n}G(x)\overline{u}_j(x) \end{bmatrix}$$

for every $i,j \in \{1,\ldots,m-r\}$ such that $1 \leqslant i \leqslant j \leqslant m-r$, where $\overline{u}_1(x),\ldots,\overline{u}_{m-r}(x) \in \mathbb{R}^m$ denote the columns of $\mathcal{U}(x)$. We define an extension of the so-called in [6] *Weak Constant Rank* property as follows:

Definition 5.4.1 (Definition 5 of [48]). Let $\overline{x}$ be a feasible point of (NSDP) and $r \doteq \operatorname{rank}(G(\overline{x}))$. We say that $\overline{x}$ satisfies the *Weak Constant Rank* (WCR) property when there exists some matrix $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ and a neighborhood $\mathcal{V}$ of $\overline{x}$ such that the family

$$\{\overline{v}_{ij}(x) \colon 1 \leqslant i \leqslant j \leqslant m-r\}$$

remains with the same rank for every $x \in \mathcal{V}$.

Also, in the following lemma we prove that WCR as in Definition 5.4.1 is equivalent to the inner semicontinuity at $\overline{x}$ of the mapping

$$x \mapsto S(x,\overline{x}) \doteq \{d \in \mathbb{R}^n \colon \mathcal{U}(x)^\top DG(x)[d]\mathcal{U}(x) = 0\},$$

which will be called the *perturbed critical subspace* at x centered at $\overline{x}$. In particular, note that $S(\overline{x},\overline{x}) = \operatorname{lin}(C_T(\overline{x}))$.

Lemma 5.4.2 (Lemma 4 of [48]). *A feasible point $\overline{x}$ satisfies WCR if, and only if, the set-valued mapping $x \mapsto S(x,\overline{x})$ is inner semicontinuous at $\overline{x}$.*

Proof. First, we shall prove that, for every $x \in \mathbb{R}^n$,

$$S(x,\overline{x}) = \left\{d \in \mathbb{R}^n \colon \overline{v}_{ij}(x)^\top d = 0,\ 1 \leqslant i \leqslant j \leqslant m-r\right\}. \tag{5.15}$$

Note that for each $\ell \in \{1,\ldots,n\}$ we have

$$\mathcal{U}(x)^\top D_{x_\ell}G(x)\mathcal{U}(x) = \begin{bmatrix} \overline{u}_1(x)^\top D_{x_\ell}G(x)\overline{u}_1(x) & \cdots & \overline{u}_1(x)^\top D_{x_\ell}G(x)\overline{u}_{m-r}(x) \\ \vdots & \ddots & \vdots \\ \overline{u}_{m-r}(x)^\top D_{x_\ell}G(x)\overline{u}_1(x) & \cdots & \overline{u}_{m-r}(x)^\top D_{x_\ell}G(x)\overline{u}_{m-r}(x) \end{bmatrix}$$

and considering that $\overline{u}_j(x)^\top D_{x_\ell}G(x)\overline{u}_i(x) = \overline{u}_i(x)^\top D_{x_\ell}G(x)\overline{u}_j(x) = (\overline{v}_{ij}(x))_\ell$ for every i,j,ℓ, we obtain

$$\mathcal{U}(x)^\top DG(x)[d]\mathcal{U}(x) = \sum_{\ell=1}^{n} \mathcal{U}(x)^\top D_{x_\ell}G(x)\mathcal{U}(x)d_\ell = \begin{bmatrix} \overline{v}_{11}(x)^\top d & \cdots & \overline{v}_{1m-r}(x)^\top d \\ \vdots & \ddots & \vdots \\ \overline{v}_{1m-r}(x)^\top d & \cdots & \overline{v}_{m-r,m-r}(x)^\top d \end{bmatrix}$$

whence follows (5.15).

Now, because $D_{x_\ell}G$ and $\mathcal{U}$ are continuous, $\overline{v}_{ij}$ is also continuous, then a result from Facchinei and Pang [45, Proposition 3.2.9] tells us that WCR is equivalent to the outer semicontinuity of the mapping $x \mapsto S(x,\overline{x})^\circ$ at $\overline{x}$, where

$$S(x,\overline{x})^\circ = \operatorname{span}(\{\overline{v}_{ij}(x) \colon 1 \leqslant i \leqslant j \leqslant m-r\})$$

using the characterization in (5.15). Then, the desired result follows from [27, Theorem 1.1.8], which states that the inner semicontinuity of a set-valued mapping at a given point is equivalent to the outer semicontinuity of its polar at that point. □

Clearly, WCR as in Definition 5.4.1 is implied by Nondegeneracy, in view of Theorem 5.1.1. Also, when G is structurally diagonal as in (5.10), Definition 5.4.1 recovers the NLP definition of WCR for the constraint $\mathrm{diag}(G(x)) \geqslant 0$. In [22, Example 5.2] the authors exhibit an NLP problem with a feasible point that satisfies both Mangasarian-Fromovitz' CQ and WCR, but not LICQ, and the above discussion tells us that it can be used again to prove that Nondegeneracy is strictly stronger than the joint condition "Robinson's CQ + WCR."

5.4.2 Differentiability properties of projections

In order to achieve second-order optimality conditions under "Robinson's CQ + WCR," we will resort to a second-order version of Theorem 3.1.1 which, as we will see later on, involves the computation of the second-order derivative of the feasibility function

$$\mathcal{F}(x) \doteq \|\Pi_{\mathbb{S}^m_+}(-G(x))\|^2.$$

However, although the projection operator is convex, it is not necessarily differentiable at the boundary of $\mathbb{S}^m_+$. Let us denote the set where $\Pi_{\mathbb{S}^m_+}$ is twice differentiable by $\mathcal{D}$. The so-called *Clarke generalized Jacobian* of $\Pi_{\mathbb{S}^m_+}$ at a point $Y \in \mathbb{S}^m$ is the convex hull of the set of all limiting derivatives of $\Pi_{\mathbb{S}^m_+}$ at Y, denoted by

$$\partial\Pi_{\mathbb{S}^m_+}(Y) \doteq \mathrm{conv}\left(\left\{V \colon \exists\{Y^k\}_{k\in\mathbb{N}} \subset \mathcal{D},\ Y^k \to Y,\ D\Pi_{\mathbb{S}^m_+}(Y^k) \to V\right\}\right). \tag{5.16}$$

For every $Y \in \mathbb{S}^m$, the set $\partial\Pi_{\mathbb{S}^m_+}(Y)$ is convex, compact, and it is a singleton when $Y \in \mathcal{D}$. Also, as an abuse of notation we will define the *generalized Hessian* of $\mathcal{F}$ at $x \in \mathbb{R}^n$ as

$$\partial^2\mathcal{F}(x) \doteq \partial\nabla\mathcal{F}(x),$$

which is the convex hull of the set of all limiting Hessian matrices of $\mathcal{F}$ at x, defined analogously to (5.16). Now, following Páles and Zeidan [73], we can characterize $\partial^2\mathcal{F}(x)$ as follows:

Theorem 5.4.3. *For every $x, d \in \mathbb{R}^n$ and every $M \in \partial^2\mathcal{F}(x)$, it holds that:*

$$d^\top Md \in -d^\top D^2G(x)^*[\Pi_{\mathbb{S}^m_+}(-G(x))]d + \left\langle DG(x)[d],\ \partial\Pi_{\mathbb{S}^m_+}(-G(x))[DG(x)[d]]\right\rangle$$

where $\partial\Pi_{\mathbb{S}^m_+}(-G(x))[DG(x)[d]] \doteq \{V[DG(x)[d]] \colon V \in \Pi_{\mathbb{S}^m_+}(-G(x))\}$.

The final fundamental piece of our approach is the characterization by Sun [86] of the Clarke generalized Jacobian of the projection onto the semidefinite cone. To make a proper reference, we define the following matrix:

$$\mathcal{B}(\lambda(M)) \doteq \left[\frac{\max\{\lambda_i(M), 0\} + \max\{\lambda_j(M), 0\}}{|\lambda_i(M)| + |\lambda_j(M)|}\right]_{i,j\in\{1,\ldots,m\}},$$

where $0/0$ is set as 1. Next, we make a slightly adapted transcription of a proposition by Qi [79, Proposition 2.5] summarizing Sun's result (see also [48, Corollary 1]):

Proposition 5.4.4. *Let $M \in \mathbb{S}^m$ and let α, β and γ be the sets of indices of the positive, zero and negative eigenvalues of M, respectively. Without loss of generality, assume those three* *blocks are separated and let* $U \doteq [u_i(M)]_{i\in\alpha\cup\beta\cup\gamma} \doteq [U_\alpha, U_\beta, U_\gamma]$*. Then, for any* $V \in \partial\Pi_{\mathbb{S}^m_+}(-M)$ *there exists a* $V_{|\beta|} \in \partial\Pi_{\mathbb{S}^{|\beta|}_+}(0)$ *such that*

$$V[H] = U \begin{bmatrix} 0 & 0 & U_\alpha^\top H U_\gamma \odot \mathcal{B}(\lambda(M))_{\alpha\gamma} \\ 0 & V_{|\beta|}[U_\beta^\top H U_\beta] & U_\beta^\top H U_\gamma \\ U_\gamma^\top H U_\alpha \odot \mathcal{B}(\lambda(M))_{\alpha\gamma}^\top & U_\gamma^\top H U_\beta & U_\gamma^\top H U_\gamma \end{bmatrix} U^\top \tag{5.17}$$

for every $H \in \mathbb{S}^m$*, where* $\odot$ *denotes the entry-wise (Hadamard) product. Conversely, for every* $V_{|\beta|} \in \partial\Pi_{\mathbb{S}^{|\beta|}_+}(0)$*, there exists some* $V \in \partial\Pi_{\mathbb{S}^m_+}(-M)$ *such that* (5.17) *holds.*

Even though we assume the eigenvalues are separated by sign, the ordering inside each partition is not relevant. Following Hiriart-Urruty et al. [55], the second-order necessary optimality condition for unconstrained minimizers of $\mathcal{F}$ when it is differentiable, is the following:

Theorem 5.4.5. *If* $\overline{x} \in \mathbb{R}^n$ *is a local minimizer of a differentiable function* $F\colon \mathbb{R}^n \to \mathbb{R}$ *such that* ∇F *is locally Lipschitz, then* $\nabla F(\overline{x}) = 0$*, and for each* $d \in \mathbb{R}^n$*, there exists some* $M \in \partial^2 F(\overline{x})$ *such that* $d^\top M d \geqslant 0$*. In other words,*

$$\forall d \in \mathbb{R}^n, \quad \sup_{M\in\partial^2 F(\overline{x})} d^\top M d \geq 0.$$

We refer to [55, Theorem 3.1] for a proof. As observed by Hiriart-Urruty et al., it is not true that $d^\top M d \geqslant 0$ for all $M \in \partial^2 F(\overline{x})$, in general, and not even this holds for some fixed M and all d.

5.4.3 WSOC under WCR and Robinson's CQ

We are now ready to proceed to the main result of this section:

Theorem 5.4.6 (Theorem 11 of [48]). *If* $\overline{x} \in \Omega$ *is a local minimizer such that Robinson's CQ and the WCR property hold, then there exists some Lagrange multiplier* $\overline{\mu} \in \Lambda(\overline{x})$ *satisfying WSOC.*

Proof. If $\overline{x}$ is a local minimizer of (NSDP), a straightforward adaptation of Theorem 3.1.1 tells us that for any given $\{\rho_k\}_{k\in\mathbb{N}} \to +\infty$, there is some sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ such that, for every $k \in \mathbb{N}$, x^k is a local minimizer of the penalty function:

$$\begin{aligned} F_k(x) &\doteq f(x) + \frac{\rho_k}{2}\|\Pi_{\mathbb{S}^m_-}(G(x))\|^2 + \frac{1}{4}\|x-\overline{x}\|^4. \\ &= f(x) + \frac{\rho_k}{2}\|\Pi_{\mathbb{S}^m_+}(-G(x))\|^2 + \frac{1}{4}\|x-\overline{x}\|^4. \end{aligned}$$

Hence, it satisfies the first-order stationarity condition

$$\nabla F(x^k) = \nabla f(x^k) + \|x^k - \overline{x}\|^2(x^k - \overline{x}) - DG(x^k)^*[\rho_k \Pi_{\mathbb{S}^m_+}(-G(x^k))] = 0$$

and due to Theorems 5.4.3 and 5.4.5, for each $d \in \mathbb{R}^n$ there exists some $V \in \partial\Pi_{\mathbb{S}^m_+}(-G(x^k))$, such that

$$\begin{aligned} d^\top \nabla^2 F(x^k) d = d^\top \Big(\nabla^2 f(x^k) - D^2 G(x^k)^*[\rho_k \Pi_{\mathbb{S}^m_+}(-G(x^k))]\Big)\, d \;+ \\ + \rho_k \left\langle DG(x^k)[d],\ V[DG(x^k)[d]]\right\rangle + d^\top \Delta^k d \geqslant\ 0 \end{aligned} \tag{5.18}$$

where $\Delta^k \doteq \|x^k - \overline{x}\|^2 \mathbb{I}_n + 2(x^k - \overline{x})(x^k - \overline{x})^\top$ and $\nabla^2 F(x^k)$ denotes the element of $\partial^2 F_k(x^k)$ that is defined in terms of V, as an abuse of notation.

Under Robinson's CQ, the sequence $\{\mu^k\}_{k\in\mathbb{N}}$ whose k-th term is defined by

$$\mu^k \doteq -\rho_k \Pi_{\mathbb{S}^m_+}(-G(x^k))$$

is bounded, implying it has a convergent subsequence, which we will consider to be itself from now on, without loss of generality. That is we assume that $\{\mu^k\}_{k\in\mathbb{N}} \to \overline{\mu}$ for some $\overline{\mu} \in \mathbb{S}^m_-$. By the discussion that follows Theorem 3.1.1, we see that $\overline{\mu} \in \Lambda(\overline{x})$.

Now, let $d \in \operatorname{lin}(C_T(\overline{x})) = S(\overline{x}, \overline{x})$. By WCR there is a sequence $\{d^k\}_{k\in\mathbb{N}} \to d$ such that $d^k \in S(x^k, \overline{x})$ for every $k \in \mathbb{N}$, and with (5.18) in mind, we see that for all such d^k there exists some V^k such that

$$(d^k)^\top \nabla^2_x L(x^k, \mu^k) d^k + \rho_k \left\langle DG(x^k)[d^k],\ V^k \left[DG(x^k)[d^k]\right] \right\rangle \geqslant -\delta^k, \tag{5.19}$$

where $\delta^k \doteq (d^k)^\top \Delta^k d^k \to 0$. The following paragraphs prove that (5.19) implies

$$d^\top \nabla^2_x L(\overline{x}, \overline{\mu}) d + 2 \left\langle DG(\overline{x})[d],\ \overline{\mu} DG(\overline{x})[d] G(\overline{x})^\dagger \right\rangle \geqslant 0, \tag{5.20}$$

which is enough to complete the proof since

$$2 \left\langle \overline{\mu},\ DG(\overline{x})[d] G(\overline{x})^\dagger DG(\overline{x})[d] \right\rangle = \sigma(\overline{\mu},\ T^2_{\mathbb{S}^m_+}(G(\overline{x}), DG(\overline{x})[d])),$$

for every $d \in \mathbb{R}^n$ due to a celebrated characterization by Shapiro [83]. To complete that task, we analyse the behaviour of the sequence $\{\rho_k \langle DG(x^k)[d^k],\ V^k[DG(x^k)[d^k]]\rangle\}_{k\in\mathbb{N}}$ in distinct cases.

Recall that, by construction, the columns of $\mathcal{U}(x^k)$ span the eigenspace associated with the $m-r$ smallest eigenvalues of $G(x^k)$, for all k sufficiently large, and denote the matrix that has in its columns eigenvectors associated with the $m-r$ smallest eigenvalues of $G(x^k)$ by E^k, and since $d^k \in S(x^k, \overline{x})$, we have

$$(E^k)^\top DG(x^k)[d^k] E^k = 0$$

for every k large enough. We proceed by analysing a few cases:

1. If $G(\overline{x}) \succ 0$, then $-G(x^k) \prec 0$ for k sufficiently large. For such k, since $\gamma(x^k) = \beta(x^k) = 0$ and $\alpha(x^k) = m$, we obtain that $V^k[DG(x^k)[d^k]] = 0$ from Proposition 5.4.4, which implies

$$\rho_k \langle DG(x^k)[d^k],\ V^k[DG(x^k)[d^k]]\rangle = 0.$$

   Also, note that $\sigma(\overline{x}, \overline{\mu}) = 0$ because $\langle G(\overline{x}), \overline{\mu}\rangle = 0$ implies $\overline{\mu} = 0$ in this case;

2. If $G(\overline{x}) = 0$, then $G(\overline{x})^\dagger = 0$ and $\sigma(\overline{x}, \overline{\mu}) = 0$ as well. On the other hand, note that

$$\begin{aligned}\langle DG(x^k)[d^k],\ V^k[DG(x^k)[d^k]]\rangle &= \langle (E^k)^\top DG(x^k)[d^k] E^k,\ (E^k)^\top V^k[DG(x^k)[d^k]] E^k\rangle \\ &= 0,\end{aligned}$$

   because $m - r = m$ and E^k is orthogonal in this case;

3. If $G(\overline{x}) \succeq 0$, but $G(\overline{x}) \neq 0$, let $\alpha \doteq \alpha(x^k)$, $\beta \doteq \beta(x^k)$, and $\gamma \doteq \gamma(x^k)$ be the set of indices of the positive, zero and negative eigenvalues of $G(x^k)$, respectively. Further, we partition α into α_+ and α_-, where α_+ correspond to the positive eigenvalues of $G(x^k)$ that converge to positive eigenvalues of $G(\overline{x})$ and α_- to the ones that converge to zero. Then, consider the following spectral decomposition of $G(x^k)$:

$$G(x^k) = U^k \begin{bmatrix} \Phi_+^k & 0 & 0 & 0 \\ 0 & \Phi_-^k & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Psi^k \end{bmatrix} (U^k)^\top, \tag{5.21}$$

where we separate the eigenvalues of $G(x^k)$ primarily by their sign and, secondarily, by their limit points. For instance, $\Phi_+^k \in \mathbb{S}^{|\alpha_+|}$ corresponds to α_+ whereas $\Phi_-^k \in \mathbb{S}^{|\alpha_-|}$ corresponds to α_-. This partitioning is possible by the pigeonhole principle. The squared block of zeros in the diagonal of (5.21) is of dimension $|\beta|$ and $\Psi^k \in \mathbb{S}^{|\gamma|}$ contains the negative eigenvalues of $G(x^k)$. Also, $|\alpha_+| + |\alpha_-| + |\beta| + |\gamma| = m$. Recall that U^k simultaneously diagonalizes $G(x^k)$ and μ^k, by definition of μ^k. In order to simplify the notation, define

$$H^k \doteq DG(x^k)[d^k]$$

and

$$B^k_{\alpha\pm\gamma} \doteq (U^k_{\alpha\pm})^\top H U^k_\gamma \odot \mathcal{B}(\lambda(-G(x^k)))_{\alpha\pm\gamma}.$$

Using the characterization of V^k provided in (5.17) from Proposition 5.4.4, we obtain

$$V^k[H^k] = U^k \begin{bmatrix} 0 & 0 & 0 & B^k_{\alpha+\gamma} \\ 0 & 0 & 0 & B^k_{\alpha-\gamma} \\ 0 & 0 & V_{|\beta|}[(U^k_\beta)^\top H U^k_\beta] & (U^k_\beta)^\top H U^k_\gamma \\ (B^k_{\alpha+\gamma})^\top & (B^k_{\alpha-\gamma})^\top & (U^k_\gamma)^\top H U^k_\beta & (U^k_\gamma)^\top H U^k_\gamma \end{bmatrix} (U^k)^\top.$$

Because

$$\rho_k \langle H^k, V^k[H^k] \rangle = \langle (U^k)^\top H^k U^k,\ \rho_k (U^k)^\top V^k[H^k] U^k \rangle,$$

it is fundamental to note that for every $i \in \alpha_+$ and $j \in \gamma$,

$$\left(\rho_k \mathcal{B}(\lambda(-G(x^k)))\right)_{ij} = \frac{\rho_k \lambda_j(-G(x^k))}{\lambda_j(-G(x^k)) - \lambda_i(-G(x^k))} \to \frac{\lambda_j(\overline{\mu})}{\lambda_i(G(\overline{x}))}, \tag{5.22}$$

because $\rho_k \lambda_j(-G(x^k)) = \lambda_j(\mu^k)$ and $\lambda_j(-G(x^k)) \to 0$. Also, keep in mind that

$$\lambda_i(-G(x^k)) = -\lambda_i(G(x^k)).$$

The blocks indexed by α_-, β, and γ, are all blocks of zeros because if k is large enough, we must have $|\alpha_-| + |\beta| + |\gamma| = m - r$ and $U^k_{\alpha_- \cup \beta \cup \gamma} = E^k$, but since $d^k \in S(x^k, \overline{x})$ we have

$$(E^k)^\top H E^k = (E^k)^\top DG(x^k)[d^k] E^k = 0.$$

Thus, defining $\overline{\alpha}$ and $\overline{\gamma}$ as limits of $\alpha_+(x^k)$ and $\gamma(x^k)$ as $k \to \infty$, respectively, we obtain

$$\lim_{k\to\infty} \rho_k \langle H^k, V^k[H^k] \rangle = \left\langle H,\ U \begin{bmatrix} 0 & 0 & A \odot U_{\overline{\alpha}}^\top H U_{\overline{\gamma}} \\ 0 & 0 & 0 \\ A^\top \odot U_{\overline{\gamma}}^\top H U_{\overline{\alpha}} & 0 & 0 \end{bmatrix} U^\top \right\rangle \tag{5.23}$$

where $A \doteq \left[\lambda_j(\overline{\mu}) \lambda_i(G(\overline{x}))^{-1}\right]_{i\in\overline{\alpha}, j\in\overline{\gamma}}$ and U is any limit point of U^k and we recall that $\odot$ is an entry-wise product. On the other hand, considering the spectral decompositions

$$G(\overline{x}) = U \begin{bmatrix} \Phi & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} U^\top \quad \text{and} \quad \overline{\mu} = U \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Gamma \end{bmatrix} U^\top,$$

where $\mathbb{S}^{|\overline{\alpha}|} \ni \Phi \succ 0$ and $\mathbb{S}^{|\overline{\gamma}|} \ni \Gamma \preceq 0$ are diagonal matrices, let $\overline{\beta}$ be the limiting set of $\alpha_-(x^k) \cup \beta(x^k)$, denote $H \doteq DG(\overline{x})[d]$, and since $d \in S(\overline{x})$ we get

$$\begin{aligned} G(\overline{x})^\dagger H \overline{\mu} &= U \begin{bmatrix} \Phi^{-1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} U^\top U \begin{bmatrix} U_{\overline{\alpha}}^\top H U_{\overline{\alpha}} & U_{\overline{\alpha}}^\top H U_{\overline{\beta}} & U_{\overline{\alpha}}^\top H U_{\overline{\gamma}} \\ U_{\overline{\beta}}^\top H U_{\overline{\alpha}} & 0 & 0 \\ U_{\overline{\gamma}}^\top H U_{\overline{\alpha}} & 0 & 0 \end{bmatrix} U^\top U \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Gamma \end{bmatrix} U^\top \\ &= U \begin{bmatrix} 0 & 0 & \Phi^{-1} U_{\overline{\alpha}}^\top H U_{\overline{\gamma}} \Gamma \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} U^\top. \end{aligned}$$

Conveniently, it is elementary to notice that

$$\Phi^{-1} U_{\overline{\alpha}}^\top H U_{\overline{\gamma}} \Gamma = \left[\frac{\lambda_j(\overline{\mu})}{\lambda_i(G(\overline{x}))} U_i^\top H U_j \right]_{i \in \overline{\alpha}, j \in \overline{\gamma}} = A \odot U_{\overline{\alpha}}^\top H U_{\overline{\gamma}} \tag{5.24}$$

and thus Equation (5.23) tells us that

$$\lim_{k \to \infty} \rho_k \langle H^k, V^k[H^k] \rangle = 2\langle H, G(\overline{x})^\dagger H \overline{\mu} \rangle = 2\langle H G(\overline{x})^\dagger H, \overline{\mu} \rangle = \sigma(\overline{\mu},\, T^2_{\mathbb{S}^m_+}(G(\overline{x}), DG(\overline{x})[d]))$$

Consequently, (5.19) implies (5.20), which means $\overline{x}$ satisfies the WSOC with respect to the multiplier $\overline{\mu}$.

□

In the presence of Nondegeneracy $\Lambda(\overline{x})$ is a singleton and Theorem 5.4.6 recovers the classical result of [83], but even without assuming uniqueness of the Lagrange multiplier it ensures there will be at least one multiplier satisfying WSOC. Moreover, in contrast with [58, 64, 83], our proof does not require strict complementarity; but nevertheless, if it does hold, then the proof of Theorem 5.4.6 can be significantly simplified, since in this case the sequence $\{G(x^k)\}_{k \in \mathbb{N}}$ is nonsingular and we can avoid the use of subdifferentials.

5.5 Wrap up

In this chapter, we presented alternative extensions of the Constant Rank Constraint Qualification (CRCQ) and the Constant Positive Linear Dependence (CPLD) conditions, and we showed their applications towards the convergence of algorithms; namely all methods that generate AKKT sequences. Our main contribution comes from the fact the condition we used in our analyses, called Sequential-CPLD (Seq-CPLD), is strictly weaker than the usual Robinson's CQ. Nonetheless, because Seq-CPLD implies the so-called Metric Subregularity CQ, it induces computable error bounds in the vicinity of any point that fulfils it. Going a bit further, to overcome the inherent difficulty in proving second-order optimality results at limit points of output sequences of algorithms, we mixed the techniques of the previous chapters (namely, sequential conditions and reduction mappings) to derive a Weak Constant Rank (WCR) property that is not a Constraint Qualification on its own but can be paired with Robinson's CQ to induce the so-called

Weak Second-Order Optimality Condition (WSOC) at local minimizers. At first sight, a weaker second-order condition does not seem advantageous, but in light of several works showing strong evidences that suggest impossibility of a numerical method capable of generating sequences whose accumulation points satisfy a stronger second-order condition, it becomes the unique relevant condition for that matter at this time. It is worth mentioning that this summary concerns only the Semidefinite Programming half of our papers [12, 13, 48] but [48] also comprises Second-Order Cone Programming, which is also the main topic of [15].

Bibliography

[1] J. Abadie. On the Kuhn-Tucker Theorem. In Jean Abadie, editor, *Nonlinear Programming*, pages 21–36. John Wiley, New York, 1967. 25

[2] E. D. Andersen, C. Roos, and T. Terlaky. Notes on duality in second order and p-order cone optimization. 2002. 20

[3] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. On augmented Lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, 18(4):1286–1309, 2008. 49

[4] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Second-order negative-curvature methods for box-constrained and general constrained optimization. *Computational Optimization and Applications*, 45:209–236, 2010. 81

[5] R. Andreani, C. E. Echagüe, and M. L. Schuverdt. Constant-rank condition and second-order constraint qualification. *Journal of Optimization Theory and Applications*, 146(2):255–266, February 2010. 30

[6] R. Andreani, C. E. Echagüe, and M. L. Schuverdt. Constant-rank condition and second-order constraint qualification. *Journal of Optimization theory and Applications*, 146(2):255–266, 2010. 57, 83

[7] R. Andreani, E. H. Fukuda, G. Haeser, H. Ramírez C., D. O. Santos, P. J. S. Silva, and T. P. Silveira. Erratum to: New constraint qualifications and optimality conditions for second order cone programs. *To appear in Set-Valued and Variational Analysis*, 2021. 54

[8] R. Andreani, E. H. Fukuda, G. Haeser, D. O. Santos, and L. D. Secchin. On the use of Jordan algebras for improving global convergence of an augmented Lagrangian method in nonlinear semidefinite programming. *To appear in Computational Optimization and Applications*, 2021. 46

[9] R. Andreani, W. Gómez, G. Haeser, L. M. Mito, and A. Ramos. On optimality conditions for nonlinear conic programming. *To appear in Mathematics of Operations Research*, 2021. 1, 35, 36, 37, 38, 40, 41, 42, 43, 44, 45, 46, 51, 52

[10] R. Andreani, G. Haeser, and J. M. Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60(5):627–641, 2011. 1, 36, 37, 38

[11] R. Andreani, G. Haeser, L. M. Mito, H. Ramırez, D. O. Santos, and T. P. Silveira. Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming. *To appear in Optimization Letters*, 2020. 1, 2

[12] R. Andreani, G. Haeser, L. M. Mito, and H. Ramírez C. Sequential constant rank constraint qualifications for nonlinear semidefinite programming with applications. Technical report, 2020. Available at arXiv:2106.00775. 1, 3, 70, 72, 73, 74, 75, 76, 77, 89

[13] R. Andreani, G. Haeser, L. M. Mito, and H. Ramírez C. Weak notions of nondegeneracy in nonlinear semidefinite programming. Technical report, 2020. Available at arXiv:2012.14810. 1, 2, 3, 71, 72, 76, 77, 78, 79, 89

[14] R. Andreani, G. Haeser, L. M. Mito, H. Ramírez C., and T. P. Silveira. First- and second-order optimality conditions for second-order cone and semidefinite programming under a constant rank condition. Technical report, 2021. Available at arXiv:2107.04693. 1, 2, 55, 58, 64, 66, 67, 73

[15] R. Andreani, G. Haeser, L. M. Mito, H. Ramírez C., and T. P. Silveira. Global convergence of algorithms under constant rank conditions for nonlinear second-order cone programming. Technical report, 2021. Available at arXiv:2110.12015. 1, 3, 69, 72, 89

[16] R. Andreani, G. Haeser, L. M. Mito, A. Ramos, and L. D. Secchin. On the best achievable quality of limit points of augmented Lagrangian schemes. *To appear in Numerical Algorithms*, 2021. DOI: 10.1007/s11075-021-01212-8. 1, 2, 47, 48, 49, 50, 51

[17] R. Andreani, G. Haeser, A. Ramos, and P. J. S. Silva. A second-order sequential optimality condition associated to the convergence of algorithms. *IMA Journal of Numerical Analysis*, 37(4):1902–1929, 2017. 81

[18] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. Two new weak constraint qualifications and applications. *SIAM Journal on Optimization*, 22(3):1109–1135, 2012. 2, 41, 46, 59, 60, 62, 63, 67

[19] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva. A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization*, 26(1):96–110, 2016. 50

[20] R. Andreani, J. M. Martínez, A. Ramos, and P. J. S. Silva. Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Mathematics of Operations Research*, 43(3):693–717, 2018. 42

[21] R. Andreani, J. M. Martínez, and 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. 3, 73

[22] R. Andreani, J. M. Martínez, and M. L. Schuverdt. On second-order optimality conditions for nonlinear programming. *Optimization*, 56(5-6):529–542, 2007. 30, 84

[23] R. Andreani, J. M. Martínez, and B. F. Svaiter. A new sequencial optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization*, 20(6):3533–3554, 2010. 1

[24] R. Andreani, J. M. Martínez, and B. F. Svaiter. A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization*, 20:3533–3554, 2010. 45

[25] R. Andreani, A. Ramos, A. A. Ribeiro, L. D. Secchin, and A. R. Velazco. On the convergence of augmented Lagrangian strategies for nonlinear programming. *To appear in IMA Journal of Numerical Analysis*, 2021. 43

[26] 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. 81

[27] J. P. Aubin and H. Frankowska. *Set-Valued Analysis*. Birkhauser, 1990. 84

[28] A. Baccari. On the classical necessary second-order optimality conditions. *Journal of Optimization Theory and Applications*, 123(1):213–221, 2004. 30

[29] R. Behling, G. Haeser, A. Ramos, and D.S. Viana. On a conjecture in second-order optimality conditions. *Journal of Optimization Theory and Applications*, 176(3):625–633, 2018. Extended version: arXiv:1706.07833. 66

[30] A. Ben-Tal and J. Zowe. A unified theory of first and second order conditions for extremum problems in topological vector spaces. In *Optimality and stability in mathematical programming*, pages 39–76. Springer, 1982. 27

[31] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Nashua, NH,, 1999. 9

[32] E. Birgin and J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. SIAM Publications, 2014. 38

[33] E. G. Birgin, L. F. Bueno, and J. M. Martínez. Assessing the reliability of general-purpose inexact restoration methods. *Journal of Computational and Applied Mathematics*, 282:1–16, 2015. 47

[34] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet. Characterization of Lojasiewicz inequalities and applications: subgradient flows, talweg, convexity. *Transactions of the American Mathematical Society*, 362:3319–3363, 2010. 45

[35] J. F. Bonnans, R. Cominetti, and A. Shapiro. Sensitivity analysis of optimization problems under second order regular constraints. *Mathematics of Operations Research*, 23(4):806–831, 1998. 15

[36] J. F. Bonnans, R. Cominetti, and A. Shapiro. Second order optimality conditions based on parabolic second order tangent sets. *SIAM Journal on Optimization*, 9(2):466–492, 1999. 27, 28, 30, 64

[37] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer Verlag, New York, 2000. 13, 14, 15, 16, 24, 26, 31, 32, 64, 71, 80, 82

[38] J. M. Borwein and H. Wolkowicz. Facial reduction for a cone-convex programming problem. *Journal of the Australian Mathematical Society, Series A*, 30(3):369–380, 1981. 21

[39] A. Brondsted. *An Introduction to Convex Polytopes*. Springer Berlin, 1983. 10

[40] L. N. H. Bunt. Bijdrage tot de theorie der konvekse puntverzamelingen. Technical report, 1934. PhD Thesis, University of Groningen, Amsterdam. 11

[41] L. Chen and D. Goldfarb. Interior-point 2-penalty methods for nonlinear programming with strong global convergence properties. *Mathematical Programming*, 108:1026, 2006. 46

[42] R. Cominetti. Metric regularity, tangent sets, and second-order optimality conditions. *Applied Mathematics and Optimization*, 21(1):265–287, 1990. 15, 24

[43] R. Correa and H. Ramírez C. A global algorithm for nonlinear semidefinite programming. *SIAM Journal on Optimization*, 15(1):303–318, 2004. 40, 41

[44] J. A. de Loera, R. Hemmecke, and M. Köppe. *Algebraic and Geometric Ideas in the Theory of Discrete Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2012. 9

[45] F. Facchinei and J.S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume I*. Springer Series in Operations Research. Springer. New York, 2003. 83

[46] S. Fitzpatrick and R. R. Phelps. Differentiability of the metric projection in Hilbert space. *Transactions of the American Mathematical Society*, 270(2):483–501, 1982. 12

[47] A. Forsgren. Optimality conditions for nonconvex semidefinite programming. *Mathematical Programming*, 88(1):105–128, 2000. 3, 78

[48] E. H. Fukuda, G. Haeser, and L. M. Mito. On the weak second-order optimality condition for nonlinear semidefinite and second-order cone programming. Technical report, 2020. Available at Optimization Online. 1, 83, 84, 85, 89

[49] P. E. Gill, V. Kungurtsev, and D. P. Robinson. A stabilized SQP method: global convergence. *IMA Journal of Numerical Analysis*, 37(1):407–443, 2017. DOI: 10.1093/imanum/drw004. 46, 81

[50] P. E. Gill, V. Kungurtsev, and D. P. Robinson. A shifted primal-dual penalty-barrier method for nonlinear optimization. *SIAM Journal on Optimization*, 30(2):1067–1093, 2020. 47

[51] F. J. Gould and J. W. Tolle. Optimality conditions and constraint qualifications in Banach space. *Journal of Optimization Theory and Applications*, 15(6):667–684, 1975. 20, 25

[52] N. I. M. Gould and P. L. Toint. A note on the convergence of barrier algorithms to second-order necessary points. *Mathematical programming*, 85(2):433–438, 1999. 81

[53] M. Guignard. Generalized Kunh-Tucker conditions for mathematical programming in a Banach space. *SIAM Journal of Control*, 7:232–241, 1969. 19, 20

[54] J.-B. Hiriart-Urruty and C. Lamaréchal. *Fundamentals of Convex Analysis*. Springer Berlin Heidelberg, Berlin Heldelberg, 2001. 9, 12

[55] J.-B. Hiriart-Urruty, J.-J. Strodiot, and V. H. Nguyen. Generalized Hessian matrix and second-order optimality conditions for problems with C1,1 data. *Applied Mathematics and Optimization*, 11:43–56, 1984. 85

[56] A. F. Izmailov, M. V. Solodov, and E. I. Uskov. Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. *SIAM Journal on Optimization*, 22(4):1579–1606, 2012. 41

[57] R. Janin. Direction derivate of the marginal function in nonlinear programming. *Mathematical Programming Study*, 21:110–126, 1984. 2, 30, 53, 54, 57, 61, 66

[58] F. Jarre. Elementary optimality conditions for nonlinear SDPs. In *Handbook on Semidefinite, Conic and Polynomial Optimization*. International Series in Operations Research & Management Science, 2012. 88

[59] C. Kanzow and D. Steck. An example comparing the standard and safeguarded augmented Lagrangian methods. *Operations Research Letters*, 45(6):598–603, 2017. 49

[60] H. Karimi, J. Nutini, and M. Schmidt. Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition. In P. Frasconi, N. Landwehr, G. Manco, and J. Vreeken, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 795–811, Cham, 2016. Springer International Publishing. 45

[61] T. Kato. *Perturbation theory for linear operators*. Springer, Berlin, 1995. Reprint of the 1980 edition. 82

[62] H. Kawasaki. An envelope-like effect of infinitely many inequality constraints on second-order necessary conditions for minimization problems. *Mathematical Programming*, 41(1):73–96, 1988. 24

[63] A.Y. Kruger, L. Minchenko, and J.V. Outrata. On relaxing the Mangasarian-Fromovitz constraint qualification. *Positivity*, 18(1):171–189, 2013. 59

[64] B. F. Lourenço, E. H. Fukuda, and M. Fukushima. Optimality conditions for nonlinear semidefinite programming via squared slack variables. *Mathematical Programming*, 166:1–24, 2016. 88

[65] B. F. Lourenço, V. Roshchina, and J. Saunderson. Amenable cones are particularly nice. Technical report, 2020. Available at arXiv:2011.07745. 10, 11

[66] S. Lu. Implications of the constant rank constraint qualification. *Mathematical Programming*, 126:365–392, 2011. 62

[67] J. M. Martínez and B. F. Svaiter. A practical optimality condition without constraint qualifications for nonlinear programming. *Journal of Optimization Theory and Applications*, 118(1):117–133, 2003. 1, 43

[68] L. Minchenko and S. Stakhovski. On relaxed constant rank regularity condition in mathematical programming. *Optimization*, 60(4):429–440, 2011. 54, 61

[69] L. Minchenko and S. Stakhovski. Parametric nonlinear programming problems under the relaxed constant rank condition. *SIAM Journal on Optimization*, 1:314–332, 2011. 54, 58

[70] A. Mohammadi, B. S. Mordukhovich, and M. E. Sarabi. Parabolic regularity in geometric variational analysis. *Transactions of the American Mathematical Society*, 374:1711–1763, 2021. 28

[71] J. M. Morguerza and F. J. Prieto. An augmented Lagrangian interior-point method using directions of negative curvature. *Mathematical Programming*, 95 (3):573–616, 2003. 81

[72] K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39:2:117–129, 1987. 81

[73] Z. Páles and V. Zeidan. Generalized Jacobian for functions with infinite dimensional range and domain. *Set-Valued Analysis*, 15:331–375, 2007. 84

[74] G. Pataki. The geometry of semidefinite programming. In R. Saigal, L. Vandenberghe, and H. Wolkowicz, editors, *Handbook of Semidefinite Programming.*, pages 29–65. Kluwer Academic Publishers, Waterloo, Canada, 2000. 11, 79

[75] G. Pataki. On the closedness of the linear image of a closed convex cone. *Mathematics of Operations Research*, 32(2):395–412, 2007. 10, 63

[76] G. Pataki. Strong duality in conic linear programming: Facial reduction and extended duals. *In: Bailey D. et al. (eds) Computational and Analytical Mathematics. Springer Proceedings in Mathematics & Statistics*, 50:613–634, 2013. 21, 22

[77] G. Pataki. Characterizing bad semidefinite programs: Normal forms and short proofs. *SIAM Review*, 61(4):839–859, 2019. 18

[78] G. Pataki and L. Tunçel. On the generic properties of convex optimization problems in conic form. *Mathematical Programming*, 89:449–457, 2001. 35

[79] H. D. Qi. Local duality of nonlinear semidefinite programming. *Mathematics of Operations Research*, 34(1):124–141, 2009. 84

[80] L. Qi and Z. Wei. On the constant positive linear dependence conditions and its application to SQP methods. *SIAM Journal on Optimization*, 10(4):963–981, 2000. 3, 40, 41, 46, 73

[81] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970. 6, 7, 8, 10, 16

[82] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin Heidelberg, 1 edition, 1998. 50

[83] A. Shapiro. First and second order analysis of nonlinear semidefinite programs. *SIAM Journal on Optimization*, 77(1):301–320, 1997. 35, 70, 86, 88

[84] A. Shapiro. Differentiability properties of metric projections onto convex sets. *Journal of Optimization Theory and Applications*, 169(3):953–964, 2016. 14

[85] A. Shapiro and M. K. H. Fan. On eigenvalue optimization. *SIAM Journal on Optimization*, 5(3):552–569, 1995. 26, 70

[86] D. Sun. The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Mathematics of Operations Research*, 31(4):761–776, 2006. 84

[87] H. Waki and M. Muramatsu. Facial reduction algorithms for conic optimization problems. *Journal of Optimization Theory and Applications*, 158(1):188–215, 2013. 21

[88] Y. Yamakawa and T. Okuno. Global convergence of a stabilized sequential quadratic semidefinite programming method for nonlinear semidefinite programs without constraint qualifications. *ArXiv:1909.13544*, 2019. 46

[89] H. Yamashita, H. Yabe, and K. Harada. A primal-dual interior point method for nonlinear semidefinite programming. *Mathematical Programming, Series A*, 135:89–121, 2012. 46

[90] E. H. Zarantonello. Projections on convex sets in Hilbert space and spectral theory: Part I. Projections on convex sets: Part II. Spectral theory. In E. H. Zarantonello, editor, *Contributions to Nonlinear Functional Analysis*, pages 237–424. Academic Press, 1971. 12, 15

[91] Y. Zhang and L. Zhang. New constraint qualifications and optimality conditions for second order cone programs. *Set-Valued and Variational Analysis*, 27:693–712, 2019. 54

Appendix A

External reference I

Title: On optimality conditions for nonlinear conic programming.

Status: Accepted for publication in Mathematics of Operations Research.

What is attached: Our copy of the paper.

Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

On Optimality Conditions for Nonlinear Conic Programming

Roberto Andreani, Walter Gómez, Gabriel Haeser, Leonardo M. Mito, Alberto Ramos

http://pubsonline.informs.org/journal/moor

MATHEMATICS OF OPERATIONS RESEARCH
Articles in Advance, pp. 1–26
ISSN 0364-765X (print), ISSN 1526-5471 (online)

On Optimality Conditions for Nonlinear Conic Programming

Roberto Andreani,[a] Walter Gómez,[b] Gabriel Haeser,[c] Leonardo M. Mito,[c] Alberto Ramos[d]

[a] Department of Applied Mathematics, University of Campinas, Campinas 13083-970, Brazil; [b] Department of Mathematical Engineering, Universidad de la Frontera, Temuco 4811230, Chile; [c] Department of Applied Mathematics, University of São Paulo, São Paulo 05508-090, Brazil; [d] Department of Mathematics, Federal University of Paraná, Curitiba 81530-015, Brazil

Contact: andreani@ime.unicamp.br, https://orcid.org/0000-0003-2031-4325 (RA); walter.gomez@ufrontera.cl, https://orcid.org/0000-0003-4366-5451 (WG); ghaeser@ime.usp.br, https://orcid.org/0000-0002-1195-3347 (GH); leokoto@ime.usp.br, https://orcid.org/0000-0003-2851-8285 (LMM); albertoramos@ufpr.br, https://orcid.org/0000-0003-0656-255X (AR)

Received: May 20, 2020
Revised: January 29, 2021
Accepted: July 25, 2021
Published Online in Articles in Advance: December 9, 2021

MSC2000 Subject Classification: Primary: 90C46; secondary: 90C48, 90C30, 90C22
OR/MS Subject Classification: Primary: programming/nonlinear/theory/optimality conditions; secondary: programming/nonlinear/algorithms

https://doi.org/10.1287/moor.2021.1203

Abstract. Sequential optimality conditions play a major role in proving stronger global convergence results of numerical algorithms for nonlinear programming. Several extensions are described in conic contexts, in which many open questions have arisen. In this paper, we present new sequential optimality conditions in the context of a general nonlinear conic framework, which explains and improves several known results for specific cases, such as semidefinite programming, second-order cone programming, and nonlinear programming. In particular, we show that feasible limit points of sequences generated by the augmented Lagrangian method satisfy the so-called approximate gradient projection optimality condition and, under an additional smoothness assumption, the so-called complementary approximate Karush–Kuhn–Tucker condition. The first result was unknown even for nonlinear programming, and the second one was unknown, for instance, for semidefinite programming.

Funding: This work was supported by Fondo de Fomento al Desarrollo Científico y Tecnológico [Grant 1170907], Fundação de Amparo à Pesquisa do Estado de São Paulo [Grants 2013/05475-7, 2013/07375-0, 2017/17840-2, 2017/18308-2, 2018/24293-0, and 2019/13297-8], Conselho Nacional de Desenvolvimento Científico e Tecnológico [Grants 301888/2017-5, 303427/2018-3, 404656/2018-8, 438185/2018-8, and 307270/2019-0], and Programa de Apoio aos Núcleos de Excelência - Conselho Nacional de Desenvolvimento Científico e Tecnológico/Fundação Carlos Chagas Filho de Amparo á Pesquisa do Estado do Rio de Janeiro [Grant E-26/010.001247/2016].

Keywords: nonlinear conic optimization • optimality conditions • numerical methods • global convergence • constraint qualifications

1. Introduction

We are interested in the general *nonlinear conic programming* (NCP) problem, which is usually presented in the following form:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & f(x), \\ \text{subject to} & G(x) \in \mathcal{K}, \end{array} \tag{NCP}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{E}$ are continuously differentiable mappings, $\mathbb{E}$ is a finite-dimensional vector space over $\mathbb{R}$ equipped with an inner product $\langle \cdot, \cdot \rangle$ and the norm $\|x\| := \sqrt{\langle x, x\rangle}$ induced by it, and $\mathcal{K} \subseteq \mathbb{E}$ is a nonempty closed convex cone. Let us denote its feasible set by Ω. This is a general class of optimization problems that encompasses, for instance, some well-known particular cases, such as *nonlinear programming* (NLP), *nonlinear semidefinite programming* (NLSDP), and *nonlinear second order cone programming* (NSOCP). It has applications in several areas, which include but are not restricted to control theory (Fares et al. [32]), truss design problems and combinatorial optimization (Wolkowicz et al. [77]), portfolio optimization (Lobo et al. [54]), structural optimization (Kocvara and Stingl [50]), and others. For more details, see Anjos and Lassere [14], Wolkowicz et al. [77], Yamashita and Yabe [78], and references therein. It is worth mentioning that the content of this paper can be straightforwardly extended to (NCP) with separate equality constraints in the form: $h(x) = 0$ and $G(x) \in \mathcal{K}$, but we stick to (NCP) as is for the sake of simplicity.

Algorithms for solving nonlinear optimization problems are mostly iterative, and their convergence theories are usually built around the limit points of their output sequences. However, numerical methods must employ stopping criteria to properly truncate those sequences, which are often based on necessary optimality conditions. Under a *constraint qualification* (CQ), every local minimizer of (NCP) satisfies the classical *Karush–Kuhn–Tucker*

(KKT) conditions, but even simple problems, such as minimizing x subject to $x^2 \in \{0\}$, may have minimizers that do not satisfy the KKT conditions. For NLP problems, a "sequential" alternative condition with high practical appeal was proposed in Andreani et al. [1] under the name *approximate KKT* (AKKT) condition, that holds at local minimizers independently of CQs, similarly to the *Fritz–John* condition (Mangasarian and Fromovitz [56]), but strictly stronger Andreani et al. [1, theorem 2.2].

Roughly speaking, sequential optimality conditions such as AKKT are general characterizations of feasible limit points of an algorithm's output sequence. It is proved in Andreani et al. [11] that these sequential conditions imply KKT under very weak CQs; for instance, strictly weaker than the *Mangasarian–Fromovitz constraint qualification* (MFCQ). These results assemble a simple unified tool for proving global convergence of algorithms without assuming boundedness of the Lagrange multiplier set at the limit point. Indeed, proving global convergence of an algorithm under such weak CQs reduces to proving that it generates limit points that satisfy a given sequential optimality condition. This was successfully done, for instance, in Andreani et al. [4] and Birgin and Martínez [20] for *augmented Lagrangian* methods; in Gill et al. [38] for a *shifted primal-dual penalty barrier* method; in Gill et al. [37] and Qi and Wei [67] for *sequential quadratic programming* (SQP) methods; in Andreani et al. [8], Chen and Goldfarb [30], Haeser [40], and Haeser et al. [42] for *interior point* methods; and also in Andreani et al. [8] and Birgin et al. [21] for *inexact restoration* methods—see also Andreani et al. [7, 10, 11] and references therein for more details. Conversely, sequential optimality conditions may suggest adaptations for practical algorithms that ensure a better theoretical performance; for instance, in Haeser et al. [42], the authors analyze a sequential optimality condition satisfied by interior point methods, which characterizes the effects of a certain control over the feasibility of the method on its convergence. In particular, they prove that a specific type of control guarantees that the sequence of approximate Lagrange multipliers of the method is bounded even in the absence of a constraint qualification. Moreover, in O'Neill and Wright [62], the authors develop a complexity analysis for an algorithm based on a log-barrier function, Newton's method, and conjugate gradients that converges to second order stationary points via sequential optimality conditions. This kind of complexity analysis via sequential optimality conditions can be done even for some special problems such as in Haeser et al. [43], in which the KKT conditions are not defined at the limit point because of lack of differentiability; then, the AKKT notion serves as a natural optimality condition for it.

Naturally, such an idea has been carried over for several other contexts, for example: *Nash equilibrium* problems (Bueno et al. [28]), *mathematical programs with equilibrium constraints* (Ramos [68, 69]), *mathematical programs with complementarity constraints* (Andreani et al. [9]), *nonlinear vector optimization with conic constraints* (Tuyen et al. [76]), the *multiobjective* case (Giorgi et al. [39]), *variational problems in Banach spaces* (Kanzow et al. [49]), *quasi-equilibrium-problems* (Bueno et al. [29]), and several others. The first extension to a conic context is due to Andreani et al. [2] followed by Andreani et al. [12] for NLSDP and NSOCP, respectively, which gave rise to a more theoretical range of applications of sequential optimality conditions. For instance, there is a recent work that uses sequential optimality conditions (from Andreani et al. [2]) to prove that every local minimizer of a general nonconvex NLSDP problem satisfies a second order condition that depends on a single Lagrange multiplier over the lineality space of the critical cone without assuming nondegeneracy or strict complementarity (the common assumptions for this kind of analysis) (Fukuda et al. [35]). The sequential framework also allows Andreani et al. [5] to define and study weaker variants of the nondegeneracy condition, which are designed to aid in proving global convergence of methods that rely on spectral decompositions or for problems that present some structural sparsity.

This paper aims at expanding the strongest known sequential optimality conditions from NLP to the general conic framework (NCP). The most difficult aspect of such generalizations is dealing with complementarity. For NLP constraints $g(x) \leq 0$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $g = (g_1, \ldots, g_m)$, and a Lagrange multiplier $\overline{\lambda} \in \mathbb{R}^m, \overline{\lambda} \geq 0, \overline{\lambda} = (\overline{\lambda}_1, \ldots, \overline{\lambda}_m)$, the complementarity constraint $\langle \overline{\lambda}, g(\overline{x}) \rangle = 0$ means precisely that, for every $i \in \{1, \ldots, m\}$, the multiplier $\overline{\lambda}_i$ is complementary with respect to the constraint $g_i(x) \leq 0$ in the sense that $\overline{\lambda}_i g_i(\overline{x}) = 0$ at a feasible point of interest $\overline{x}$. It turns out that, when considering perturbations of $\overline{x}$, the latter gives a stronger optimality condition. In the conic case (NCP), it is not clear how to exploit a complementarity-like structure in a statement of the form $\langle \overline{\Lambda}, G(\overline{x}) \rangle = 0$ for a Lagrange multiplier $\overline{\Lambda} \in \mathcal{K}^\circ$ (the polar of the cone $\mathcal{K}$), where $G(\overline{x}) \in \mathcal{K}$. In the context of NLSDP (Andreani et al. [2]), the eigenvalues are heavily employed to exploit a complementarity-like structure, for which one must carefully consider how to order consistently the eigenvalues of $G(\overline{x})$ and $\overline{\Lambda}$. In Andreani et al. [12], this approach is extended to so-called symmetric cones, in which an eigenvalue structure is still available, but a more elegant solution is given by making use of a so-called Jordan product $\circ$, which is inherent to the cone $\mathcal{K} = \{u \circ u : u \in \mathbb{E}\}$. Note that self-duality of $\mathcal{K}$ plays an important role in defining these optimality conditions (Andreani et al. [2, 12]).

In this paper, we propose a much more general and unified approach for defining such conditions. Here, we propose splitting $\langle \overline{\Lambda}, G(\overline{x}) \rangle = 0$, by means of Moreau's decomposition, into two complementarity-like statements of the form $\langle \overline{\Lambda}, \Pi_{\mathcal{K}} G(\overline{x}) \rangle = 0$ and $\langle \overline{\Lambda}, \Pi_{\mathcal{K}^\circ} G(\overline{x}) \rangle = 0$, where $\Pi_{\mathcal{K}}$ and $\Pi_{\mathcal{K}^\circ}$ denote orthogonal projections onto $\mathcal{K}$ and its polar, respectively. Hence, no particular structure of the cone $\mathcal{K}$ is needed. We then show that a primal-dual sequence $\{(x^k, \Lambda^k)\} \subset \mathbb{R}^n \times \mathcal{K}^\circ$ generated by an augmented Lagrangian method is such that $\langle \Lambda^k, \Pi_{\mathcal{K}} G(x^k)\rangle \to 0$. In the context of NLP, an optimality condition associated with this measure of complementarity turns out to be equivalent to the so-called approximate gradient projection (AGP) optimality condition (Andreani et al. [3]), which is strictly stronger than the more common AKKT (Andreani et al. [1]) optimality condition. The revelation of this property of the augmented Lagrangian sequence is somewhat surprising, and it was achieved as a corollary of our more general approach. Also, under an additional smoothness assumption, the other complementarity-like statement $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ} G(x^k)\rangle \to 0$ is also satisfied. This answers an open question of Andreani et al. [2] in the context of NLSDP by presenting a stronger complementarity-like structure, generated by the augmented Lagrangian method, which was not achieved in Andreani et al. [2]. In Andreani et al. [12], although an optimality condition that reveals a strong complementarity-like structure is defined for general symmetric cones, the proof that the sequence generated by the augmented Lagrangian fulfills this property was only done in the context of NSOCP.

Finally, we show that our global convergence results are strictly stronger than the ones usually employed for conic constraints, namely, in which Robinson's CQ is employed. Note that our results do not require that $\mathcal{K}$ has a nonempty interior or self-duality; hence, our results are relevant even when the constraints are linear. Also, because Robinson's CQ may fail, our results imply that, even when the set of Lagrange multipliers is unbounded at a feasible limit point of a sequence generated by the algorithm, a global convergence result is available; that is, the dual sequence $\{\Lambda^k\}$ may diverge.

This paper is organized as follows: Section 2 presents some basic definitions and a short literature review on sequential optimality conditions for NLP and NLSDP. In Section 3, we define sequential conditions for NCP and some standard properties are proven. Section 4 presents an augmented Lagrangian algorithm and its convergence theory in terms of sequential conditions. Section 5 is dedicated to a distinguished extension of AKKT and its relation to the other conditions. Section 6 is focused on contextualizing our conditions when NCP is reduced to NLP, NLSDP, and NSOCP. Section 7 introduces new constraint qualifications that can be useful for the convergence analysis of numerical methods. Finally, Section 8 is dedicated to summarizing our main contributions while presenting our prospective work.

2. Preliminaries

In this section, we recall some basic concepts and results of convex analysis, and we make a more detailed review of sequential optimality conditions for NLP and NLSDP.

2.1. Notations and Convex Analysis Background

Our notation is standard in optimization and variational analysis: $\mathbb{N}$ denotes the set of natural numbers (with $0 \in \mathbb{N}$), and $\mathbb{R}^n$ stands for the n-dimensional real Euclidean space. Let $x \in \mathbb{R}^n$, and we use $B[x, \delta]$ to denote the closed ball with the center at x and radius $\delta > 0$. For a, $b \in \mathbb{R}^n$ with components a_i and b_i, respectively, we use $\max\{a, b\}$ to represent the vector with components $\max\{a_i, b_i\}$. The vector $\min\{a, b\}$ has a similar meaning. We denote the *interior* of a set A by int A. Moreover, we recall that $\mathbb{E}$ is a finite-dimensional linear space equipped with an inner product, which we denote by $\langle \cdot, \cdot \rangle$.

Given a set-valued mapping $\Gamma : \mathbb{R}^s \rightrightarrows \mathbb{E}$, the sequential (Painlevé–Kuratowski) *outer limit* of $\Gamma(z)$ as $z \to \bar{z}$ is the set $\{\bar{w} \in \mathbb{E} : \exists\, (z^k, w^k) \to (\bar{z}, \bar{w}), w^k \in \Gamma(z^k)\}$, which is denoted by $\limsup_{z \to \bar{z}} \Gamma(z)$. Moreover, we say that Γ is outer semicontinuous at $\bar{z}$ when

$$\limsup_{z \to \bar{z}} \Gamma(z) \subseteq \Gamma(\bar{z}).$$

For a differentiable mapping $G : \mathbb{R}^n \to \mathbb{E}$, we use $DG(x)$ to denote the derivative of G at x, and $DG(x)^* : \mathbb{E} \to \mathbb{R}^n$ to denote the *adjoint* of $DG(x)$, which is characterized by the following property: $\langle DG(x)d, \Lambda\rangle = \langle d, DG(x)^*\Lambda\rangle$ for every $d \in \mathbb{R}^n$, $\Lambda \in \mathbb{E}$. For a differentiable real-valued function $f : \mathbb{R}^n \to \mathbb{R}$, we use $\nabla f(x)$ to denote the transpose of $Df(x)$, seen as a $1 \times n$ matrix.

Given a closed convex cone $K \subset \mathbb{E}$, the *polar* of K is the set $K^o := \{w \in \mathbb{E} : \langle w, k\rangle \le 0, \;\; \forall k \in K\}$ and $(K^o)^o = K$. The distance of $w \in \mathbb{E}$ to K is defined as $\mathrm{dist}_K(w) := \min\{\|w - v\| : v \in K\}$, and the *orthogonal projection* of w onto K, denoted by $\Pi_K(w)$, is the point at which the minimum is attained. Moreover, it can be proved that $\Pi_K(w)$ is nonexpansive, that is,

$$\|\Pi_K(w) - \Pi_K(v)\| \le \|w - v\|, \;\forall v, \;\forall w,$$

so it is a Lipschitz continuous function, and it can also be proved that $\text{dist}_K^2(w)$ is a continuously differentiable function whose derivative is given by (1). For a proof, see Fitzpatrick and Phelps [34].

$$D(\text{dist}_K^2)(w) = 2(w - \Pi_K(w)),\ \forall w. \tag{1}$$

The following lemma (see, e.g., Hiriart-Urruty and Lemaréchal [47, theorems 3.2.3 and 3.2.5]) encompasses other well-known properties of projections:

Lemma 1. *Let $K \subset \mathbb{E}$ be a closed convex cone and $w \in \mathbb{E}$. Then,*

1. $v = \Pi_K(w)$ *if and only if* $v \in K$, $w - v \in K^{\circ}$, *and* $\langle w - v, v\rangle = 0$;
2. $\Pi_K(\alpha w) = \alpha \Pi_K(w)$, *for every* $\alpha \geq 0$, *and* $\Pi_K(-w) = -\Pi_{-K}(w)$;
3. *(Moreau's decomposition) for every* $w \in \mathbb{E}$, *we have* $w = \Pi_K(w) + \Pi_{K^{\circ}}(w)$ *and* $\langle \Pi_K(w), \Pi_{K^{\circ}}(w)\rangle = 0$.

2.2. Sequential Optimality Conditions for NLP and NLSDP

In order to start a deeper discussion on sequential optimality conditions, we make a brief exposition of the most important results around them in NLP, in which it has been extensively studied, and a summary of some recent advances in NLSDP.

Consider the following NLP problem in standard form:

$$\begin{aligned} &\underset{x\in\mathbb{R}^n}{\text{Minimize}} && f(x),\\ &\text{subject to} && G(x) \leq 0, \end{aligned} \tag{NLP}$$

which is (NCP) with $\mathbb{E} = \mathbb{R}^m$ and $\mathcal{K} = \mathbb{R}^m_- := \{z \in \mathbb{R}^m : \forall i \in \{1, \ldots, m\}, z_i \leq 0\}$. Following Andreani et al. [1], we say that the AKKT condition holds at a feasible point $\bar{x}$ when there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m_+ := -\mathbb{R}^m_-$ such that

$$\nabla f(x^k) + \sum_{i=1}^{m} \Lambda_i^k \nabla G_i(x^k) \to 0, \tag{2}$$

and $\Lambda_i^k = 0$, whenever $G_i(\bar{x}) < 0$, for sufficiently large k. Note that AKKT allows divergence of the multiplier sequences associated with active constraints. It has been proved that, under some constraint qualifications, weaker than the *linear independence constraint qualification* and MFCQ, for example, every AKKT point also satisfies the KKT conditions (see Andreani et al. [7, 8, 10]). Then, as mentioned in the introduction, because many algorithms generate AKKT sequences, this improves their convergence theory in a unified manner. Another practical advantage of sequential conditions is their relation to natural choices of stopping criteria for algorithms; for example, it is elementary to verify that AKKT holds at $\bar{x}$ if and only if, for every $\varepsilon > 0$, there is some $x_\varepsilon \in B[\bar{x}, \varepsilon]$ and some approximate multiplier $\Lambda_\varepsilon \geq 0$ such that

$$\|\max\{0, G(x_\varepsilon)\}\| \leq \varepsilon,\ \left\|\nabla f(x_\varepsilon) + \sum_{i=1}^{m} (\Lambda_\varepsilon)_i \nabla G_i(x_\varepsilon)\right\| \leq \varepsilon,\ \|\min\{\Lambda_\varepsilon, -G(x_\varepsilon)\}\| \leq \varepsilon. \tag{3}$$

The properties that made AKKT useful motivate the following general description of a "good" sequential optimality condition that provides guidelines for defining new ones:

1. It must be a necessary optimality condition independent of the fulfillment of any CQ.
2. There must be meaningful numerical methods that generate sequences whose limit points satisfy it.
3. It must imply optimality conditions in the form "KKT or not-CQ" for very weak CQs.

The third property measures the strength of such a sequential optimality condition in comparison with standard ones, and the first one guarantees that no local minimizer is censured by it. In addition, the second property means that one must be able to employ it to formalize the convergence theory of at least one algorithm. It should be observed that, as long as they satisfy those three properties, the stronger the condition (in the logical implication sense), the better. The ability of strengthening global convergence results is of paramount importance because, otherwise, stronger optimality conditions could be derived without resorting to the sequential approach.

For improving the AKKT condition for (NLP), it is proposed in Andreani et al. [3] the so-called *complementary AKKT* (CAKKT) condition, that holds at a feasible point $\bar{x}$ when there are sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m_+$ such that (2) holds and

$$\Lambda_i^k G_i(x^k) \to 0,\ \forall i \in \{1, \ldots, m\}.$$

Indeed, the CAKKT condition is strictly stronger than the AKKT condition, but an additional property (the so-called *generalized Łojasiewicz inequality*) is needed in order to prove that the augmented Lagrangian algorithm generates CAKKT sequences.

Another interesting sequential condition is introduced in Martínez and Svaiter [60] under the name AGP, which holds at a feasible point $\bar{x}$ when there exists some sequence $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ such that

$$\|\Pi_{L(\Omega,x^k)}(-\nabla f(x^k))\| \to 0, \tag{4}$$

where $L(\Omega, x) := \{d \in \mathbb{R}^n : \min\{0, G_i(x)\} + \nabla G_i(x)^T d \leq 0,$ for all i such that $G_i(\bar{x}) = 0\}$. One of the most highlighted features of AGP is its lack of Lagrange multiplier approximations, using projections instead. This makes it useful for supporting the global convergence of numerical optimization methods in which multiplier approximations are not explicitly available; for example, algorithms based on *inexact restoration* (IR) procedures. See Bueno et al. [27], Fischer and Friedlander [33]. Martínez and Pilotta [58, 59], Martínez and Svaiter [60], and references therein for details.

Now, consider the following NLSDP problem:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & f(x), \\ \text{subject to} & G(x) \in \mathbb{S}^m_-, \end{array} \tag{NLSDP}$$

which is a particular case of (NCP) in which $\mathbb{E} = \mathbb{S}^m$ is the linear space of $m \times m$ symmetric matrices and $\mathcal{K} = \mathbb{S}^m_-$ is the cone of $m \times m$ symmetric negative semidefinite matrices. The AKKT extension for (NLSDP) presented in Andreani et al. [2, definition 3.1] holds at a feasible point $\bar{x}$ when there are sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathbb{S}^m_+ := -\mathbb{S}^m_-$ such that

$$\nabla f(x^k) + DG(x^k)^*\Lambda^k \to 0 \tag{5}$$

and $\lambda_i^{U^k}(\Lambda^k) = 0$ whenever $\lambda_i^U(G(\bar{x})) < 0$ and k is sufficiently large for some sequence of orthogonal matrices $\{U^k\}_{k\in\mathbb{N}} \to U$, where U^k diagonalizes Λ^k for each k, and U diagonalizes $G(\bar{x})$. The notation $\lambda_i^U(G(\bar{x}))$ stands for the i th eigenvalue in the diagonal of $U^T G(\bar{x}) U$, and the same goes for $\lambda_i^{U^k}(\Lambda^k)$. This is done for imbuing the notion of ordering into the eigenvalues of the multipliers and for establishing a proper correspondence with the eigenvalues of $G(\bar{x})$, which makes this extension natural from the NLP context but very dependent on the structure of $\mathbb{S}^m$. Still under the same analogy, the most natural extension of CAKKT, discussed in Andreani et al. [2], would simply require (5) and

$$\lambda_i^{S^k}(G(x^k))\lambda_i^{U^k}(\Lambda^k) \to 0, \tag{6}$$

where $\{S^k\}_{k\in\mathbb{N}} \to U$ is a sequence of orthogonal matrices that diagonalizes $G(x^k)$ for each k. However, although this is an actual optimality condition, it is not possible at the moment to provide an algorithm capable of generating sequences with these properties even under generalized Łojasiewicz. Then, instead of using the eigenvalue product, Andreani et al. [2] use the canonical inner product of $\mathbb{S}^m$ (given by the trace of the matrix product) to define a new condition called *trace AKKT* (TAKKT), that requires (5) and

$$\langle \Lambda^k, G(x^k)\rangle \to 0.$$

Surprisingly, TAKKT is proven to be completely independent of AKKT (see Andreani et al. [2, example 5.2; 12, example 3.1]), and it also requires the generalized Łojasiewicz inequality to hold for it to be generated by the augmented Lagrangian algorithm. However, observe that TAKKT can be equivalently stated in NLP using diagonal matrices, and in this context, it is strictly implied by CAKKT.

3. New Optimality Conditions for Nonlinear Conic Programming

In this section, we propose new sequential optimality conditions for general optimization problems, we prove some of their properties, and we clarify the relations among them.

Before we begin, recall (from Bonnans and Shapiro [25] and Robinson [70], for example) that the KKT conditions hold at a feasible point $\bar{x}$ of (NCP) when there exists some *Lagrange multiplier* $\overline{\Lambda} \in \mathcal{K}^\circ$ such that

$$\nabla f(\bar{x}) + DG(\bar{x})^*\overline{\Lambda} = 0, \tag{7}$$

$$\langle \overline{\Lambda}, G(\bar{x})\rangle = 0. \tag{8}$$

By taking the natural relaxation of (7) and (8), we obtain a trivial extension of the TAKKT condition (Andreani et al. [2]) from NLSDP to NCP, replacing the trace product with an arbitrary inner product.

Definition 1 (TAKKT). Let $\bar{x}$ be a feasible point of (NCP). We say that $\bar{x}$ satisfies the TAKKT condition if there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathcal{K}^\circ$ such that

$$\nabla f(x^k) + DG(x^k)^*\Lambda^k \to 0, \tag{9}$$

$$\langle \Lambda^k, G(x^k)\rangle \to 0. \tag{10}$$

Points that satisfy TAKKT are usually called "TAKKT points," and the sequences associated with them are called "TAKKT sequences." Similar names hold for the other sequential conditions. At the end of this section, we prove that TAKKT is an optimality condition for (NCP) as well. Before that, note that, at a KKT pair $(\bar{x}, \overline{\Lambda}) \in \mathbb{R}^n \times \mathcal{K}^\circ$, we also have

$$\langle \overline{\Lambda}, \Pi_{\mathcal{K}}(G(\bar{x}))\rangle = \langle \overline{\Lambda}, G(\bar{x})\rangle - \langle \overline{\Lambda}, \Pi_{\mathcal{K}^\circ}(G(\bar{x}))\rangle = \langle \overline{\Lambda}, G(\bar{x})\rangle = 0,$$

by Moreau's decomposition, so (8) can be equivalently stated as $\langle \overline{\Lambda}, \Pi_{\mathcal{K}}(G(\bar{x}))\rangle = 0$. Relaxing this alternative expression leads to the following:

Definition 2 (AGP). Let $\bar{x}$ be a feasible point of (NCP). We say that $\bar{x}$ satisfies the AGP condition if there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathcal{K}^\circ$ such that (9) holds and

$$\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k))\rangle \to 0. \tag{11}$$

A similar optimality condition appears in Kanzow et al. [49, definition 5.2], in which the authors deal with a version of (NCP) over infinite-dimensional *Banach spaces*, in which $\mathcal{K}$ is contained in a *Hilbert lattice*. However, the authors refer to it as "asymptotic KKT (AKKT)," which does not make it clear how strong their results are. We point out that AGP might be a more appropriate name for, when Definition 2 is reduced to NLP, it is equivalent to the concept with the same name introduced in Martínez and Svaiter [60], which is given by (4). What follows is a proof of our claim.

Theorem 1. *Consider (NLP), which is (NCP) with $\mathbb{E} = \mathbb{R}^m$ and $\mathcal{K} = \mathbb{R}^m_-$. Let $\bar{x}$ be a feasible point for it. Then, AGP as in Definition 2 holds at $\bar{x}$ if and only if AGP as in (4) holds at $\bar{x}$.*

Proof. Let $\bar{x}$ satisfy Definition 2. Then, there exist sequences $x^k \to \bar{x}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m_+$ such that (9) and (11) hold. Now, define $d^k := \Pi_{L(\Omega, x^k)}(-\nabla f(x^k))$, $\forall k \in \mathbb{N}$. By definition, d^k is a solution of

$$\begin{array}{ll} \underset{d\in\mathbb{R}^n}{\text{Minimize}} & \frac{1}{2}\|-\nabla f(x^k) - d\|^2, \\ \text{subject to} & \min\{0, G_i(x^k)\} + DG_i(x^k)d \le 0, \ \forall i \in \mathcal{A}(\bar{x}), \end{array} \tag{12}$$

where $\mathcal{A}(\bar{x}) := \{i \in \{1, \dots, m\} : G_i(\bar{x}) = 0\}$. Because the constraints are linear, by the KKT conditions, there exists some $\widehat{\Lambda}^k \in \mathbb{R}^{|\mathcal{A}(\bar{x})|}_+$ such that the first order conditions hold for it. Define $\overline{\Lambda}^k_i \in \mathbb{R}^m_+$ such that

$$\overline{\Lambda}^k_i := \begin{cases} \widehat{\Lambda}^k_i, & \text{if } i \in \mathcal{A}(\bar{x}), \\ 0, & \text{otherwise.} \end{cases}$$

Hence,

$$\nabla f(x^k) + d^k + DG(x^k)^*\overline{\Lambda}^k = 0 \text{ and } \langle \overline{\Lambda}^k, \min\{0, G(x^k)\} + DG(x^k)d^k\rangle = 0. \tag{13}$$

Multiplying (13) by d^k, we obtain

$$\begin{aligned} \|d^k\|^2 &= -\langle \nabla f(x^k), d^k\rangle - \langle d^k, DG(x^k)^*\overline{\Lambda}^k\rangle \\ &= -\langle \nabla f(x^k), d^k\rangle - \langle DG(x^k)d^k, \overline{\Lambda}^k\rangle \\ &= -\langle \nabla f(x^k), d^k\rangle - \langle DG(x^k)d^k + \min\{0, G(x^k)\}, \overline{\Lambda}^k\rangle + \\ &\quad + \langle \min\{0, G(x^k)\}, \overline{\Lambda}^k\rangle \\ &\le -\langle \nabla f(x^k), d^k\rangle \\ &= -\langle \nabla f(x^k) + DG(x^k)^*\Lambda^k, d^k\rangle + \langle DG(x^k)d^k + \min\{0, G(x^k)\}, \Lambda^k\rangle - \\ &\quad - \langle \min\{0, G(x^k)\}, \Lambda^k\rangle \\ &\le \|\nabla f(x^k) + DG(x^k)^*\Lambda^k\| \|d^k\| + \\ &\quad + \sum_{i\in\mathcal{A}(\bar{x})} (DG_i(x^k)d^k + \min\{0, G_i(x^k)\})\Lambda^k_i - \langle \min\{0, G(x^k)\}, \Lambda^k\rangle, \end{aligned}$$

where in the last inequality we used the Cauchy–Schwarz inequality and that d^k is feasible for (12). Moreover, because $\|d^k\| \leq \|\nabla f(x^k)\|$ and (9) and (11) hold, we obtain $d^k \to 0$, and (4) holds because $\Lambda_i^k \to 0$ for every $i \notin \mathcal{A}(\bar{x})$.

Conversely, assume that $\bar{x}$ satisfies AGP as in (4) and set $d^k := \Pi_{L(\Omega,x^k)}(-\nabla f(x^k))$. Analogously, because d^k is a global minimizer of (12), there is an analogous choice of $\overline{\Lambda}^k \in \mathbb{R}^m_+$ such that (13) holds. Then,

$$\langle \overline{\Lambda}^k, \min\{0, G(x^k)\}\rangle = -\langle \overline{\Lambda}^k, DG(x^k)d^k\rangle = -\langle DG(x^k)^*\overline{\Lambda}^k, d^k\rangle = \langle \nabla f(x^k) + d^k, d^k\rangle. \tag{14}$$

By (4), (13), and (14), we obtain $\nabla f(x^k) + DG(x^k)^*\overline{\Lambda}^k = -d^k \to 0$ and $\langle \overline{\Lambda}^k, \min\{0, G(x^k)\}\rangle \to 0$. Consequently, Definition 2 holds at $\bar{x}$. □

Surprisingly, although AGP and TAKKT look like twins, they are completely independent. The following counterexample shows that TAKKT does not imply AGP.

Example 1 (TAKKT Does Not Imply AGP). In $\mathbb{R}^2$, consider the nonlinear programming problem to minimize $-x_2$ subject to $G(x_1, x_2) \in \mathcal{K}$, where $G(x_1, x_2) := (-x_1, x_1 \exp(x_2))$, $\mathcal{K} := \mathbb{R}^2_-$, and the feasible point $\bar{x} = (0, 1)$. In this case, $\Lambda^k := (\lambda_1^k, \lambda_2^k) \in \mathbb{R}^2_+$ and $\nabla f(x^k) + DG(x^k)^*\Lambda^k \to 0$ reduces to $-\lambda_1^k + \lambda_2^k \exp(x_2^k) \to 0$ and $-1 + \lambda_2^k x_1^k \exp(x_2^k) \to 0$.

TAKKT holds at $\bar{x}$: take $x_1^k := 1/k$, $x_2^k := 1$, $\lambda_2^k := (x_1^k \exp(x_2^k))^{-1}$, $\lambda_1^k := \exp(x_2^k)\lambda_2^k$. It is elementary to verify that $\{x^k = (x_1^k, x_2^k)\}$ is a TAKKT sequence.

AGP fails at $\bar{x}$: assume that there is an AGP sequence $\{x^k\}$. We observe that the approximate complementarity condition $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k))\rangle \to 0$ implies that $\lambda_1^k \min\{-x_1^k, 0\} + \lambda_2^k \exp(x_2^k) \min\{x_1^k, 0\} \to 0$. If there is an AGP sequence with $x_1^k > 0$ for an infinite set of indices, from the complementarity condition, we have that $-x_1^k\lambda_1^k \to 0$, and thus, $x_1^k\lambda_2^k \exp(x_2^k) \to 0$, which is a contradiction with $-1 + \lambda_2^k x_1^k \exp(x_2) \to 0$. Similar results are obtained if there is an AGP sequence with $x_1^k < 0$ (or $x_1^k = 0$) for an infinite set of indices. □

Now, we show that AGP does not imply TAKKT either.

Example 2 (AGP Does Not Imply TAKKT). Consider the nonlinear programming problem in $\mathbb{R}^2$ to minimize x_2 subject to $G(x_1, x_2) := x_2 h(x_1) \in \mathcal{K} = \{0\} \subset \mathbb{R}$, where $h : \mathbb{R} \to \mathbb{R}$ is the C^1 function introduced in Andreani et al. [3], defined as

$$h(z) := \begin{cases} z^4 \sin(z^{-1}) & \text{if } z \neq 0; \\ 0 & \text{if } z = 0. \end{cases} \tag{15}$$

Consider the point $\bar{x} := (0, 1)$. Following Andreani et al. [3], we see that there exists a sequence $\{z^k\}_{k\in\mathbb{N}} \subset \mathbb{R}$ such that $z^k \to 0$, $h'(z^k) = -(z^k)^5$ and $\sin(1/z^k) \to 1$.

AGP holds at $\bar{x}$. First, choose a sequence $x^k := (z^k, 1)$ with $\Lambda := -(z^k)^{-4} \in \mathcal{K}^\circ = \mathbb{R}$. Now, observe that $\nabla f(x^k) + DG(x^k)\Lambda^k$ goes to zero because

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \Lambda^k \begin{pmatrix} h'(z^k) \\ h(z^k) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{-1}{(z^k)^4} \begin{pmatrix} -(z^k)^5 \\ (z^k)^4 \sin(1/z^k) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} z^k \\ -\sin(1/z^k) \end{pmatrix} \to \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{16}$$

Finally, the approximate complementarity condition trivially holds because $\Pi_{\mathcal{K}}(G(x^k)) = 0$.

TAKKT fails at $\bar{x}$. Suppose that there exists a sequence $\{x^k := (x_1^k, x_2^k)\}_{k\in\mathbb{N}}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathbb{R}$ conforming to the definition of TAKKT. Then,

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \Lambda^k \begin{pmatrix} x_2^k h'(x_1^k) \\ h(x_1^k) \end{pmatrix} \to \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{17}$$

The approximate complementarity condition of TAKKT implies $\Lambda^k G(x_1^k, x_2^k) = \Lambda^k x_2^k h(x_1^k) \to 0$. Because $x_2^k \to 1$, we get $\Lambda^k h(x_1^k) \to 0$, which is a contradiction with (17). □

Through Definition 2 and Theorem 1, it is possible to see AGP as an incomplete CAKKT condition in (NLP), which is a different interpretation from Martínez and Svaiter [60]. Indeed, note that AGP holds at $\bar{x}$ with sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m_+$ if and only if $\Lambda_i^k G_i(x^k) \to 0$ whenever $G_i(x^k) \leq 0$. Because both AGP and CAKKT push Λ_i^k to zero when $G_i(\bar{x}) < 0$, they only differ when $G_i(\bar{x}) = 0$. Even though the CAKKT condition allows divergence of Λ_i^k in this case, it demands it to go to infinity slower than $G_i(x^k)$ goes to zero although AGP may allow a faster growth as long as $\{x^k\}_{k\in\mathbb{N}}$ violates $G_i(x^k) \leq 0$. From this point of view, CAKKT improves AGP by introducing some control in the behavior of the multiplier sequences associated with the infeasible part of the constraints, using a quantitative measure of such infeasibility. Generalizing this reasoning, we obtain the following:

Definition 3 (CAKKT). Let $\bar{x}$ be a feasible point. We say that $\bar{x}$ satisfies the CAKKT condition if there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathcal{K}^\circ$ such that (9) and (11) hold, and

$$\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle \to 0. \tag{18}$$

Observe that CAKKT as in Definition 3 is indeed a generalization of the condition with the same name from Andreani et al. [3] for (NLP) because $\Lambda_i^k G_i(x^k) \to 0$ in this case, independently on the sign of $G_i(x^k)$. Moreover, note that CAKKT is essentially AGP upgraded with (18). Therefore, in view of Moreau's decomposition, CAKKT clearly implies both AGP and TAKKT. But, because they are independent, the implications are strict. We proceed by showing that CAKKT is a genuine necessary optimality condition, that is, a property that must be satisfied by every local minimizer, even the ones that do not satisfy any constraint qualification.

Theorem 2. *If $\bar{x}$ is a local minimizer of (NCP), then $\bar{x}$ satisfies the CAKKT condition.*

Proof. Let $\bar{x}$ be a local minimizer of (NCP) in $B[\bar{x},\delta]$ for some $\delta > 0$. Then, $\bar{x}$ is a global minimizer of

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & f(x) + \frac{1}{2}\|x-\bar{x}\|^2, \\ \text{subject to} & G(x) \in \mathcal{K} \\ & \|x-\bar{x}\| \le \delta. \end{array}$$

Let $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$, and for each $k \in \mathbb{N}$, consider the penalized optimization problem

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & f(x) + \frac{1}{2}\|x-\bar{x}\|^2 + \rho_k \frac{1}{2}\|\Pi_{\mathcal{K}^\circ}(G(x))\|^2, \\ \text{subject to} & \|x-\bar{x}\| \le \delta. \end{array} \tag{19}$$

Denote by x^k a global solution of (19). Using analogous arguments as in the standard external penalty algorithm convergence proof Nocedal and Wright [61], we see that $x^k \to \bar{x}$, and thus, $\|x^k - \bar{x}\| < \delta$ for k large enough. Using Fermat's rule applied to (19), we have

$$\nabla f(x^k) + (x^k - \bar{x}) + DG(x^k)^*\Lambda^k = 0, \text{ where } \Lambda^k := \rho_k \Pi_{\mathcal{K}^\circ}(G(x^k)) \in \mathcal{K}^\circ,$$

and the expression for the derivative follows from (1) along with the definition of orthogonal projection and Moreau's decomposition. Thus, (9) holds. Furthermore, by the definition of Λ^k and Lemma 1, we see that $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k))\rangle = \langle \rho_k \Pi_{\mathcal{K}^\circ}(G(x^k)), \Pi_{\mathcal{K}}(G(x^k))\rangle = 0$; that is, (11) holds. We proceed to show that (18) is satisfied. First, from the optimality of x^k, we have

$$f(x^k) + \frac{1}{2}\|x^k-\bar{x}\|^2 + \frac{1}{2}\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle = f(x^k) + \frac{1}{2}\|x^k-\bar{x}\|^2 + \rho_k\frac{1}{2}\|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \le f(\bar{x}),$$

which leads to

$$0 \le \langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle = \rho_k\|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \le 2(f(\bar{x}) - f(x^k)) - \|x^k-\bar{x}\|^2,$$

and because $x^k \to \bar{x}$ and $f(x^k) \to f(\bar{x})$, we see that $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle \to 0$. Thus, CAKKT holds at $\bar{x}$. □

Consequently, from our previous discussion, we get the following:

Corollary 1. *If $\bar{x}$ is a local minimizer of (NCP), then $\bar{x}$ satisfies AGP and TAKKT.*

In the next section, we propose an augmented Lagrangian algorithm for NCP based on projections onto $\mathcal{K}^\circ$, and we build its convergence theory using the new conditions.

4. An Augmented Lagrangian Algorithm

Employing augmented Lagrangian methods to find the solution of optimization problems is a very successful technique for solving finite-dimensional problems, and it is described in several textbooks on continuous optimization, for example, Bertsekas [18, 19], Birgin and Martínez [20], Nocedal and Wright [61], and Schnabel [74], to cite a few of them. In this section, we show that a variant of the *Powell–Hestenes–Rockafellar* (Hestenes [46], Powell [65], Rockafellar [72]) augmented Lagrangian algorithm generates AGP sequences without any additional condition and also CAKKT (and TAKKT) sequences under the so-called generalized Łojasiewicz inequality (see (28) for the definition). The augmented Lagrangian variant that we consider is a direct generalization of the one considered in Andreani et al. [2, 3] and in the book Birgin and Martínez [20], called Algencan. Roughly speaking, the main distinguishing features of Algencan is the use of a step control strategy (step 3) and a safeguarding strategy (step 2). The latter consists of solving the inner subproblem of the method (step 1) using a bounded Lagrange multiplier estimate to mitigate numerical instability. A detailed study of the effects of these modifications over the "quality of convergence" of the method is presented, at least for the particular case of NLP, in Andreani et al. [13] and Kanzow and Steck [48].

Given $\rho > 0$, let $L_\rho : \mathbb{R}^n \times \mathcal{K}^\circ \to \mathbb{R}$ be the *augmented Lagrangian function* of (NCP), defined as

$$L_\rho(x, \Lambda) := f(x) + \frac{\rho}{2}\left[\left\|\Pi_{\mathcal{K}^\circ}\left(G(x) + \frac{\Lambda}{\rho}\right)\right\|^2 - \left\|\frac{\Lambda}{\rho}\right\|^2\right],$$

whose partial derivative with respect to x is given by

$$\nabla_x L_\rho(x, \Lambda) = \nabla f(x) + DG(x)^*\left(\rho \Pi_{\mathcal{K}^\circ}\left(G(x) + \frac{\Lambda}{\rho}\right)\right). \tag{20}$$

The expression of the derivative in (20) is what motivates the particular choice of Lagrange multiplier update in the following algorithm:

Algorithm 1 (General Framework: Augmented Lagrangian)

Inputs: A sequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$ of positive scalars such that $\varepsilon_k \to 0$; a nonempty convex compact set $\mathcal{B} \subset \mathcal{K}^\circ$; real parameters $\tau > 1$, $\sigma \in (0, 1)$, and $\rho_0 > 0$; and initial points $(x^{-1}, \widehat{\Lambda}^0) \in \mathbb{R}^n \times \mathcal{B}$. Also, define $\|V^{-1}\| = \infty$.

For every $k \in \mathbb{N}$,

1. Compute some point x^k such that

$$\|\nabla_x L_{\rho_k}(x^k, \widehat{\Lambda}^k)\| \le \varepsilon_k; \tag{21}$$

2. Update the multiplier

$$\Lambda^k := \rho_k \Pi_{\mathcal{K}^\circ}\left(G(x^k) + \frac{\widehat{\Lambda}^k}{\rho_k}\right), \tag{22}$$

and compute some $\widehat{\Lambda}^{k+1} \in \mathcal{B}$ (typically, the projection of Λ^k onto $\mathcal{B}$);

3. Define

$$V^k := \frac{\widehat{\Lambda}^k}{\rho_k} - \Pi_{\mathcal{K}^\circ}\left(G(x^k) + \frac{\widehat{\Lambda}^k}{\rho_k}\right); \tag{23}$$

4. If $\|V^k\| \le \sigma\|V^{k-1}\|$, set $\rho_{k+1} := \rho_k$. Otherwise, choose some $\rho_{k+1} \ge \tau\rho_k$.

In Algorithm 1, the role of the set $\mathcal{B}$ is to bound the Lagrange multiplier estimates used in step 1, as part of the so-called safeguarding strategy; in practice, $\mathcal{B}$ should be any set that allows projections to be computed easily, such as a box. Note that step 4 implies that either $\rho^k \to \infty$ or there is some $k_0 \in \mathbb{N}$ such that $\rho_k = \rho_{k_0}$ for every $k > k_0$. In the latter case, it holds also that $\|V^k\| \to 0$. With this in mind, we proceed by showing that Algorithm 1 generates sequences whose limit points satisfy AGP.

Theorem 3. *Let $\bar{x}$ be a feasible limit point of a sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by the augmented Lagrangian method. Then, $\bar{x}$ satisfies AGP.*

Proof. First, from (21) together with (22) and (20), we get

$$\|\nabla f(x^k) + DG(x^k)^*\Lambda^k\| \le \varepsilon_k, \quad \text{with} \quad \Lambda^k = \rho_k \Pi_{\mathcal{K}^\circ}(G(x^k) + \rho_k^{-1}\widehat{\Lambda}^k), \tag{24}$$

which implies (9) and $\Lambda^k \in \mathcal{K}^\circ$. Taking a subsequence if necessary, we can suppose that $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$. We consider two cases depending on whether the sequence $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded or not:

1. Suppose that $\rho_k \to \infty$. By (24) and Lemma 1, item 3, we have $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k) + \rho_k^{-1}\widehat{\Lambda}^k)\rangle = 0$ for every $k \in \mathbb{N}$, which yields

$$\begin{aligned} |\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k))\rangle| &= |\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) - \Pi_{\mathcal{K}}(G(x^k) + \rho_k^{-1}\widehat{\Lambda}^k)\rangle| \\ &\le \|\Lambda^k\| \|\Pi_{\mathcal{K}}(G(x^k)) - \Pi_{\mathcal{K}}(G(x^k) + \rho_k^{-1}\widehat{\Lambda}^k)\| \\ &\le \|\Lambda^k\| \|\rho_k^{-1}\widehat{\Lambda}^k\|, \end{aligned}$$

where the inequalities follow from the Cauchy–Schwarz inequality and the nonexpansiveness of the projection, respectively. But note that $\|\rho_k^{-1}\Lambda^k\|\,\|\widehat{\Lambda}^k\| = \|\Pi_{\mathcal{K}^\circ}\left(G(x^k)+\rho_k^{-1}\widehat{\Lambda}^k\right)\|\,\|\widehat{\Lambda}^k\|$ converges to zero by the continuity of the projection, so $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k))\rangle \to 0$ and AGP holds at $\bar{x}$.

2. If $\{\rho_k\}_{k\in\mathbb{N}}$ is a bounded sequence, it must be constant for sufficiently large k. Note that

$$\begin{aligned}\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k))\rangle &= \rho_k^{-1}\langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)+\widehat{\Lambda}^k), \Pi_{\mathcal{K}}(\rho_k G(x^k))\rangle \\ &= \rho_k^{-1}\langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)+\widehat{\Lambda}^k), \Pi_{\mathcal{K}}(\rho_k G(x^k)) - \Pi_{\mathcal{K}}(\rho_k G(x^k)+\widehat{\Lambda}^k)\rangle,\end{aligned} \tag{25}$$

where, in the second equality, we use $\langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)+\widehat{\Lambda}^k), \Pi_{\mathcal{K}}(\rho_k G(x^k)+\widehat{\Lambda}^k)\rangle = 0$. It remains to show that the right-hand side of (25) goes to zero. In fact, because $\Lambda^k = \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)+\widehat{\Lambda}^k)$ is a bounded sequence, we only need to prove that $\Pi_{\mathcal{K}}(\rho_k G(x^k)) - \Pi_{\mathcal{K}}(\rho_k G(x^k)+\widehat{\Lambda}^k)$ converges to zero. Indeed, the reasoning is as follows:
By step 4 of Algorithm 1, we see that V^k converges to zero, and hence, $\rho_k V^k \to 0$. Using the definition of V^k we get that

$$\Lambda^k - \widehat{\Lambda}^k = \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)+\widehat{\Lambda}^k) - \widehat{\Lambda}^k = -\rho_k V^k \to 0,$$

and thus, $\Lambda^k = \widehat{\Lambda}^k - \rho_k V^k$ characterizes a bounded sequence. To show that $\Pi_{\mathcal{K}}(\rho_k G(x^k)) - \Pi_{\mathcal{K}}(\rho_k G(x^k)+\widehat{\Lambda}^k)$ converges to zero, consider the next expression

$$\begin{aligned}\rho_k G(x^k) - \Pi_{\mathcal{K}}(\rho_k G(x^k)+\widehat{\Lambda}^k) &= \rho_k G(x^k) + \widehat{\Lambda}^k - \Pi_{\mathcal{K}}(\rho_k G(x^k)+\widehat{\Lambda}^k) - \widehat{\Lambda}^k \\ &= \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)+\widehat{\Lambda}^k) - \widehat{\Lambda}^k = \Lambda^k - \widehat{\Lambda}^k \to 0.\end{aligned} \tag{26}$$

Using this expression and the nonexpansiveness of the projection onto $\mathcal{K}$, we get that

$$\|\Pi_{\mathcal{K}}(\rho_k G(x^k)) - \Pi_{\mathcal{K}}(\rho_k G(x^k)+\widehat{\Lambda}^k)\| \le \|\rho_k G(x^k) - \Pi_{\mathcal{K}}(\rho_k G(x^k)+\widehat{\Lambda}^k)\| \to 0. \tag{27}$$

Thus, we obtain that $\bar{x}$ is an AGP point associated with $\{x^k\}_{k\in\mathbb{N}}$. □

Remark 1. IR algorithms are well-known methods for solving NLP problems (see Birgin et al. [23], Martínez and Pilotta [58, 59] for details). The philosophy behind them consists of dealing with feasibility and optimality in different stages. Hence, IR methods fit well in difficult problems whose structure allows the implementation of an efficient feasibility restoration procedure (Bueno et al. [27]). The AGP condition plays a pivotal role in obtaining global convergence results for IR methods (Birgin et al. [21]), but its applicability beyond that class of algorithms was still unclear. Theorem 3 solves this issue by showing that the convergence theory of the augmented Lagrangian is also supported by AGP, which is not an obvious result. From this point of view, Theorems 1 and 3 show that IR methods generate solution candidates at least as good as augmented Lagrangian methods for NLP.

Next, we show that Algorithm 1 generates CAKKT sequences under an additional condition called generalized Łojasiewicz inequality, that is satisfied by a point $\bar{x}$ and a function Ψ when there exist some $\delta > 0$ and a continuous function $\psi(x): B(\bar{x},\delta) \subset \mathbb{R}^n \to \mathbb{R}$ such that $\psi(x) \to 0$ when $x \to \bar{x}$, and

$$|\Psi(x) - \Psi(\bar{x})| \le \psi(x)\|D\Psi(x)\| \text{ for every } x \in B(\bar{x},\delta). \tag{28}$$

This property coincides with the inequality with the same name that is proposed in Andreani et al. [3]. Such types of property have been extensively used in optimization methods, complexity theory, stability of gradient systems etc. See, for instance Attouch et al. [15, 16], Bolte et al. [24], Chill and Mildner [31], Kurdyka [52], Li and Pong [53], Łojasiewicz [55], and references therein. For instance, all analytic functions satisfy it, and so does every function that satisfies the classical Łojasiewicz [55] inequality. Now, we may resume our results:

Theorem 4. *Let $\bar{x}$ be a feasible limit point of a sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by the augmented Lagrangian algorithm. If $\bar{x}$ satisfies (28) for $\Psi(x) = (1/2)\|\Pi_{\mathcal{K}^\circ}(G(x))\|^2$, then $\bar{x}$ satisfies CAKKT.*

Proof. For the sake of simplicity, we can suppose that $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$. By Theorem 3, $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k))\rangle$ converges to zero. Thus, it suffices to show that $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle$ converges to zero as well. Similarly to the proof of the previous theorem, we split this proof into two cases, depending on whether $\{\rho_k\}_{k\in\mathbb{N}}$ is a bounded sequence or not:

1. Suppose that $\{\rho_k\}_{k\in\mathbb{N}}$ is unbounded. We start by showing that $\rho_k\|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \to 0$.
From the generalized Łojasiewicz inequality, there exists some function ψ such that

$$\frac{1}{2}\|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 = |\Psi(x^k)| \le \psi(x^k)\|D\Psi(x^k)\| = \psi(x^k)\|DG(x^k)^*\Pi_{\mathcal{K}^\circ}(G(x^k))\|,$$

and we obtain the following inequality

$$\rho_k \|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \leq 2\psi(x^k)\|\rho_k DG(x^k)^* \Pi_{\mathcal{K}^\circ}(G(x^k))\|.$$

Now, we proceed by finding an upper bound for the sequence $\|\rho_k DG(x^k)^* \Pi_{\mathcal{K}^\circ}(G(x^k))\|$:

$$\begin{aligned}\|\rho_k DG(x^k)^* \Pi_{\mathcal{K}^\circ}(G(x^k))\| &\leq \|DG(x^k)^*(\Lambda^k - \rho_k \Pi_{\mathcal{K}^\circ}(G(x^k)))\| + \|DG(x^k)^*\Lambda^k\|\\ &\leq \|DG(x^k)^*\|\,\|\Lambda^k - \rho_k\Pi_{\mathcal{K}^\circ}(G(x^k)))\| + \|DG(x^k)^*\Lambda^k\|\\ &\leq \|DG(x^k)^*\|\,\|\Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k) - \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)))\| + \|DG(x^k)^*\Lambda^k\|.\end{aligned}$$

Furthermore, from $\|\Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k) - \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)))\| \leq \|\widehat{\Lambda}^k\|$, we see that

$$\begin{aligned}\|\rho_k DG(x^k)^* \Pi_{\mathcal{K}^\circ}(G(x^k))\| &\leq \|DG(x^k)^*\|\,\|\widehat{\Lambda}^k\| + \|DG(x^k)^*\Lambda^k\|\\ &\leq \|DG(x^k)^*\|\,\|\widehat{\Lambda}^k\| + \|\nabla f(x^k)\| + \varepsilon_k,\end{aligned} \tag{29}$$

where, in the second inequality, we use that $\|\nabla f(x^k) + DG(x^k)^*\Lambda^k\| \leq \varepsilon_k$ (step 1 of Algorithm 1). Thus, (29) is bounded by some scalar $M > 0$ because of the continuity of DG and ∇f near $\bar{x}$. Thus, $\rho_k\|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \leq 2\psi(x^k)M$. Using the fact that $\psi(x^k) \to 0$, we get that $\rho_k\|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \to 0$. We proceed by computing $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle$. Indeed,

$$\begin{aligned}\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle &= \langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k), \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle\\ &= \langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k) - \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)), \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle +\\ &\quad + \langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)), \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle\\ &= \langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k) - \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)), \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle +\\ &\quad + \rho_k\|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2.\end{aligned} \tag{30}$$

Because $\rho_k\|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 \to 0$, we only need to show that the first expression of (30) goes to zero. Now, because $\Pi_{\mathcal{K}^\circ}(G(x^k)) \to \Pi_{\mathcal{K}^\circ}(G(\bar{x})) = 0$ and from the boundedness of $\{\widehat{\Lambda}^k\}$, we get that

$$|\langle \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k) + \widehat{\Lambda}^k) - \Pi_{\mathcal{K}^\circ}(\rho_k G(x^k)), \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle| \leq \|\widehat{\Lambda}^k\|\,\|\Pi_{\mathcal{K}^\circ}(G(x^k))\| \to 0.$$

Thus, $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle \to 0$, and as a consequence, CAKKT holds at $\bar{x}$.

2. Suppose that $\{\rho_k\}_{k\in\mathbb{N}}$ is a bounded sequence. By the proof of Theorem 3, we see that $\{\Lambda^k\}_{k\in\mathbb{N}}$ is a bounded sequence, and hence, $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle$ goes to zero because $\Pi_{\mathcal{K}^\circ}(G(x^k)) \to \Pi_{\mathcal{K}^\circ}(G(\bar{x})) = 0$.

In both cases, we show that $\bar{x}$ is a CAKKT point associated with $\{x^k\}_{k\in\mathbb{N}}$. □

The augmented Lagrangian method presented in Andreani et al. [2, algorithm 1] for NLSDP is proven to generate TAKKT sequences under generalized Łojasiewicz (Andreani et al. [2, theorem 5.2]). Hence, Theorem 4 improves this result not only in terms of generality, but also in terms of refinement of the convergence theory. If (28) is not satisfied, then Algorithm 1 may generate sequences that do not satisfy CAKKT as is shown by the counterexample after Andreani et al. [3, theorem 5.1] (note that Algorithm 1 is a direct generalization of the augmented Lagrangian presented in Andreani et al. [3]).

For completing the convergence theory of Algorithm 1, it is necessary to know how likely it is to reach feasible points. In fact, even though one cannot guarantee that every limit point of the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by Algorithm 1 is always feasible, at least it is possible to prove that it has the tendency of finding feasible points in the following sense:

Proposition 1. *Every limit point $\bar{x}$ of a sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by Algorithm 1 is a stationary point of*

$$\underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad \|\Pi_{\mathcal{K}^\circ}(G(x))\|^2. \tag{31}$$

Proof. Taking a subsequence if necessary, suppose that $x^k \to \bar{x}$. If $\{\rho^k\}_{k\in\mathbb{N}}$ is bounded, it must converge to some $\bar{\rho}$ and then $\widehat{\Lambda}^k \to \widehat{\Lambda}$. Also, in this case $V^k \to 0$, which means $\widehat{\Lambda} = \Pi_{\mathcal{K}^\circ}(\widehat{\Lambda} + \bar{\rho}G(\bar{x}))$. Then, by Lemma 1, item 1, we get $\bar{\rho}G(\bar{x}) \in \mathcal{K}$, so $\Pi_{\mathcal{K}^\circ}(G(\bar{x})) = 0$ and $\bar{x}$ is a global solution of (31). On the other hand, if $\{\rho^k\}_{k\in\mathbb{N}}$ is unbounded, by step 1 and (20), note that

$$\frac{1}{\rho^k}(\nabla f(x^k) + DG(x^k)^*\Lambda^k) = \frac{\nabla f(x^k)}{\rho^k} + DG(x^k)^*\Pi_{\mathcal{K}^\circ}\left(\frac{\widehat{\Lambda}^k}{\rho^k} + G(x^k)\right) \to 0.$$

Then, because $\nabla f(x^k) \to \nabla f(\bar{x})$ and $\{\widehat{\Lambda}^k\}_{k\in\mathbb{N}}$ is bounded, we obtain $DG(\bar{x})^*\Pi_{\mathcal{K}^\circ}(G(\bar{x})) = 0$, which means $\bar{x}$ is a stationary point of (31). □

Remark 2. We highlight that our convergence theory for Algorithm 1 allows the set $\mathcal{K}$ to have an empty interior, and it does not demand self-duality. For instance, it can be applied to optimization problems involving the classical set of *Euclidean distance matrices* (EDM) of dimension m, which is defined as

$$\mathcal{E}^m := \left\{M \in \mathbb{S}^m : \exists p_1, \ldots, p_m \in \mathbb{R}^r,\ \forall i,j \in \{1,\ldots,m\}, M_{ij} = \|p_i - p_j\|_2^2\right\}.$$

It is a closed convex cone because it can be seen as the image of $\mathbb{S}^m_+$ through the linear operator

$$\mathcal{T}(Y) := diag(Y)e^T + e\,diag(Y)^T - 2Y,$$

where e is a vector of ones and $diag(M) := (M_{11}, \ldots, M_{mm})$. Also, $\mathcal{E}^m$ is not self-dual. Because every $M \in \mathcal{E}^m$ is *hollow*, that is, $diag(M) = 0$, the EDM cone has an empty interior. In this case, it is possible to build a general convergence theory of algorithms over the EDM cone via sequential conditions even when Robinson's CQ does not hold. In particular, algorithms based on projections onto $\mathcal{E}^m$ or its polar, such as the augmented Lagrangian method we propose, can benefit from the recent advances toward efficient numerical methods to compute projections while keeping a fixed *embedding dimension* (Qi and Yuan [66]). More details about Euclidean distance matrices can be found in Krislock and Wolkowicz [51].

5. The AKKT Condition

Until this point, we have presented generalizations of CAKKT and AGP via projections onto $\mathcal{K}$ and $\mathcal{K}^\circ$, but we have not addressed yet a generalization of AKKT, which is the most natural and simple condition in NLP. However, in Sections 3 and 4, everything was built starting from TAKKT instead of AKKT. Historically, AKKT was born in NLP as a natural way of representing limit points of sequences generated by algorithms and studying their properties. But, in NCP, we present it arising from a much more theoretical field, which is the theory of perturbations in optimization problems.

Let us recall the KKT conditions at a point $\bar{x} \in \mathbb{R}^n$ with a multiplier $\overline{\Lambda} \in \mathcal{K}^\circ$ in the form of a generalized equation (in the sense of Robinson [71]):

$$\underbrace{\begin{pmatrix} \nabla f(\bar{x}) + DG(\bar{x})^*\overline{\Lambda} \\ G(\bar{x}) \end{pmatrix}}_{\mathcal{F}(\bar{x},\overline{\Lambda})} \in \underbrace{\begin{bmatrix} \{0\} \\ \{Y \in \mathcal{K} : \langle Y, \overline{\Lambda}\rangle = 0\} \end{bmatrix}}_{\mathcal{N}(\overline{\Lambda})}.$$

Given some $\varepsilon > 0$, the standard perturbation theory (see, for example, Bonnans and Shapiro [25], Hager and Gowda [44], Hager and Mico-Umutesi [45]) can be used as inspiration to say that a point $x \in B(\bar{x}, \varepsilon)$ satisfies the KKT conditions with error ε when there is a multiplier $\Lambda \in \mathcal{K}^\circ$ and some perturbation vector $\xi \in \mathbb{R}^n \times \mathbb{E}$ such that $\mathcal{F}(x,\Lambda) + \xi \in \mathcal{N}(\Lambda)$ and $\|\xi\| \leq \varepsilon$. This strongly suggests a sequential optimality condition:

Definition 4 (AKKT). A feasible point $\bar{x}$ satisfies the AKKT condition when there exist sequences $\{y^k\}_{k\in\mathbb{N}} \to 0$, $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$, and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathcal{K}^\circ$ such that (9) holds, $G(x^k) + y^k \in \mathcal{K}$, and

$$\langle \Lambda^k, G(x^k) + y^k\rangle = 0,\ \forall k \in \mathbb{N}. \tag{32}$$

It turns out that Definition 4 coincides with Qi and Wei [67, definition 2.5], which, to the best of our knowledge, employed for the first time perturbed KKT ideas to improve global convergence results of algorithms (for instance, SQP methods). At the time, the authors did not prove it was an optimality condition. Note that AKKT as in Definition 4 is distinguished for not directly relying on projections, eigenvalues, or other similar objects, but giving some degree of freedom to the approximation instead, makes it much more versatile and simple than the others. On the other hand,because the perturbation is inside the inner product, AKKT has a more solid structure to work with when compared with the others. Also, when Definition 4 is specialized to the NLP, the NLSDP, or the NSOCP contexts, it is consistent with the existing concepts with the same name from Andreani et al. [1, section 2; 2, definition 3.1; 12, definition 3.3], respectively. This is clarified in Section 6.1.

For relating AKKT with the other conditions in NCP, we begin by proving that AGP implies AKKT:

Proposition 2. *If $\bar{x}$ satisfies the AGP condition, then it must also satisfy AKKT.*

Proof. If $\bar{x}$ satisfies AGP, then there are sequences $\{\tilde{\Lambda}^k\}_{k\in\mathbb{N}} \subset \mathcal{K}^\circ$ and $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ satisfying (9) and (11). Denote by y^k the global solution of

$$\begin{aligned} \underset{y\in\mathbb{E}}{\text{Minimize}} \quad & \frac{1}{2}\|y\|^2 - \langle \tilde{\Lambda}^k, G(x^k) + y\rangle, \\ \text{subject to} \quad & G(x^k) + y \in \mathcal{K} \end{aligned} \tag{33}$$

and consider the feasible point $y := -\Pi_{\mathcal{K}^\circ}(G(x^k))$ of (33). Then,

$$\begin{aligned} (1/2)\|y^k\|^2 - \langle \tilde{\Lambda}^k, G(x^k) + y^k\rangle &\le (1/2)\|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 - \langle \tilde{\Lambda}^k, G(x^k) - \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle \\ &= (1/2)\|\Pi_{\mathcal{K}^\circ}(G(x^k))\|^2 - \langle \tilde{\Lambda}^k, \Pi_{\mathcal{K}}(G(x^k))\rangle. \end{aligned} \tag{34}$$

Taking $k \to \infty$ in (34), because $\Pi_{\mathcal{K}^\circ}(G(x^k)) \to 0$ and AGP holds, we see that $y^k \to 0$ and $\langle \tilde{\Lambda}^k, G(x^k) + y^k\rangle \to 0$ because $\tilde{\Lambda}^k \in \mathcal{K}^\circ$ and $G(x^k) + y^k \in \mathcal{K}$. Now, the necessary optimality condition for y^k in (33) implies

$$0 \in y^k - \tilde{\Lambda}^k + \{\Theta \in \mathcal{K}^\circ : \langle \Theta, G(x^k) + y^k\rangle = 0\}.$$

Now, choosing $\Lambda^k := \tilde{\Lambda}^k - y^k$ for all $k \in \mathbb{N}$, we obtain that AKKT holds at $\bar{x}$ with the sequences $\{x^k\}_{k\in\mathbb{N}}$, $\{\Lambda^k\}_{k\in\mathbb{N}}$, and $\{y^k\}_{k\in\mathbb{N}}$. □

Though, the converse is not necessarily true because of Andreani et al. [1, counterexample 3.1], which shows that AKKT does not imply AGP in NLP (and Theorem 1). Also, Andreani et al. [2, example 5.2; 12, example 3.1] show that TAKKT and AKKT do not imply each other in NLP; hence, the same is valid for NCP.

Because AKKT is a weak condition, in comparison with AGP and CAKKT, we expect it to be more commonly generated by algorithms designed for solving (NCP). An interesting fact is that the vector y^k of Definition 4 formalizes the idea of seeing joint feasibility–complementarity measures, which are commonly used in algorithms, as certificates of approximate optimality along with the Lagrangian residue. For example, it is easy to verify that Algorithm 1 generates sequences $\{x^k\}_{k\in\mathbb{N}}$ and $\{\Lambda^k\}_{k\in\mathbb{N}}$ such that every feasible limit point $\bar{x}$ of $\{x^k\}_{k\in\mathbb{N}}$ satisfies AKKT. In this case, without loss of generality, we can assume $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$. Then, AKKT can be verified by taking the sequence $\{y^k\}_{k\in\mathbb{N}}$ such that $y^k = V^k$ for every k; then, $G(x^k) + y^k \in \mathcal{K}$ and

$$\langle \Lambda^k, G(x^k) + y^k\rangle = \rho_k \left\langle \Pi_{\mathcal{K}^\circ}\left(G(x^k) + \frac{\widehat{\Lambda}^k}{\rho_k}\right), \left(G(x^k) + \frac{\widehat{\Lambda}^k}{\rho_k}\right) - \Pi_{\mathcal{K}^\circ}\left(G(x^k) + \frac{\widehat{\Lambda}^k}{\rho_k}\right)\right\rangle = 0,$$

by Lemma 1. Condition (9) follows directly from step 1 because $\nabla_x L_{\rho^k}(x^k, \Lambda^k) = \nabla f(x^k) + DG(x^k)^* \Lambda^k$ for the multiplier choice of step 2. Also, if $\{\rho_k\}_{k\in\mathbb{N}}$ is unbounded, it follows that $y^k \to -\Pi_{\mathcal{K}^\circ}(G(\bar{x})) = 0$; if it is bounded, then $\|y^k\| \to 0$ because of step 4. Thus, it can be said that AKKT is the most simple and natural sequential condition among the ones we present when viewed from the algorithmic and theoretical perspectives.

When $\mathcal{K}$ is a product of closed convex cones, say $\mathcal{K} = \mathcal{K}_1 \times \ldots \times \mathcal{K}_r$ and $\mathbb{E} = \mathbb{E}_1 \times \ldots \times \mathbb{E}_m$, where $\mathcal{K}_i \subset \mathbb{E}_i$ for all $i \in \{1, \ldots, r\}$, it is possible to see that Definition 4 resembles the classical AKKT from NLP via the following lemma:

Lemma 2. *Let $\bar{x}$ be a feasible point of (NCP). If $\bar{x}$ satisfies AKKT, then there are sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathcal{K}^\circ$ such that (9) holds and $\Lambda_i^k = 0$ whenever $G_i(\bar{x}) \in \operatorname{int} \mathcal{K}_i$ for sufficiently large k.*

Proof. Let $\bar{x}$ be an AKKT point associated with the sequences $\{(y_1, \ldots, y_r)^k\}_{k\in\mathbb{N}} \to 0$, $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$, and $\{(\Lambda_1, \ldots, \Lambda_r)^k\}_{k\in\mathbb{N}} \subset \mathcal{K}^\circ$. Proving (9) is trivial. Then, for every index $i \in \{1, \ldots, r\}$ such that $G_i(\bar{x}) \in \operatorname{int} \mathcal{K}_i$, there is some $k_0 \in \mathbb{N}$ such that $z_i^k := G_i(x^k) + y_i^k \in \operatorname{int} \mathcal{K}_i$ for every $k > k_0$. Hence, for every such k and i, there is some $\alpha_i^k > 0$ such that $z_i^k + \alpha_i^k \Lambda_i^k \in \mathcal{K}_i$, so we have $0 \ge \langle \Lambda_i^k, z_i^k + \alpha_i^k \Lambda_i^k\rangle = \alpha_i^k \|\Lambda_i^k\|^2$. Thus, $\Lambda_i^k = 0$ when $G_i(\bar{x}) \in \operatorname{int} \mathcal{K}_i$ for all $k > k_0$. □

Note that the converse holds in NLP when $\mathcal{K} = \mathbb{R}_+^m$ is seen as the Cartesian product of m copies of $\mathbb{R}_+$. Therefore, Definition 4 is consistent with the usual form of AKKT in NLP. In the next section, we contextualize the other conditions in other classical particular cases of (NCP) as well.

Remark 3. In Steck's [75] PhD thesis, it is introduced a new sequential optimality condition, which is further developed in Börgens et al. [26, definitions 3.1 and 3.2]. In view of Börgens et al. [26, remark 4.1], because $\mathbb{E}$ is finite-dimensional and $\mathcal{K}$ is a closed convex cone, their condition can be defined for (NCP) similarly to Definition 1, but replacing (10) by

$$\liminf_{k\to\infty} \langle \Lambda^k, G(x^k)\rangle \geq 0. \tag{35}$$

Following Börgens et al. [26], we call this condition s-AKKT. It is immediate to see that s-AKKT is implied by TAKKT. Moreover, if AGP holds at a feasible point $\bar{x}$ and $\mathcal{K}$ is self-dual, then Moreau's decomposition tells us that

$$\liminf_{k\to\infty} \langle \Lambda^k, G(x^k)\rangle = \liminf_{k\to\infty} -\langle \Lambda^k, \Pi_{\mathcal{K}}(-G(x^k))\rangle \geq 0.$$

Then, if $\mathcal{K}$ is self-dual, AGP implies s-AKKT. However, because AGP and TAKKT are independent, then s-AKKT is strictly implied by both of them in this case. Because TAKKT does not imply AKKT, then s-AKKT does not imply AKKT either. Conversely, Example 2 shows that AKKT does not imply s-AKKT; in fact, any sequence that satisfies (17) must have $\liminf_{k\to\infty} \Lambda^k G(x_1^k, x_2^k) = \Lambda^k x_2^k h(x_1^k) \to -1 < 0$ because $x_2^k \to 1$. In summary, if $\mathcal{K}$ is self-dual, s-AKKT is strictly weaker than TAKKT, strictly weaker than AGP, but independent of AKKT. If $\mathcal{K}$ is not self-dual, Example 2 also shows that AGP is independent of s-AKKT, but the other relations still hold. It is worth mentioning that all notions of convergence considered in Börgens et al. [26] are equivalent in a finite dimensional setting, which is why we only address s-AKKT.

The relations between the new sequential optimality conditions of this paper are presented in Figure 1, in which the arrows indicate (strict) implications.

6. Contextualization in Some Particular Cases of NCP

In this section, we specialize CAKKT, AGP, and AKKT in the contexts of NLP, NLSDP, and NSOCP to illustrate the stopping criteria associated with each of them in more practical terms. The TAKKT condition does not acquire any specific format when reduced to any context, so it is not included in this section.

6.1. Nonlinear Programming

Consider the standard nonlinear programming problem with q inequality constraints and p equality constraints:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & h(x) = 0, \\ & g(x) \leq 0. \end{aligned} \tag{NLP}$$

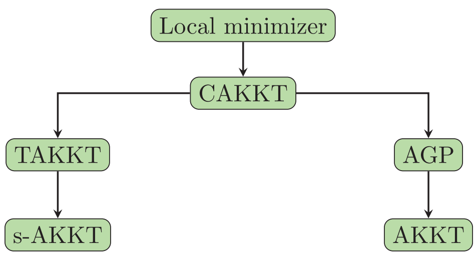
The most straightforward way of viewing (NLP) as a particular case of (NCP) is directly phrasing the constraints as $(h(x), g(x)) \in \{0\}^p \times \mathbb{R}^q_-$, but Moreau's decomposition is meaningless when $\mathcal{K} = \{0\}^p$, and CAKKT reduces to TAKKT in this case. However, the CAKKT from Andreani et al. [3] is strictly stronger than TAKKT, which indicates that this is not the best way of modeling (NLP) as a particular case of (NCP) concerning CAKKT. The AKKT and AGP conditions do not present the same behavior. It is noteworthy that if one considers (NCP) with separate equality constraints, that is, with two blocks of constraints: $h(x) = 0$ and $G(x) \in \mathcal{K}$, then every definition from this paper can be straightforwardly extended to this new problem and, in this case, $\mathcal{K}$ can be taken as $\mathbb{R}^q_-$. Then, we consider an auxiliary formulation of (NLP), in which equality constraints are treated as two inequalities:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & G(x) := (h(x), -h(x), g(x)) \in \mathbb{R}^p_- \times \mathbb{R}^p_- \times \mathbb{R}^q_-. \end{aligned} \tag{NLP-Auxiliary}$$

The Lagrange multipliers are then written as

$$\Lambda^k = \left(\omega^k_{1+}, \ldots, \omega^k_{p+}, \omega^k_{1-}, \ldots, \omega^k_{p-}, \mu^k_1, \ldots, \mu^k_q\right) \in \mathbb{R}^p_+ \times \mathbb{R}^p_+ \times \mathbb{R}^q_+.$$

Figure 1. (Color online) Relationship of the new sequential optimality conditions.



Our next result concerns the specializations of CAKKT, AGP, and AKKT from (NCP) to (NLP), but we use (NLP-Auxiliary) as an intermediary step for recovering the original (NLP) definitions of such conditions.

Proposition 3. *Let $\bar{x}$ be a feasible point of (NLP). Then, $\bar{x}$ satisfies*

a. *AKKT for (NLP-Auxiliary) if and only if $\mu_j^k = 0$ whenever $g_j(\bar{x}) < 0$ for every sufficiently large k and some sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$, $\{\omega^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p$, and $\{\mu^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^q_+$ such that*

$$\nabla f(x^k) + \sum_{i=1}^{p} \omega_i^k \nabla h_i(x^k) + \sum_{j=1}^{q} \mu_j^k \nabla g_j(x^k) \to 0; \tag{36}$$

b. *AGP for (NLP-Auxiliary) if and only if $\mu_j^k \min\{0, g_j(x^k)\} \to 0$ and $\liminf_{k\to\infty} \omega_i^k h_i(x^k) \geq 0$ for every $i \in \{1,\ldots,p\}$ and every $j \in \{1,\ldots,q\}$; for some sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$, $\{\mu^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^q_+$, and $\{\omega^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p$ such that (36) holds;*

c. *CAKKT for (NLP-Auxiliary) if and only if $\mu_j^k g_j(x^k) \to 0$ and $\omega_i^k h_i(x^k) \to 0$ for every $i \in \{1,\ldots,p\}$ and every $j \in \{1,\ldots,q\}$; for some sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$, $\{\mu^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^q_+$, and $\{\omega^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p$ such that (36) holds.*

Proof.

a. AKKT: If AKKT holds for (NLP-Auxiliary), there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to x^k$, $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathcal{K}^\circ$, and $\{Y^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^q$, where $Y^k = (z_+^k, z_-^k, y^k) \to 0$ such that $h(x^k) + z_+^k \leq 0$, $-h(x^k) + z_-^k \leq 0$, and $g(x^k) + y^k \leq 0$ for every k, and

$$\langle G(x^k) + Y^k, \Lambda^k \rangle = \sum_{i=1}^{p} \omega_{i+}^k (h_i(x^k) + z_{i+}^k) + \sum_{i=1}^{p} \omega_{i-}^k (-h_i(x^k) + z_{i-}^k) + \sum_{j=1}^{q} \mu_j^k (g_j(x^k) + y_j^k) = 0. \tag{37}$$

Because all terms are nonnegative, they are all zero, and consequently, whenever $g_j(\bar{x}) < 0$, we have $g_j(x^k) + y_j^k < 0$ for sufficiently large k. That implies $\mu_j^k = 0$ for every such k and j. Take $\omega_i^k := \omega_{i+}^k - \omega_{i-}^k$ to obtain (36) from (9).
Conversely, consider the following natural choice: $Y^k := (-h(x^k), h(x^k), \widehat{g}_j(x^k)) \to 0,$ where

$$\widehat{g}_j(x^k) := \begin{cases} 0, \text{if } g_j(\bar{x}) < 0 \\ -g_j(x^k), \text{ otherwise,} \end{cases}$$

and then, for sufficiently large k, we have (37) and $G(x^k) + Y^k \in \mathcal{K}$, and AKKT is satisfied with $\omega_{i+}^k := \max\{0, \omega_i^k\}$ and $\omega_{i-}^k := -\min\{0, \omega_i^k\}$.

b. AGP: In this case, $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle \to 0$ is equivalent to saying that

$$\sum_{i=1}^{p} \omega_{i+}^k \min\{0, h_i(x^k)\} + \sum_{i=1}^{p} \omega_{i-}^k \min\{0, -h_i(x^k)\} + \sum_{j=1}^{q} \mu_j^k \min\{0, g_j(x^k)\} \to 0.$$

Because each part of the sum has the same sign, we get that $\omega_{i-}^k \min\{0, -h_i(x^k)\} \to 0$, $\omega_{i+}^k \min\{0, h_i(x^k)\} \to 0$, and $\mu_j^k \min\{0, g_j(x^k)\} \to 0$. Hence, we have that $\mu_j^k \min\{0, g_j(x^k)\} \to 0$ for every $j \in \{1,\ldots,q\}$, and defining $\omega_i^k := \omega_{i+}^k - \omega_{i-}^k$, we observe that

$$\omega_i^k h_i(x^k) = \omega_{i+}^k \min\{0, h_i(x^k)\} - \omega_{i-}^k \min\{0, h_i(x^k)\} - \omega_{i+}^k \min\{0, -h_i(x^k)\} + \omega_{i-}^k \min\{0, -h_i(x^k)\}. \tag{38}$$

Then, note that the first and last terms of the right-hand side of (38) both vanish in the limit, and the middle terms are nonnegative. Hence, $\liminf_{k\to\infty} \omega_i^k h_i(x^k) \geq 0$ for every $i \in \{1,\ldots,p\}$.
Conversely, set $\omega_{i+}^k := \max\{0, \omega_i^k\}$ and $\omega_{i-}^k := -\min\{0, \omega_i^k\}$. Then, for each k, only one term of the right-hand side of (38) can be nonzero, so its first and last terms must converge to zero because they are nonpositive.

c. CAKKT: Here, we get

$$\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle = \langle \omega_+^k, \min\{h(x^k), 0\} \rangle + \langle \omega_-^k, \min\{-h(x^k), 0\} \rangle + \langle \mu^k, \min\{g(x^k), 0\} \rangle$$

and

$$\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle = \langle \omega_+^k, \max\{h(x^k), 0\} \rangle + \langle \omega_-^k, \max\{-h(x^k), 0\} \rangle + \langle \mu^k, \max\{g(x^k), 0\} \rangle,$$

so if both tend to zero, define $\widehat{\omega}^k := \omega_+^k - \omega_-^k$, and we obtain $\widehat{\omega}_i^k h_i(x^k) \to 0$ and $\mu_j^k g_j(x^k) \to 0$.

The converse is analogous to the previous item with the same choice of multipliers:

$$\omega_{i+}^k := \max\{0, \omega_i^k\} \text{ and } \omega_{i-}^k := -\min\{0, \omega_i^k\}. \quad \square$$

Note that, from this point of view, our definitions are consistent with the original AKKT from Andreani et al. [1], the AGP from Martínez and Svaiter [60] (which follows from Theorem 1), and the CAKKT from Andreani et al. [3], respectively.

We summarize our results in Table 1.

Because the sequential conditions mostly differ in how they deal with approximate complementarity, only this condition is made explicit in Table 1.

6.2. Nonlinear Semidefinite Programming

Here, we recall the classical form of an NLSDP problem:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & G(x) \in \mathbb{S}^m_-, \end{aligned} \qquad \text{(NLSDP)}$$

which is (NCP) with $\mathbb{E} = \mathbb{S}^m$, $\langle A, B\rangle := \text{tr}(AB)$ is the (Frobenius) inner product given by the trace of AB, and $\mathcal{K} = \mathbb{S}^m_-$. We recall that $\mathcal{K}^\circ = -\mathcal{K}$ and every symmetric matrix $A \in \mathbb{E}$ has a spectral decomposition, that is, there exists an orthogonal matrix U such that $A = UDU^T$, where $D = \text{Diag}\,(\lambda_1^U(A), \dots, \lambda_m^U(A))$ is a diagonal matrix and $\lambda_i^U(A)$ are eigenvalues of A ordered according to the eigenvectors in the columns of U. Moreover,

$$\Pi_{\mathcal{K}}(A) = U\text{Diag}(\min\{\lambda_1^U(A), 0\}, \dots, \min\{\lambda_m^U(A), 0\})U^T,$$

and a similar relation holds for $\Pi_{\mathcal{K}^\circ}(A)$ with max instead of min.

When no order is specified, we consider $\lambda_1(A) \le \dots \le \lambda_m(A)$. Note that for every $i \in \{1, \dots, m\}$, we have $\lambda_i(-A) = -\lambda_{m-i+1}(A)$. Also, the following inequality is important to our analyses: for every $A, B \in \mathbb{S}^m_-$, we have the inequality

$$\sum_{i=1}^m \lambda_i(A)\lambda_{m-i+1}(B) \le \text{tr}(AB) \le \sum_{i=1}^m \lambda_i(A)\lambda_i(B). \qquad (39)$$

For its proof see Marcus [57].

Now, we specialize our conditions from (NCP) to (NLSDP):

Proposition 4. *Let $\bar{x}$ be a feasible point of (NLSDP). Then, $\bar{x}$ satisfies*

a. *AKKT if and only if there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$, $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathcal{K}^\circ$, and a sequence of orthogonal matrices $S^k \to U$, where U diagonalizes $G(\bar{x})$ and each S^k diagonalizes Λ^k such that (9) holds and $\lambda_i^{S^k}(\Lambda^k) = 0$ if $\lambda_i^U(G(\bar{x})) < 0$ for sufficiently large k.*

b. *AGP implies that, for every i,*

$$\min\{0, \lambda_i(G(x^k))\}\lambda_i(\Lambda^k) \to 0,$$

for some sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathcal{K}^\circ$ such that (9) holds.

c. *CAKKT implies that, for every i,*

$$\min\{0, \lambda_i(G(x^k))\}\lambda_i(\Lambda^k) \to 0 \text{ and } \min\{0, -\lambda_{m-i+1}(G(x^k))\}\lambda_i(\Lambda^k) \to 0,$$

for some sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathcal{K}^\circ$ such that (9) holds.

Proof.

a. AKKT: If there is a sequence $\{y^k\}_{k\in\mathbb{N}} \to 0$ such that $G(x^k) + y^k \in \mathcal{K}$ and $\langle \Lambda^k, G(x^k) + y^k\rangle = 0$ for every k, the latter implies that $G(x^k) + y^k$ and Λ^k are simultaneously diagonalizable; that is, for every k, there is a matrix S^k such that $G(x^k) + y^k = S^k\Theta^k(S^k)^T$ and $\Lambda^k = S^k\Gamma^k(S^k)^T$, where $\Theta^k = Diag(\lambda_i^{S^k}(G(x^k) + y^k))$ and $\Gamma^k = Diag(\lambda_i^{S^k}(\Lambda^k))$. The continuity of G and the convergence of $\{x^k\}_{k\in\mathbb{N}}$ imply that $S^k \to U$ for some orthogonal matrix U. Then, U diagonalizes $G(\bar{x})$ because $G(x^k) + y^k \to G(\bar{x})$, and if $\lambda_i^U(G(\bar{x})) < 0$, then for sufficiently large k, we have $\lambda_i^{S^k}(G(x^k) + y^k) < 0$ as well. Then, $\lambda_i^{S^k}(\Lambda^k) = 0$ for those k. Conversely, let $\bar{x}$ be a feasible point associated with the sequences $\{\Lambda^k\}_{k\in\mathbb{N}}$, $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$, and $\{S^k\}_{k\in\mathbb{N}} \to U$, where each S^k diagonalizes Λ^k and U diagonalizes $G(\bar{x})$ such that $\lambda_i^{S^k}(\Lambda^k) = 0$

Table 1. Sequential conditions when specialized to (NLP).

	Approximate complementarity condition
AKKT	$g_j(\bar{x}) < 0 \Rightarrow \mu_j^k = 0$, $\forall j$ and k sufficiently large
AGP	$\mu_j^k \min\{g_j(x^k), 0\} \to 0$ and $\liminf_{k\to\infty} \omega_i^k h_i(x^k) \ge 0$, $\forall i$, $\forall j$
CAKKT	$\mu_j^k g_j(x^k) \to 0$ and $\omega_i^k h_i(x^k) \to 0$, $\forall i$, $\forall j$.

whenever $\lambda_i^U(G(\bar{x})) < 0$. Without loss of generality and to simplify the notation, we can suppose that U leaves the eigenvalues of $G(\bar{x})$ increasingly ordered. Also, suppose that there are α negative eigenvalues and β zero eigenvalues in $G(\bar{x})$ with $\alpha + \beta = m$, and for some indexed matrix A^k, define $\tilde{A}^k := (S^k)^T A^k S^k$. Then, choose

$$y^k := S^k \begin{bmatrix} 0 & -\tilde{G}(x^k)_{\alpha\beta} \\ -\tilde{G}(x^k)_{\beta\alpha} & -\tilde{G}(x^k)_{\beta\beta} \end{bmatrix} (S^k)^T, \text{ where } \tilde{G}(x^k) = \begin{bmatrix} \tilde{G}(x^k)_{\alpha\alpha} & \tilde{G}(x^k)_{\alpha\beta} \\ \tilde{G}(x^k)_{\beta\alpha} & \tilde{G}(x^k)_{\beta\beta} \end{bmatrix},$$

and the partition $\alpha\alpha$ refers to the first α rows of the matrix, for example. Then, if $\Gamma^k := \tilde{\Lambda}^k$, we get

$$\langle \Lambda^k, G(x^k) + y^k \rangle = \langle \Gamma^k, \tilde{G}(x^k) + \tilde{y}^k \rangle = 0,$$

for sufficiently large k because $\Gamma^k_{\alpha\alpha} = 0$ for large enough k. Because every block of $\tilde{G}(x^k)$ converges to zero except for $\tilde{G}(x^k)_{\alpha\alpha}$, we know that $y^k \to 0$. Also, because $\tilde{G}(x^k) + \tilde{y}^k \in \mathbb{S}^m_-$, we get $G(x^k) + y^k \in \mathbb{S}^m_-$.

b. AGP: In this case, observe that $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle = \text{tr}(\Lambda^k \Pi_{\mathcal{K}}(G(x^k))) \to 0$ can be simplified using the left inequality of (39) for $A = \Pi_{\mathcal{K}}(G(x^k))$, $B = -\Lambda^k$. With this, we obtain $\lambda_i(\Pi_{\mathcal{K}}(G(x^k)))\lambda_{m-i+1}(-\Lambda^k) \to 0$ for all i. Because $\lambda_i(\Lambda^k) = -\lambda_{m-i+1}(-\Lambda^k)$, we conclude that AGP implies $\lambda_i(\Pi_{\mathcal{K}}(G(x^k)))\lambda_i(\Lambda^k) \to 0$, for every $i \in \{1, \ldots, m\}$. This means

$$\min\{0, \lambda_i(G(x^k))\}\lambda_i(\Lambda^k) \to 0.$$

c. CAKKT: Similarly to the previous item, from $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle \to 0$ we get $\min\{0, \lambda_i(G(x^k))\}\lambda_i(\Lambda^k) \to 0$. Also, from $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle \to 0$ and (39), we obtain $\lambda_i(-\Pi_{\mathcal{K}^\circ}(G(x^k)))\lambda_i(\Lambda^k) \to 0$, for every $i \in \{1, \ldots, m\}$, which is equivalent to $\lambda_i(\Pi_{\mathcal{K}}(-G(x^k)))\lambda_i(\Lambda^k) \to 0$. Then,

$$\min\{0, -\lambda_{m-i+1}(G(x^k))\}\lambda_i(\Lambda^k) \to 0. \quad \square$$

Note that the characterizations for AGP and CAKKT are unilateral in this case, but because the purpose of specializing our conditions is to define stopping criteria, this is not an issue. For instance, if an algorithm employs the stopping criterion related to AGP (given by item (b) of Proposition 4), the proposition states that its feasible limit points are at least as good as AGP. Nevertheless, we point out that the converse statements hold when Λ^k and $G(x^k)$ commute for every k.

We summarize our results for (NLSDP) in Table 2.

Recall from Definition 3 that CAKKT incorporates the idea of controlling the behavior of the Lagrange multiplier through a vanishing measure of infeasibility. In NLP, this control can be understood in terms of growth, but in more general contexts, such as NLSDP, it can have different meanings. As mentioned in the introduction, Andreani et al. [2] conjecture that the ideal definition of CAKKT should control the growth of all eigenvalues of the multiplier. In our case, only $\max\{0, m - 2r\}$ eigenvalues have their growths controlled, and r is the number of nonzero eigenvalues of $G(\bar{x})$. This suggests that, even though our definition of CAKKT generalizes one of the multiple interpretations of the nonlinear programming CAKKT, it is still imperfect. We conjecture that our definition, the one presented in Andreani et al. [12], and (6) are all independent. If this is the case, then there would be multiple correct ways of generalizing CAKKT. However, we are not able to find examples that support our claim at this moment.

That AGP was not yet defined for NLSDP and AKKT is consistent with the definition presented in Andreani et al. [2, definition 3.1]. Also, employing analogous reasoning, it is possible to recover AKKT from Andreani et al. [12] in symmetric cones after imbuing a Jordan product into $\mathbb{E}$ by making use of the *spectral theorem* from Baes [17].

Table 2. Sequential optimality conditions when specialized to (NLSDP). Recall that, for $C \in \mathbb{S}^m$, the symbols $\lambda_1(C), \ldots, \lambda_m(C)$ represent the eigenvalues of C, increasingly ordered.

	Approximate complementarity condition
AKKT	$\lambda_i^U(G(\bar{x})) < 0 \Rightarrow \lambda_i^{S^k}(\Lambda^k) = 0$, $\forall i$, $\forall k$ sufficiently large,where S^k diagonalizes Λ^k, for every k, and U diagonalizes $G(\bar{x})$
AGP	$\min\{0, \lambda_i(G(x^k))\}\lambda_i(\Lambda^k) \to 0$, $\forall i$
CAKKT	$\min\{0, \lambda_i(G(x^k))\}\lambda_i(\Lambda^k) \to 0$ and $\max\{0, \lambda_{m-i+1}(G(x^k))\}\lambda_i(\Lambda^k) \to 0$, $\forall i$

6.3. Nonlinear Second Order Cone Programming

Consider the following particular case of (NCP):

$$\begin{aligned} &\underset{x\in\mathbb{R}^n}{\text{Minimize}} && f(x), \\ &\text{subject to} && G_i(x)\in\mathcal{K}_i, i\in\{1,\dots,r\}, \end{aligned} \tag{NSOCP}$$

where $\mathbb{E}=\mathbb{R}^{m_1}\times\ldots\times\mathbb{R}^{m_r}$ with $m_1+\ldots+m_r=m$, and each $K_i\subset\mathbb{R}^{m_i}$ is a *second order cone* (or *Lorentz cone*), that is,

$$\mathcal{K}_i:=\{(z_0,\widehat{z})\in\mathbb{R}\times\mathbb{R}^{m_i-1}:\|\widehat{z}\|\le z_0\},\ i\in\{1,\dots,r\}.$$

Denote $G(x):=(G_1(x),\dots,G_r(x))$ and $\mathcal{K}:=\mathcal{K}_1\times\ldots\times\mathcal{K}_r$. From Fukushima et al. [36], the interior and the boundary of $\mathcal{K}_i$ are described by

$$\begin{aligned} \text{int } \mathcal{K}_i &:= \{(z_0,\widehat{z})\in\mathbb{R}\times\mathbb{R}^{m_i-1}:\|\widehat{z}\|<z_0\},\\ \text{bd } \mathcal{K}_i &:= \{(z_0,\widehat{z})\in\mathbb{R}\times\mathbb{R}^{m_i-1}:\|\widehat{z}\|=z_0\}. \end{aligned}$$

Moreover, consider the following sets of indices:

$$\begin{aligned} \mathcal{I}_{\text{int}} &:= \{i\in\{1,\dots,r\}:G_i(\overline{x})\in\text{int }\mathcal{K}_i\},\\ \mathcal{I}_{\text{bd}_+} &:= \{i\in\{1,\dots,r\}:G_i(\overline{x})\in\text{bd }\mathcal{K}_i\backslash\{0\}\}. \end{aligned} \tag{40}$$

It is well known that $\mathcal{K}_i^\circ=-\mathcal{K}_i$, $\forall i$, and hence, $\mathcal{K}^\circ=-\mathcal{K}$. In Fukushima et al. [36], there is a formula for the projection onto a single Lorentz cone $\mathcal{K}_i$. Following Fukushima et al. [36], every $v=(v_0,\widehat{v})\in\mathbb{R}\times\mathbb{R}^{m_i-1}$ can be decomposed as

$$v=\mu_1(v)c_1(v)+\mu_2(v)c_2(v), \tag{41}$$

where $\mu_\ell(v)\in\mathbb{R}$ and $c_\ell(v)\in\mathcal{K}$ for $\ell\in\{1,2\}$ are given by the following expressions:

$$\mu_\ell(v)=v_0+(-1)^\ell\|\widehat{v}\| \text{ and } c_\ell(v)=\begin{cases}(1/2)(1,(-1)^\ell\widehat{v}\|\widehat{v}\|^{-1}) & \text{if } \widehat{v}\neq 0\\ (1/2)(1,(-1)^\ell w) & \text{if } \widehat{v}=0,\end{cases} \tag{42}$$

where w is any unitary vector in $\mathbb{R}^{m_i-1}$. Clearly, we always have that $\mu_1(v)\le\mu_2(v)$ and $0\le\langle c_i(v),c_j(w)\rangle\le 1$ for every v,w and $i,j\in\{1,2\}$. Now, for every $v=(v_0,\widehat{v})\in\mathbb{R}\times\mathbb{R}^{m_i-1}$, we have $\Pi_{\mathcal{K}_i}(v):=\max\{\mu_1(v),0\}c_1(v)+\max\{\mu_2(v),0\}c_2(v)$. Using $\mathcal{K}_i^\circ=-\mathcal{K}_i$, we obtain

$$\Pi_{\mathcal{K}_i^\circ}(v):=\min\{\mu_1(v),0\}c_1(v)+\min\{\mu_2(v),0\}c_2(v).$$

Finally, for $v=(v_1,\dots,v_r)\in\mathbb{E}$ with $v_i\in\mathbb{R}\times\mathbb{R}^{m_i-1}$, we have that $\Pi_{\mathcal{K}}(v)=(\Pi_{\mathcal{K}_1}(v_1),\dots,\Pi_{\mathcal{K}_r}(v_r))$ and $\Pi_{\mathcal{K}^\circ}(v)=(\Pi_{\mathcal{K}_1^\circ}(v_1),\dots,\Pi_{\mathcal{K}_r^\circ}(v_r))$.

We recall from Peng et al. [64, lemma 6.2.3] that, for every $z_i,v_i\in\mathbb{R}^{m_i}$,

$$\mu_1(z_i)\mu_2(v_i)+\mu_2(z_i)\mu_1(v_i)\le 2\langle z_i,v_i\rangle\le\mu_1(z_i)\mu_1(v_i)+\mu_2(z_i)\mu_2(v_i). \tag{43}$$

Now, let $\overline{x}$ be a feasible point of (NSOCP) and $\{x^k\}_{k\in\mathbb{N}}$ be a sequence with $x^k\to\overline{x}$ associated with some sequential optimality condition with multipliers $\{\Lambda^k\}_{k\in\mathbb{N}}$; then, (9) can be stated as

$$\nabla f(x^k)+\sum_{i=1}^r DG_i(x^k)^T\lambda_i^k\to 0, \tag{44}$$

where $\Lambda^k=(\lambda_1^k,\dots,\lambda_r^k)\in\mathcal{K}_1^\circ\times\ldots\times\mathcal{K}_r^\circ$.

Similarly to the previous sections, the following result exhibits the formats of our sequential condition when specialized to NSOCP:

Proposition 5. *Let $\overline{x}$ be a feasible point of (NSOCP). Then, $\overline{x}$ satisfies*

a. *AKKT if and only if there exist sequences $\{x^k\}_{k\in\mathbb{N}}\to\overline{x}$, $\{\lambda_i^k\}_{k\in\mathbb{N}}\subset\mathcal{K}^\circ$ such that (44) holds, $\lambda_i^k=0$ for every $i\in\mathcal{I}_{int}$ and sufficiently large k, and when $i\in\mathcal{I}_{bd+}$, then $-\lambda_i^k\in\text{bd }\mathcal{K}_i$ and either $\lambda_i^k\to 0$ or*

$$\frac{\widehat{\lambda_i^k}}{\|\widehat{\lambda_i^k}\|}\to\frac{\widehat{G}_i(\overline{x})}{\|\widehat{G}_i(\overline{x})\|}; \tag{45}$$

b. *AGP implies that*

$$\mu_\ell(\lambda_i^k)\max\{\mu_\ell(G_i(x^k)),0\}\to 0,$$

for every $\ell\in\{1,2\}$ and every $i\in\{1,\dots,r\}$ for some sequences $\{x^k\}_{k\in\mathbb{N}}\to\overline{x}$ and $\{\lambda_i^k\}_{k\in\mathbb{N}}\subset\mathcal{K}^\circ$ such that (44) holds;

c. *CAKKT implies that*

$$\mu_\ell(\lambda_i^k)\max\{\mu_\ell(G_i(x^k)),0\}\to 0,\ \ell\in\{1,2\}\ \text{and}\ \ \mu_\ell(\lambda_i^k)\ \min\{\mu_s(G_i(x^k)),0\}\to 0,$$

for every $\ell, s\in\{1,2\}$ *with* $\ell\neq s$, *and every* $i\in\{1,\dots,r\}$ *for some sequences* $\{x^k\}_{k\in\mathbb{N}}\to\bar{x}$ *and* $\{\Lambda^k\}_{k\in\mathbb{N}}\subset\mathcal{K}^\circ$ *such that* (44) *holds.*

Proof.

a. AKKT: If there are sequences $\{y_i^k\}_{k\in\mathbb{N}}\to 0$ such that $z_i^k := G_i(x^k)+y_i^k\in\mathcal{K}_i$ for every i and

$$\sum_{i=1}^r\langle\lambda_i^k, G_i(x^k)+y_i^k\rangle = 0,$$

then we have, for every i, that $\langle\lambda_i^k, G_i(x^k)+y_i^k\rangle=0$. So if $G_i(\bar{x})\in\operatorname{int}\mathcal{K}_i$, we get $\lambda_i^k=0$ for k sufficiently large because of Lemma 2. But, also, if $G_i(\bar{x})\in\operatorname{bd}_+\mathcal{K}_i$, then $\|\widehat{G}_i(\bar{x})\|=(G_i(\bar{x}))_0>0$, so for large enough k, we must have $(z_i^k)_0>0$ as well. Now, note that

$$0=\langle\lambda_i^k,z_i^k\rangle=\langle\widehat{\lambda}_i^k,\widehat{z}_i^k\rangle+(\lambda_i^k)_0(z_i^k)_0\le\|\widehat{\lambda}_i^k\|\|\widehat{z}_i^k\|+(\lambda_i^k)_0(z_i^k)_0\le(\|\widehat{\lambda}_i^k\|+(\lambda_i^k)_0)(z_i^k)_0,$$

but because $(z_i^k)_0>0$, we get $\|\widehat{\lambda}_i^k\|\ge-(\lambda_i^k)_0$. On the other hand, because $-\lambda_i^k\in\mathcal{K}_i$, we know that $\|\widehat{\lambda}_i^k\|\le-(\lambda_i^k)_0$. Hence, $\|\widehat{\lambda}_i^k\|=-(\lambda_i^k)_0$, which means $-\lambda_i^k\in\operatorname{bd}\mathcal{K}_i$. If $\lambda_i^k\not\to 0$, then $(\lambda_i^k)_0\not\to 0$, but we still have $\langle\lambda_i^k,z_i^k\rangle=0$, whence

$$1=\lim_{k\to\infty}\left\langle\frac{\widehat{\lambda}_i^k}{-(\lambda_i^k)_0},\frac{\widehat{z}_i^k}{(z_i^k)_0}\right\rangle=\lim_{k\to\infty}\left\langle\frac{\widehat{\lambda}_i^k}{\|\widehat{\lambda}_i^k\|},\frac{\widehat{G}_i(x^k)}{(G_i(x^k))_0}\right\rangle=\lim_{k\to\infty}\left\langle\frac{\widehat{\lambda}_i^k}{\|\widehat{\lambda}_i^k\|},\frac{\widehat{G}_i(\bar{x})}{\|\widehat{G}_i(\bar{x})\|}\right\rangle,$$

which means $\widehat{\lambda}_i^k/\|\widehat{\lambda}_i^k\|\to\widehat{G}_i(\bar{x})/\|\widehat{G}_i(\bar{x})\|$. In order to check this, keep in mind that both vectors are unitary, so the cosine of the angle between them must tend to one.

Conversely, without loss of generality, we can suppose that every λ_i^k such that $i\in\mathcal{I}_{\text{bd}_+}$ and $\lambda_i^k\to 0$ is equal to zero for k sufficiently large and set

$$y_i^k:=\begin{cases}-G_i(x^k), & \text{if } G_i(\bar{x})=0,\\ -\Pi_{\mathcal{K}^\circ}(G_i(x^k)), & \text{if } i\in\mathcal{I}_{\text{int}} \text{ or } i\in\mathcal{I}_{\text{bd}_+} \text{ with } \lambda_i^k\to 0,\\ \dfrac{\|\widehat{G}_i(x^k)\|}{(\lambda_i^k)_0}\left((\lambda_i^k)_0,-\widehat{\lambda}_i^k\right)-G_i(x^k), & \text{if } i\in\mathcal{I}_{\text{bd}_+} \text{ with } \lambda_i^k\not\to 0.\end{cases}$$

Then, we have $G_i(x^k)+y_i^k\in\mathcal{K}_i$ for every $i\in\{1,\dots,r\}$ because $((\lambda_i^k)_0,-\widehat{\lambda}_i^k)\in\mathcal{K}^\circ$ and $(\lambda_i^k)_0\le 0$. If $G_i(\bar{x})\in\{0\}\cup\operatorname{int}\mathcal{K}_i$, we clearly have $\langle G_i(x^k)+y_i^k,\lambda_i^k\rangle=0$ for k sufficiently large. If $i\in\mathcal{I}_{\text{bd}_+}$, because $-\lambda_i^k\in\operatorname{bd}\mathcal{K}_i$, we get

$$\langle G_i(x^k)+y_i^k,\lambda_i^k\rangle=\frac{\|\widehat{G}_i(x^k)\|}{(\lambda_i^k)_0}\left((\lambda_i^k)_0^2-\|\widehat{\lambda}_i^k\|^2\right)=0.$$

Also, note that

$$(y_i^k)_0=\|\widehat{G}_i(x^k)\|-(G_i(x^k))_0\ \text{and}\ \widehat{y_i^k}=\frac{\|\widehat{G}_i(x^k)\|}{\|\widehat{\lambda}_i^k\|}\widehat{\lambda}_i^k-\widehat{G}_i(x^k),$$

so in case $\lambda_i^k\not\to 0$, we get $y_i^k\to 0$ from $G_i(\bar{x})\in\operatorname{bd}\mathcal{K}_i\backslash\{0\}$ and (45).

b. AGP: In this case, because

$$\langle\Lambda^k,\Pi_{\mathcal{K}}(G(x^k))\rangle=\sum_{i=1}^r\langle\lambda_i^k,\Pi_{\mathcal{K}_i}(G_i(x^k))\rangle\to 0$$

and $\mathcal{K}^\circ=-\mathcal{K}$, we obtain $\langle\lambda_i^k,\Pi_{\mathcal{K}_i}(G_i(x^k))\rangle\to 0$ for every $i\in\{1,\dots,r\}$. Now, using the spectral decomposition (41) and the right-hand side of (43), we obtain

$$2\langle\lambda_i^k,\Pi_{\mathcal{K}_i}(G_i(x^k))\rangle\le\mu_1(\lambda_i^k)\mu_1(\Pi_{\mathcal{K}_i}(G_i(x^k)))+\mu_2(\lambda_i^k)\mu_2(\Pi_{\mathcal{K}_i}(G_i(x^k)))\le 0.$$

Hence, taking $k \to \infty$ in the preceding expression, we see that

- $\mu_1(\lambda_i^k)\mu_1(\Pi_{\mathcal{K}_i}(G_i(x^k))) = \mu_1(\lambda_i^k)\max\{\mu_1(G_i(x^k)),0\} \to 0$.
- $\mu_2(\lambda_i^k)\mu_2(\Pi_{\mathcal{K}_i}(G_i(x^k))) = \mu_2(\lambda_i^k)\max\{\mu_2(G_i(x^k)),0\} \to 0$.

Furthermore, from the preceding and because $\mu_1(\lambda^k) \le \mu_2(\lambda^k) \le 0$, we see that $\mu_2(\lambda_i^k)\max\{\mu_1(G_i(x^k)),0\} \to 0$. Thus, $\langle \lambda_i^k, \Pi_{\mathcal{K}_i}(G_i(x^k))\rangle \to 0$ implies $\mu_\ell(\lambda_i^k)\max\{\mu_\ell(G_i(x^k)),0\} \to 0$ for $\ell \in \{1,2\}$ and $\forall i \in \{1,\dots,r\}$.

CAKKT: Again, $\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k))\rangle \to 0$ and $\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k))\rangle \to 0$ are equivalent to $\langle \lambda_i^k, \Pi_{\mathcal{K}_i}(G_i(x^k))\rangle \to 0$ and $\langle \lambda_i^k, \Pi_{\mathcal{K}_i^\circ}(G_i(x^k))\rangle \to 0$ for every $i \in \{1,\dots,r\}$.

From $\langle \lambda_i^k, \Pi_{\mathcal{K}_i^\circ}(G_i(x^k))\rangle \to 0$, using (41) and the left-hand side of (43), we have

$$0 \le \mu_1(\lambda_i^k)\mu_2(\Pi_{\mathcal{K}_i^\circ}(G_i(x^k))) + \mu_2(\lambda_i^k)\mu_1(\Pi_{\mathcal{K}_i^\circ}(G_i(x^k))) \le 2\langle \lambda_i^k, \Pi_{\mathcal{K}_i^\circ}(G_i(x^k))\rangle.$$

Then, from (41) and taking $k \to \infty$ in the preceding expression, we see that

- $\mu_1(\lambda_i^k)\mu_2(\Pi_{\mathcal{K}_i^\circ}(G_i(x^k))) = \mu_1(\lambda_i^k)\min\{\mu_2(G_i(x^k)),0\} \to 0$.
- $\mu_2(\lambda_i^k)\mu_1(\Pi_{\mathcal{K}_i^\circ}(G_i(x^k))) = \mu_2(\lambda_i^k)\min\{\mu_1(G_i(x^k)),0\} \to 0$.

Furthermore, from the preceding, we see that $\mu_2(\lambda_i^k)\min\{\mu_2(G_i(x^k)),0\} \to 0$. Thus, $\langle \lambda_i^k, \Pi_{\mathcal{K}_i^\circ}(G_i(x^k))\rangle \to 0$ implies $\mu_\ell(\lambda_i^k)\min\{\mu_s(G_i(x^k)),0\} \to 0$ for $\ell, s \in \{1,2\}$, $s \neq \ell$, and $\forall i \in \{1,\dots,r\}$. See Table 3. □

Note that AKKT is consistent with Andreani et al. [12, definition 3.3] in view of Andreani et al. [12, theorem 4.1], which gives the exact same characterization as item (a) of Proposition 5, and AGP was not yet defined for NSOCP. Also, our version of CAKKT comprises eigenvalue products, which is similar to what we expected to obtain in the NLSDP case.

Table 3 summarizes our results.

In NSOCP, the relation between Definition 3 and CAKKT as in Andreani et al. [12, definition 3.4] is not clear. We conjecture that they are independent, which may endorse the possibility of the existence of multiple independent extensions of CAKKT. Observe that the most important feature of our approach is its simplicity and its generality because it only uses inner products and projections. On the other hand, Andreani et al. [12, definition 3.4] relies on the Jordan algebra structure, which is limited to symmetric cones, but it is more elegant than our approach in certain aspects.

7. Strength of the Sequential Optimality Conditions

A sequential optimality condition carries the convergence properties of the algorithms supported by them, and this is what gives them a practical meaning. However, even though we compared sequential conditions among themselves, we have not yet shown any improvement regarding the usual convergence theory of any algorithm. In other words, to complete our results, we still need to clarify the relation between our sequential conditions and other optimality conditions of the form "KKT or not-CQ" for some CQ. This section is dedicated to filling this gap.

Recall that the classical Robinson's [70] CQ holds at some feasible point x when

$$0 \in \operatorname{int}(G(x) + DG(x)\mathbb{R}^n - \mathcal{K}), \tag{46}$$

where $DG(x)\mathbb{R}^n := \{DG(x)d : d \in \mathbb{R}^n\}$. It is widely known that Robinson's CQ generalizes the MFCQ from NLP (see Mangasarian and Fromovitz [56]). We proceed by reproving the classical convergence results to KKT points under Robinson's CQ via sequential conditions.

Table 3. Sequential optimality conditions when specialized to (NSOCP). We use $\{\mu_\ell(v), \ell \in \{1,2\}\}$ to denote spectral values of $v \in \mathcal{K}_i$ (see (42)). For the definitions of $\mathcal{I}_{bd+}$ and $\mathcal{I}_{int}$, see (40).

	Approximate complementarity condition
AKKT	$i \in \mathcal{I}_{int}$, $\lambda_i^k \to 0$, and for $i \in \mathcal{I}_{bd+}$, $-\lambda_i^k \in \operatorname{bd} \mathcal{K}_i$ and either $\lambda_i^k \to 0$ or $\widehat{\lambda}_i^k/\lVert\widehat{\lambda}_i^k\rVert \to \widehat{G_i(\bar{x})}/\lVert\widehat{G_i(\bar{x})}\rVert$
AGP	$\mu_\ell(\lambda_i^k)\max\{\mu_\ell(G_i(x^k)),0\} \to 0$ for $\forall \ell \in \{1,2\}$, $\forall i$
CAKKT	$\mu_\ell(\lambda_i^k)\max\{\mu_\ell(G_i(x^k)),0\} \to 0$ and $\mu_\ell(\lambda_i^k)\min\{\mu_s(G_i(x^k)),0\} \to 0, \forall \ell, s \in \{1,2\}$ $(\ell \neq s)$, $\forall i$

Proposition 6. *If Robinson's CQ holds at an AKKT point $\bar{x}$ associated with the sequences $\{x^k\}_{k\in\mathbb{N}}$ and $\{\Lambda^k\}_{k\in\mathbb{N}}$, then $\bar{x}$ satisfies the KKT conditions for (NCP).*

Proof. We begin by proving that $\{\Lambda^k\}_{k\in\mathbb{N}}$ is bounded. In order to do that, by contradiction, suppose not. Then, we can assume $\|\Lambda^k\| \to \infty$, but $\widehat{\Lambda^k} := \Lambda^k/\|\Lambda^k\| \to \overline{\Lambda} \in \mathcal{K}^\circ$. Then, from (9), we get $\nabla f(x^k)/\|\Lambda^k\| + DG(x^k)^*\widehat{\Lambda}^k \to 0$ and, consequently, $DG(\bar{x})^*\overline{\Lambda} = 0$. Moreover,

$$0 = \langle \Lambda^k, G(x^k) + y^k \rangle = \langle \widehat{\Lambda}^k, G(x^k) + y^k \rangle \to \langle \overline{\Lambda}, G(\bar{x}) \rangle = 0,$$

and by Robinson's CQ, there exists some small $\alpha > 0$ such that $-\alpha\overline{\Lambda} \in (G(\bar{x}) + DG(\bar{x})\mathbb{R}^n - \mathcal{K})$. Let $d \in \mathbb{R}^n$ and $z \in \mathcal{K}$ be such that $-\alpha\overline{\Lambda} = G(\bar{x}) + DG(\bar{x})d - z$, so we have

$$-\alpha\langle \overline{\Lambda}, \overline{\Lambda} \rangle = \langle \overline{\Lambda}, G(\bar{x}) \rangle + \langle \overline{\Lambda}, DG(\bar{x})d \rangle - \langle \overline{\Lambda}, z \rangle = \langle DG(\bar{x})^*\overline{\Lambda}, d \rangle - \langle \overline{\Lambda}, z \rangle \geq 0,$$

which implies $\overline{\Lambda} = 0$. Because this contradicts the definition of $\overline{\Lambda}$, we conclude that $\{\Lambda^k\}_{k\in\mathbb{N}}$ must be bounded. Hence, without loss of generality, we can assume it converges to $\overline{\Lambda} \in \mathcal{K}^\circ$. Trivially, $\nabla f(\bar{x}) + DG(\bar{x})^*\overline{\Lambda} = 0$ and

$$|\langle \overline{\Lambda}, G(\bar{x}) \rangle| = \lim_{k\to\infty} |\langle \Lambda^k, G(x^k) \rangle| = |\lim_{k\to\infty} \langle \Lambda^k, y^k \rangle| = 0.$$

Also, because $\mathcal{K}$ is closed and $\mathcal{K} \ni G(x^k) + y^k \to G(\bar{x})$, we have $G(\bar{x}) \in \mathcal{K}$. Thus, $(\bar{x}, \overline{\Lambda})$ is a KKT pair of (NCP). □

Analogously, it can be proved that TAKKT also satisfies the KKT conditions under Robinson's CQ. Moreover, because AKKT is implied by CAKKT and AGP, the same holds for both. That means every algorithm that is supported by one of our sequential conditions converges to KKT points under Robinson's CQ, but sequential conditions tell us more than that. Following Andreani et al. [11], for each sequential optimality condition (OC), it is possible to define conditions, so-called *strict constraint qualifications* (SCQ) such that "$OC + SCQ \Rightarrow KKT$" and, among them, characterize the weakest one.

We use the same nomenclature style of Andreani et al. [11]. For instance, the weakest SCQ associated with the AKKT condition is called *AKKT-regularity*, and similar names are given for the other sequential optimality conditions presented in Section 3.

Definition 5. Consider the following sets:

1. $\mathbb{K}_A(x,r) := \{DG(x)^*\Lambda : |y| \leq r,\ \Lambda \in \mathcal{K}^\circ, G(x) + y \in \mathcal{K} \cap \{\Lambda\}^\perp\}$;
2. $\mathbb{K}_T(x,r) := \{DG(x)^*\Lambda : |\langle \Lambda, G(x) \rangle| \leq r,\ \Lambda \in \mathcal{K}^\circ\}$;
3. $\mathbb{K}_{AGP}(x,r) := \{DG(x)^*\Lambda : |\langle \Lambda, \Pi_{\mathcal{K}}(G(x)) \rangle| \leq r,\ \Lambda \in \mathcal{K}^\circ\}$;
4. $\mathbb{K}_C(x,r) := \{DG(x)^*\Lambda : \max\{|\langle \Lambda, \Pi_{\mathcal{K}}(G(x)) \rangle|, |\langle \Lambda, \Pi_{\mathcal{K}^\circ}(G(x)) \rangle|\} \leq r,\ \Lambda \in \mathcal{K}^\circ\}$.

We say that the AKKT-regularity condition holds at $\bar{x}$ if the set-valued mapping $(x,r) \longmapsto \mathbb{K}_A(x,r)$ is outer semicontinuous at $(\bar{x}, 0)$. The constraint qualification conditions TAKKT-, AGP-, and CAKKT-regularity have analogous definitions using the sets $\mathbb{K}_T(x,r), \mathbb{K}_{AGP}(x,r)$, and $\mathbb{K}_C(x,r)$, respectively.

Remark 4. For a feasible point $\bar{x}$ of (NCP), we see that, at $(\bar{x}, 0)$, all the sets from Definition 5 coincide with $\{DG(x)^*\Lambda : \langle \Lambda, G(x) \rangle = 0, \Lambda \in \mathcal{K}^\circ\}$. Thus, given an objective function f for (NCP), the KKT conditions hold at $\bar{x}$ if and only if $-\nabla f(x) \in \mathbb{K}_A(\bar{x}, 0)$. Similar statements hold for the other sets.

The next theorem states that each SCQ is, indeed, the weakest SCQ associated with each optimality condition.

Theorem 5. *A feasible point $\bar{x}$ for (NCP) satisfies CAKKT-regularity if and only if for every continuously differentiable objective function, the CAKKT condition at $\bar{x}$ implies the KKT conditions. Similar conclusions are valid for AKKT-, AGP-, and TAKKT-regularity.*

Proof. We use the same techniques as Andreani et al. [10, 11]. We just prove the statement for CAKKT-regularity because the other ones are analogous.

Suppose that CAKKT-regularity holds at $\bar{x}$ and take any objective function f having $\bar{x}$ as a CAKKT point. By definition, there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and $\{\Lambda^k\}_{k\in\mathbb{N}} \subset \mathcal{K}^\circ$ such that

$$w^k := \nabla f(x^k) + DG(x^k)^*\Lambda^k \to 0,\ \langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle \to 0, \text{ and } \langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle \to 0.$$

Set $r_k := \max\{|\langle \Lambda^k, \Pi_{\mathcal{K}}(G(x^k)) \rangle|, |\langle \Lambda^k, \Pi_{\mathcal{K}^\circ}(G(x^k)) \rangle|\}$, $\forall k \in \mathbb{N}$. Thus, $-\nabla f(x^k) + w^k \in \mathbb{K}_C(x^k, r_k)$. Taking limits in the last expression, we get $-\nabla f(\bar{x}) \in \mathbb{K}_C(\bar{x}, 0)$ from the outer semicontinuity of $\mathbb{K}_C$, and hence, the KKT conditions hold at $\bar{x}$. Conversely, suppose that, for every continuously differentiable objective function, the CAKKT condition at $\bar{x}$ implies the KKT conditions. We show that CAKKT-regularity holds at $\bar{x}$. Take $\omega \in \limsup_{(x,r)\to(\bar{x},0)} \mathbb{K}_C(x,r)$. Then, there exists an infinite subset $I \subseteq \mathbb{N}$ and sequences $\{x^k\}_{k\in I} \to \bar{x}$, $\{r_k\}_{k\in I} \to 0$, and $\{w^k\}_{k\in I} \to w$ such that $w^k \in \mathbb{K}_C(x^k, r_k)$ for every $k \in I$, which means each w^k is associated with some $\Lambda^k \in \mathcal{K}^\circ$. Now, define the function $f(x) := -\langle \omega, x\rangle$, and let $i(k)$ denote the element of I that is closest to a given $k \in \mathbb{N}$; in particular, when $k \in I$, then $i(k) = k$. Defining $\tilde{x}^k := x^{i(k)}$ and $\tilde{\Lambda}^k := \Lambda^{i(k)}$ for every $k \in \mathbb{N}$, we see that $\bar{x}$ is a CAKKT point for f associated with the sequences $\{\tilde{x}^k\}_{k\in\mathbb{N}}$ and $\{\tilde{\Lambda}^k\}_{k\in\mathbb{N}}$. Then, by assumption, the KKT conditions hold, which is equivalent to $-\nabla f(\bar{x}) = w \in \mathbb{K}_C(\bar{x}, 0)$, and thus, CAKKT-regularity holds at $\bar{x}$. □

From Theorem 6 and Definition 5, we observe that Robinson's CQ implies AKKT-regularity, which strictly implies AGP-regularity, and the latter strictly implies CAKKT-regularity because of Theorem 5 and the relations among the sequential conditions from Sections 3 and 5. Thus, our previous considerations show that the algorithms supported by the CAKKT condition are guaranteed to converge to KKT points under assumptions that are weaker than Robinson's CQ, for example.

Now, to conclude our analyses, we make explicit the relation between our SCQs and the very weak *Abadie's* CQ, which holds at a feasible point $\bar{x}$ if and only if $T_\Omega(\bar{x}) = L_\Omega(\bar{x})$ and the set $\mathbb{K}_C(\bar{x}, 0) = \{DG(\bar{x})^*\Lambda : \langle \Lambda, G(\bar{x})\rangle = 0, \Lambda \in \mathcal{K}^\circ\}$ is closed, where

$$T_\Omega(\bar{x}) := \{d \in \mathbb{R}^n : \exists t_k \downarrow 0, d^k \to d \text{ with } \bar{x} + t_k d^k \in \Omega\}$$

is the *tangent cone* to Ω at $\bar{x}$ and

$$L_\Omega(\bar{x}) := \{d \in \mathbb{R}^n : DG(\bar{x})d \in T_{\mathcal{K}}(G(\bar{x}))\}$$

is the so-called *linearized tangent cone* to Ω at $\bar{x}$.

The only affirmation that requires a proof is "CAKKT-regularity implies Abadie's CQ," but note that even in the finite-dimensional setting of (NCP) the set $\mathbb{K}_C(\bar{x}, 0)$ may not be closed (Pataki [63]). Therefore, the first thing we prove is that this condition is guaranteed under CAKKT-regularity.

Lemma 3. *Let $\bar{x}$ be a feasible point of (NCP) that satisfies CAKKT-regularity. Then, $\mathbb{K}_C(\bar{x}, 0)$ is closed, which, in turn, implies that it coincides with $L_\Omega(\bar{x})^\circ$.*

Proof. Recall that $\limsup_{(x,r)\to(\bar{x},0)} \mathbb{K}_C(x, r)$ is always a closed set, so if it coincides with $\mathbb{K}_C(\bar{x}, 0)$, then the latter is also closed. Moreover, it follows directly from Rockafellar and Wets [73, corollary 11.25(d)] that, if $\mathbb{K}_C(\bar{x}, 0)$ is closed, then it coincides with $L_\Omega(\bar{x})^\circ$. □

In light of Lemma 3, the rest of the proof follows the same recipe as in NLP. To present it, we first recall the *regular normal cone* to Ω at $\bar{z} \in \Omega$, which is defined as

$$\widehat{N}_\Omega(\bar{z}) := \left\{ w \in \mathbb{R}^n : \limsup_{z\to\bar{z}, z\in\Omega} \|z - \bar{z}\|^{-1} \langle w, z - \bar{z}\rangle \leq 0 \right\},$$

and the *limiting normal cone* to Ω at $\bar{x} \in \Omega$, which is $N_\Omega(\bar{z}) := \limsup_{z\to\bar{z}, z\in\Omega} \widehat{N}_\Omega(z)$. Now, we present a technical lemma.

Lemma 4. *We always have that $N_\Omega(\bar{x}) \subset \limsup_{(x,r)\to(\bar{x},0)} \mathbb{K}_C(x, r)$.*

Proof. Analogous to the proof of Andreani et al. [11, lemma 4.3]. □

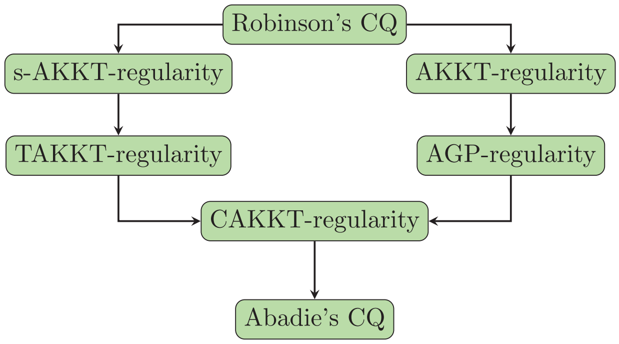
And, finally, the result:

Theorem 6. *CAKKT-regularity implies Abadie's CQ.*

Proof. Let $\bar{x}$ be a feasible point such that CAKKT-regularity holds at $\bar{x}$. Using the definition of outer semicontinuity of $\mathbb{K}_C(x, r)$ and Lemma 4, we get that $N_\Omega(\bar{x}) \subset \mathbb{K}_C(\bar{x}, 0)$, and by Lemma 3, we have that $\mathbb{K}_C(\bar{x}, 0)$ is closed and $\mathbb{K}_C(\bar{x}, 0) = L_\Omega(\bar{x})^\circ$. Therefore, $N_\Omega(\bar{x}) \subset L_\Omega(\bar{x})^\circ$. Now, because $T_\Omega(\bar{x}) \subset L_\Omega(\bar{x})$ always holds for every set, to show that Abadie's CQ holds at $\bar{x}$, it suffices to prove the inclusion $L_\Omega(\bar{x}) \subset T_\Omega(\bar{x})$. Now, from the inclusion $N_\Omega(\bar{x}) \subset L_\Omega(\bar{x})^\circ$, we get that $L_\Omega(\bar{x}) \subset (L_\Omega(\bar{x})^\circ)^\circ \subset N_\Omega(\bar{x})^\circ \subset T_\Omega(\bar{x})$, where the first inclusion follows from the polarity theorem because $L_\Omega(\bar{x})$ is a closed convex cone because $\mathcal{K}$ is also a closed convex cone, and the last inclusion comes from Rockafellar and Wets [73, theorems 6.28(b) and 6.26]. □

Remark 5. Börgens et al. [26, theorem 4.6] states that *s-AKKT-regularity* Börgens et al. [26, definition 4.4] is the weakest constraint qualification that guarantees equivalence between KKT and s-AKKT, similarly to Theorem 5.

Figure 2. (Color online) Relationship between the new (strict) constraint qualifications and existing ones.



Therefore, s-AKKT-regularity is independent of AKKT-regularity, independent of AGP-regularity, and strictly stronger than TAKKT-regularity. When $\mathcal{K}$ is self-dual, however, it implies AGP regularity strictly. Moreover, we obtain from Börgens et al. [26, theorem 5.2 and example 5.4], that s-AKKT-regularity is strictly implied by Robinson's CQ.

For summarizing our results, we illustrate the position of the new SCQs among the existing CQs: s-AKKT-regularity, Abadie's CQ, and Robinson's CQ, in the diagram of Figure 2.

8. Final Remarks

Powerful modeling languages and other recent technological advances extended the possibilities for solving complex real-life problems. Such complexity is often translated in terms of (NCP), which is a large family of optimization problems, that generalizes NLP, NLSDP, and NSOCP, for example. In this paper, we extended to the NCP context some of the so-called sequential optimality conditions, which have been useful in particular cases of NCP for improving the global convergence analysis of several practical algorithms in a unified manner. Also, we presented a variant of the augmented Lagrangian method for NCP, whose global convergence theory was built via sequential optimality conditions. We proved that every feasible limit point of this method satisfies AGP, and under an additional smoothness assumption, it also satisfies CAKKT, which is a strictly stronger condition. The meaning of such results lies in the fact that every CAKKT (respectively, AGP) point also satisfies the KKT conditions in the presence of a constraint qualification called CAKKT-regularity (AGP-regularity), which is strictly weaker than Robinson's condition. That means, for instance, that Algorithm 1 is at least as strong as the classical variants of the augmented Lagrangian method despite being much more general. To the best of our knowledge, the convergence of the augmented Lagrangian to CAKKT points was only known in NLP and, more recently, in NSOCP, but its convergence to AGP points was not yet known even in NLP.

Intuitively, one may expect general environments to be more complicated or to be less likely to achieve strong results in comparison with more structured ones. However, in this work, we see the opposite because we are able to recover and improve most of the existing results from NLP, NLSDP, and NSOCP while employing simpler techniques in our analyses. In fact, we limit ourselves to using only somewhat simple structures, such as inner products and projections, to make our results as applicable as possible. Our efforts lead us to believe that NCP encompasses most of the fundamental aspects of the classical optimization theory in a natural way, which may encourage further research in this field. For instance, the relation between CAKKT and the concept with the same name, from Andreani et al. [12], is still unclear. Another subject of further investigation is the role of sequential conditions in perturbation theory and error estimation, which may clarify their value as a theoretical local optimality analysis tool as an alternative to the punctual KKT conditions. Second order sequential conditions have recently appeared in Andreani et al. [6], Birgin et al. [22], Haeser [40], and Haeser and Ramos [41] for NLP, and we intend to extend them to more general contexts as well. Moreover, as mentioned in the introduction, there are several algorithms for NLP that have had their convergence theories (re)built via sequential optimality conditions of first and second order, such as Andreani et al. [8], Birgin et al. [21], Chen and Goldfarb [30], Gill et al. [37, 38], Haeser [40], Haeser et al. [42], Qi and Wei [67], and we believe that this work can be useful for extending such algorithms to NCP as well.

Acknowledgments
The authors are thankful to the editors and two anonymous referees for their remarkable suggestions. In particular, one referee pointed out Börgens et al. [26] and suggested Remarks 3 and 5.

References
[1] Andreani R, Haeser G, Martínez JM (2011) On sequential optimality conditions for smooth constrained optimization. *Optim.* 60(5):627–641.
[2] Andreani R, Haeser G, Viana DS (2020) Optimality conditions and global convergence for nonlinear semidefinite programming. *Math. Programming* 180(1–2):203–235.
[3] Andreani R, Martínez JM, Svaiter BF (2010) A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM J. Optim.* 20(6):3533–3554.
[4] Andreani R, Birgin EG, Martínez JM, Schuverdt ML (2008) Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. *Math. Programming* 111(1–2):5–32.
[5] Andreani R, Haeser G, Mito LM, Ramírez H (2020) Weak notions of nondegeneracy in nonlinear semidefinite programming. Technical report. Accessed December 29, 2020, http://www.optimization-online.org/DB_HTML/2020/12/8182.html.
[6] Andreani R, Haeser G, Ramos A, Silva PJS (2017) A second-order sequential optimality condition associated to the convergence of optimization algorithms. *IMA J. Numerical Anal.* 37(4):1902–1929.
[7] Andreani R, Haeser G, Schuverdt ML, Silva PJS (2012a) A relaxed constant positive linear dependence constraint qualification and applications. *Math. Programming* 135(1–2):255–273.
[8] Andreani R, Haeser G, Schuverdt ML, Silva PJS (2012b) Two new weak constraint qualifications and applications. *SIAM J. Optim.* 22(3): 1109–1135.
[9] Andreani R, Haeser G, Secchin LD, Silva PJS (2019b) New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences. *SIAM J. Optim.* 29(4):3201–3230.
[10] Andreani R, Martínez JM, Ramos A, Silva PJS (2016) A cone-continuity constraint qualification and algorithmic consequences. *SIAM J. Optim.* 26(1):96–110.
[11] Andreani R, Martínez JM, Ramos A, Silva PJS (2018) Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Math. Oper. Res.* 43(3):693–717.
[12] Andreani R, Fukuda EH, Haeser G, Santos DO, Secchin LD (2019) Optimality conditions for nonlinear second-order cone programming and symmetric cone programming. Technical report. Accessed October 17, 2019, http://www.optimization-online.org/DB_HTML/2019/10/7436.html.
[13] Andreani R, Haeser G, Mito LM, Ramos A, Secchin LD (2020) On the best achieavable quality of limit points of augmented Lagrangian schemes. *Numer. Algorithms.* http://doi.org.br/10.1007/s11075-021-01212-8.
[14] Anjos MF, Lassere JB (2012) *Handbook on Semidefinite, Conic and Polynomial Optimization*, International Series in Operations Research and Management Science (Springer, New York).
[15] Attouch H, Bolte J, Svaiter BF (2013) Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting and regularized Gauss-Seidel methods. *Math. Programming* 137(1–2):91–129.
[16] Attouch H, Bolte J, Redont P, Soubeyran A (2010) Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Math. Oper. Res.* 35(2):438–457.
[17] Baes M (2007) Convexity and differentiability properties of spectral functions and spectral mappings on Euclidean Jordan algebras. *Linear Algebra Appl.* 422(2):664–700.
[18] Bertsekas DP (1982) *Constrained Optimization and Lagrange Multiplier Methods* (Academic Press, New York).
[19] Bertsekas DP (1999) *Nonlinear Programming* (Athenas Scientific, Belmont, MA).
[20] Birgin E, Martínez JM (2014) *Practical Augmented Lagrangian Methods for Constrained Optimization* (SIAM Publications, Philadelphia).
[21] Birgin EG, Bueno LF, Martínez JM (2015) Assessing the reliability of general-purpose inexact restoration methods. *J. Comput. Appl. Math.* 282(C):1–16.
[22] Birgin EG, Haeser G, Ramos A (2018a) Augmented Lagrangians with constrained subproblems and convergence to second-order stationary points. *Comput. Optim. Appl.* 69(1):51–75.
[23] Birgin EG, Krejić N, Martínez JM (2018b) On the employment of inexact restoration for the minimization of functions whose evaluation is subject to errors. *Math. Comput.* 87:1307–1326.
[24] Bolte J, Daniilidis A, Ley O, Mazet L (2010) Characterization of Łojasiewicz inequalities and applications: Subgradient flows, talweg, convexity. *Trans. Amer. Math. Soc.* 362(6):3319–3363.
[25] Bonnans JF, Shapiro A (2000) *Pertubation Analysis of Optimization Problems* (Springer-Verlag, Berlin).
[26] Börgens E, Kanzow C, Mehlitz P, Wachsmuth G (2020b) New constraint qualifications for optimization problems in Banach spaces based on asymptotic KKT conditions. *SIAM J. Optim.* 30(4):2956–2982.
[27] Bueno LF, Haeser G, Martínez JM (2015) A flexible inexact restoration method for constrained optimization. *J. Optim. Theory Appl.* 165(1): 188–208.
[28] Bueno LF, Haeser G, Rojas FN (2019) Optimality conditions and constraint qualifications for generalized Nash equilibrium problems and their practical implications. *SIAM J. Optim.* 29(1):31–54.
[29] Bueno LF, Haeser G, Lara F, Rojas FN (2020) An augmented Lagrangian method for quasi-equilibrium problems. *Comput. Optim. Appl.* 76:737–766.
[30] Chen L, Goldfarb D (2006) Interior-point 2-penalty methods for nonlinear programming with strong global convergence properties. *Math. Programming* 108:1–36.
[31] Chill R, Mildner S (2018) The Kurdyka-Łojasiewicz-Simon inequality and stabilisation in nonsmooth infinite-dimensional gradient systems. *Proc. Amer. Math. Soc.* 146:4307–4314.

[32] Fares B, Apkarian P, Noll D (2001) An augmented Lagrangian method for a class of LMI-constrained problems in robust control theory. *Internat. J. Control* 74(4):348–360.
[33] Fischer A, Friedlander A (2010) A new line search inexact restoration approach for nonlinear programming. *Comput. Optim. Appl.* 46(2):333–346.
[34] Fitzpatrick S, Phelps RR (1982) Differentiability of the metric projection in Hilbert space. *Trans. Amer. Math. Soc.* 270(2):483–501.
[35] Fukuda EH, Haeser G, Mito LM (2020) On the weak second-order optimality condition for nonlinear semidefinite and second-order cone programming. Technical report. Accessed August 4, 2020, http://www.optimization-online.org/DB_HTML/2020/08/7951.html.
[36] Fukushima M, Luo ZQ, Tseng P (2006) Smoothing functions for second-order-cone complementarity problems. *SIAM J. Optim.* 12(2):436–460.
[37] Gill PE, Kungurtsev V, Robinson DP (2017) A stabilized SQP method: Global convergence. *IMA J. Numerical Anal.* 37(1):407–443.
[38] Gill PE, Kungurtsev V, Robinson DP (2020) A shifted primal-dual penalty-barrier method for nonlinear optimization. *SIAM J. Optim.* 30(2):1067–1093.
[39] Giorgi G, Jiménez B, Novo V (2016) Approximate Karush-Kuhn-Tucker condition in multiobjective optimization. *J. Optim. Theory Appl.* 171(1):70–89.
[40] Haeser G (2018) A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms. *Comput. Optim. Appl.* 70(2):615–639.
[41] Haeser G, Ramos A (2020) New constraint qualifications with second-order properties in nonlinear optimization. *J. Optim. Theory Appl.* 182(2):494–506.
[42] Haeser G, Hinder O, Ye Y (2021) On the behavior of Lagrange multipliers in convex and nonconvex infeasible interior point methods. *Math. Programming* 186:257–288.
[43] Haeser G, Liu H, Ye Y (2019) Optimality condition and complexity analysis for linearly-constrained optimization without differentiability on the boundary. *Math. Programming* 178(1):263–299.
[44] Hager WW, Gowda MS (1999) Stability in the presence of degeneracy and error estimation. *Math. Programming* 85(1):181–192.
[45] Hager WW, Mico-Umutesi D (2014) Error estimation in nonlinear optimization. *J. Global Optim.* 59(2–3):327–341.
[46] Hestenes MR (1969) Multiplier and gradient methods. *J. Optim. Theory Appl.* 4:303–320.
[47] Hiriart-Urruty JB, Lemaréchal C (2001) *Fundamentals of Convex Analysis* (Springer-Verlag, Berlin).
[48] Kanzow C, Steck D (2017) An example comparing the standard and safeguarded augmented Lagrangian methods. *Oper. Res. Lett.* 45(6):598–603.
[49] Kanzow C, Steck D, Wachsmuth D (2018) An augmented Lagrangian method for optimization problems in Banach spaces. *SIAM J. Control Optim.* 56(1):272–291.
[50] Kocvara M, Stingl M (2004) Solving nonconvex SDP problems of structural optimization with stability control. *Optim. Methods Software* 19(5):595–609.
[51] Krislock N, Wolkowicz H (2012) Euclidean distance matrices and applications. Anjos MF, Lasserre JB, eds. *Handbook on Semidefinite, Conic and Polynomial Optimization*. International Series in Operations Research and Management Science, vol. 166 (Springer, Boston), 879–914. 10.1007/978-1-4614-0769-0_30.
[52] Kurdyka K (1998) On gradients of functions definable in o-minimal structures. *Ann. Inst. Fourier (Grenoble)* 48(3):769–783.
[53] Li G, Pong TK (2018) Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its appplications to linear convergence of first-order methods. *Foundations Comput. Math.* 18(5):1199–1232.
[54] Lobo MS, Vandenberghe L, Boyd S, Lebret H (1998) Applications of second-order cone programming. *Linear Algebra Appl.* 284(1–3):193–228.
[55] Łojasiewicz S (1984) Sur les trajectories du gradient d'une fonction analytique. *in Seminari di Geometria 1982-1983, Universita di Bologna, Istituto di Geometria, Dipartamento di Matematica*, 115–117.
[56] Mangasarian OL, Fromovitz S (1967) The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *J. Math. Anal. Appl.* 17(1):37–47.
[57] Marcus M (1956) An eigenvalue inequality for the product of normal matrices. *Amer. Math. Monthly* 63(3):173–174.
[58] Martínez JM, Pilotta EA (2000) Inexact restoration algorithms for constrained optimization. *J. Optim. Theory Appl.* 104:135–163.
[59] Martínez JM, Pilotta EA (2005) Inexact restoration methods for nonlinear programming: advances and perspectives. Qi L, Teo K, Yang X, eds. *Optimization and Control with Applications* (Springer, Boston), 271–292.
[60] Martínez JM, Svaiter BF (2003) A practical optimality condition without constraint qualifications for nonlinear programming. *J. Optim. Theory Appl.* 118:117–133.
[61] Nocedal J, Wright S (2006) *Numerical Optimization*, Springer Series in Operations Research (Springer-Verlag, New York).
[62] O'Neill M, Wright SJ (2021) A log-barrier Newton-CG method for bound constrained optimization with complexity guarantees. *IMA J. Numerical Anal.* 41(1):84–121.
[63] Pataki G (2007) On the closedness of the linear image of a closed convex cone. *Math. Oper. Res.* 32(2):395–412.
[64] Peng J, Roos C, Terlaky T (2002) *Self-Regularity: A New Paradigm for Primal-Dual Interior Point Algorithms*, Princeton Series in Applied Mathematics (Princeton University Press, Chichester, UK).
[65] Powell MJD (1969) A method for nonlinear constraints in minimization problems. Fletcher R, ed. *Optimization* (Academic Press, New York), 283–298.
[66] Qi HD, Yuan X (2013) Computing the nearest Euclidean distance matrix with low embedding dimensions. *Math. Programming* 147(1–2):351–389.
[67] Qi L, Wei Z (2000) On the constant positive linear dependence conditions and its application to SQP methods. *SIAM J. Optim.* 10(4):963–981.
[68] Ramos A (2019) Two new weak constraint qualifications for mathematical programs with equilibrium constraints and applications. *J. Optim. Theory Appl.* 183:566–591.
[69] Ramos A (2021) Mathematical programs with equilibrium constraints: A sequential optimality condition, new constraint qualifications and algorithmic consequences. *Optim. Methods Software* 36(1):45–81.

[70] Robinson SM (1976) Stability theory for systems of inequalities, part II: Differentiable nonlinear systems. *SIAM J. Numerical Anal.* 13(4): 497–513.
[71] Robinson SM (1982) Generalized equations and their solutions, part II: Applications to nonlinear programming. *Math. Programming Stud.* 19:200–221.
[72] Rockafellar RT (1974) Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM J. Control Optim.* 12(2): 268–285.
[73] Rockafellar RT, Wets R (2009) *Variational Analysis* (Springer-Verlag, Berlin).
[74] Schnabel RB (1982) Determining feasibility of a set of nonlinear inequality constraints. *Math. Programming Stud.* 16:137–148.
[75] Steck D (2018) Lagrange multiplier methods for constrained optimization and variational problems in Banach spaces. Unpublished PhD thesis, Institute of Mathematics, Universität Würzburg, Germany.
[76] Tuyen NV, Xiao YB, Son TQ (2019) On AKKT optimality conditions for cone-constrained vector optimization problems. Preprint, submitted February 20. Accessed August 10, 2020, https://arxiv.org/abs/1902.07416.
[77] Wolkowicz H, Saigal R, Vandenberghe L (2000) *Handbook of Semidefinite Programming: Theory, Algorithms, an Applications*, International Series in Operations Research and Management Science (Springer, New York).
[78] Yamashita H, Yabe H (2015) A survey of numerical methods for nonlinear semidefinite programming. *J. Oper. Res. Soc. Japan* 58(1):24–60.

Appendix B

External reference II

Title: On the best achievable quality of limit points of augmented Lagrangian schemes.

Status: Accepted for publication in Numerical Algorithms.

What is attached: Our copy of the paper and an erratum with a few misprint corrections that was published afterwards.

Numerical Algorithms
https://doi.org/10.1007/s11075-021-01212-8

ORIGINAL PAPER

On the best achievable quality of limit points of augmented Lagrangian schemes

Roberto Andreani[1] · Gabriel Haeser[2] · Leonardo M. Mito[2] · Alberto Ramos[3] · Leonardo D. Secchin[4]

Received: 9 March 2021 / Accepted: 29 September 2021 / Published online: 30 October 2021

Abstract

The optimization literature is vast in papers dealing with improvements on the global convergence of augmented Lagrangian schemes. Usually, the results are based on weak constraint qualifications, or, more recently, on *sequential* optimality conditions obtained via penalization techniques. In this paper, we propose a somewhat different approach, in the sense that the algorithm itself is used in order to formulate a new optimality condition satisfied by its feasible limit points. With this tool at hand, we present several new properties and insights on limit points of augmented Lagrangian schemes, in particular, characterizing the strongest possible global convergence result for the safeguarded augmented Lagrangian method.

✉ Roberto Andreani
andreani@unicamp.br

Gabriel Haeser
ghaeser@ime.usp.br

Leonardo M. Mito
leokoto@ime.usp.br

Alberto Ramos
albertoramos@ufpr.br

Leonardo D. Secchin
leonardo.secchin@ufes.br

[1] Department of Applied Mathematics, University of Campinas, Rua Sérgio Buarque de Holanda, 651, 13083-859, Campinas, SP, Brazil

[2] Department of Applied Mathematics, University of São Paulo, Rua do Matão, 1010, Cidade Universitária, 05508-090, São Paulo, SP, Brazil

[3] Department of Mathematics, Federal University of Paraná, 81531-980, Curitiba, PR, Brazil

[4] Department of Applied Mathematics, Federal University of Espírito Santo, Rodovia BR 101, Km 60, 29932-540, São Mateus, ES, Brazil

Keywords Nonlinear optimization · Augmented Lagrangian methods · Optimality conditions

Mathematics Subject Classification (2010) 90C46 · 90C30 · 65K05

1 Introduction

In this paper, we deal with the following problem:

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad h(x) = 0, \; g(x) \leqslant 0, \end{aligned} \tag{NLP}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, and $g : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions.

There is a variety of ways in which the *quality* of a general scheme for solving (NLP) may be measured. Most often, numerical stability, execution time, convergence rate, or other measures would be employed. However, in this paper, we are interested in the quality of the limit points of the sequences generated by the scheme, in the sense of how close they are to being necessarily optimal. This is an important complementary analysis of the reliability of an algorithm. Here, we restrict our analysis and conclusions only to augmented Lagrangian schemes. Usually, in most descriptions of algorithms for (NLP), the question of the quality of its limit points is answered in a simple way: a constraint qualification (CQ) such as the linear independence CQ or the Mangasarian-Fromovitz CQ is assumed at all feasible points of (NLP) and it is shown that all feasible limit points of a sequence generated by the algorithm satisfies the Karush-Kuhn-Tucker (KKT) conditions. Several weaker CQs have been considered in the recent years yielding stronger global convergence results based on weaker constraint qualifications such as RCPLD [10], CPG [11], CCP [15], and Quasinormality [4].

One may restate a global convergence result that a KKT point is achieved under a given CQ by saying that feasible limit points of an algorithm satisfy "KKT or not-CQ," meaning either KKT holds or that particular CQ is violated. This has the advantage that the latter is a true necessary optimality condition, satisfied at all local minimizers of (NLP), while KKT on its own may fail for specific problems. This simple approach has lead to the definition of different but genuine necessary optimality conditions that imply "KKT or not-CQ" but that are more tailored to global convergence analysis of algorithms. In particular, these have been called *sequential optimality conditions* [4, 7, 17, 40] and the sequences needed for checking the validity of the condition are precisely the primal and dual sequences generated by the algorithm.

The simplest example of a sequential optimality condition is the so-called Approximate-KKT (AKKT) condition, which is said to hold at a feasible point x^* whenever one may find a primal sequence $\{x^k\} \subset \mathbb{R}^n$, $x^k \to x^*$ and a dual sequence $\{(\lambda^k, \mu^k)\} \subset \mathbb{R}^m \times \mathbb{R}^p_+$ such that

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j: g_j(x^*)=0} \mu_j^k \nabla g_j(x^k) \to 0.$$

The simple observation that most reasonable algorithms generate sequences with this property is enough to recover most global convergence results to a KKT point under a constraint qualification. However, this is a poor analysis of the quality of x^*, as the sequences generated by the algorithm have a much more specific form. For a concrete example, consider the problem of minimizing $(x_1 - 1)^2 + (x_2 - 1)^2$ subject to $x_1 \geqslant 0, x_2 \geqslant 0, x_1 x_2 \leqslant 0$. Here, the only solutions are (1, 0) and (0, 1) but all feasible points satisfy AKKT [4]. However, it is known that an augmented Lagrangian method can only accumulate around (1, 0), (0, 1), or (0, 0) [32]. Also, even though an AKKT sequence is always generated, different augmented Lagrangian schemes will have different convergence properties (see, e.g., [36]) and in this paper we will investigate the impact on the quality of the limit point in view of several variants of augmented Lagrangian schemes, in particular taking into account how the dual sequence is computed.

In some sense, there is a mismatch between the sequences generated by the augmented Lagrangian algorithm and the sequences proved to exist converging to a local minimizer, since the latter is frequently built using a pure external penalty method on a regularized problem, and not the augmented Lagrangian. In this paper, we will use *the sequence itself*, generated by the augmented Lagrangian algorithm, to attest optimality of a limit point. More precisely, we show that *all* local minimizers are limit points of a sequence generated by the algorithm. This is a somewhat different approach from previous global convergence results via sequential optimality conditions, in the sense that there is no need to employ an optimality condition disconnected from the algorithm. The algorithm itself is used to study the quality of its limit points. With this tool at hand, we are able to investigate further what are the important features of the algorithm, in the sense that this property may be lost when modifying it. Surprisingly, we show that the algorithm may fail to achieve some local minimizers if exact stationary points are found in each subproblem or exact feasible points, hence, it is paramount to allow some degree of error when solving the subproblems. We also show that some simplified variants of the augmented Lagrangian method are equivalent to the standard algorithm in terms of the quality of its limit points, while a variant including a penalty parameter for each constraint induces a weaker optimality condition.

1.1 Contributions of this article

In this paper, we introduce a mathematical description of all feasible limit points of an augmented Lagrangian algorithm that also takes into account the sequences it generates (Definition 1). Then, using this description:

- we prove that being a feasible limit point of the algorithm is a necessary condition for local optimality (Theorem 2);
- we characterize the weakest constraint qualification required for proving global convergence to KKT points (Section 4);
- we prove that a growth control over the penalty parameter does not affect the quality of its limit points (Section 3.1), even though it provides a computational gain of stability;
- we provide strong evidences that suggest that the safeguarded augmented Lagrangian method is equivalent to the classical external penalty method in terms of the quality of their feasible limit points (Section 3.2 and Corollary 2);
- we show that using a different penalty parameter for each constraint is theoretically worse than using a common parameter (Section 3.4);
- we discuss the theoretical implications of employing a safeguarding technique for updating Lagrange multipliers (Theorem 5);
- we show that forcing exact feasibility or stationarity over the subproblems may possibly cause undesired effects on its convergence;
- we discuss other implementation details and their effects on the quality of limit points.

1.2 Notation

Our work environment is $\mathbb{R}^n$ equipped with the Euclidean norm, which is defined as $\|x\|_2 \doteq \sqrt{x_1^2 + \cdots + x_n^2}$ for every $x \in \mathbb{R}^n$. To attest algorithmic convergence, we employ the norm $\|x\|_\infty \doteq \max\{|x_i| : i \in \{1, \ldots, n\}\}$. The nonnegative orthant of $\mathbb{R}^n$ is denoted by $\mathbb{R}^n_+ \doteq \{x \in \mathbb{R}^n : x_i \geqslant 0, \forall i \in \{1, \ldots, n\}\}$ and, similarly, its positive orthant is $\mathbb{R}^n_{++} \doteq \{x \in \mathbb{R}^n : x_i > 0, \forall i \in \{1, \ldots, n\}\}$. It is well-known that the projection of a point $x \in \mathbb{R}^n$ onto $\mathbb{R}^n_+$ is given by $[x]_+ \doteq (\max\{0, x_1\}, \ldots, \max\{0, x_n\})$, and that $\nabla\|[x]_+\|_2^2 = 2[x]_+$, for every x.

The *Lagrangian function* of (NLP) is defined as $L(x, \lambda, \mu) \doteq f(x) + h(x)^T\lambda + g(x)^T\mu$, where $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$. Also, its gradient with respect to x is given by $\nabla_x L(x, \lambda, \mu) \doteq \nabla f(x) + \nabla h(x)^T\lambda + \nabla g(x)^T\mu$.

We denote sequences labeled by a variable x and indexed by a set $J \subseteq \mathbb{N}$ by $\{x^k\}_{k\in J}$, and in this context, $x^k \rightsquigarrow x^j$ means that x^j is the successor of x^k, where $j, k \in J$.

2 The PHR augmented Lagrangian algorithm

There exist many augmented Lagrangian variants in the literature, but the main subject of our analyses is the one known as the *Powell-Hestenes-Rockafellar* (PHR) algorithm [30, 43, 47], which is characterized by the following shifted penalty function:

$$L_{\rho,\bar{\lambda},\bar{\mu}}(x) \doteq f(x) + \frac{\rho}{2}\left[\left\|\frac{\bar{\lambda}}{\rho} + h(x)\right\|_2^2 + \left\|\left[\frac{\bar{\mu}}{\rho} + g(x)\right]_+\right\|_2^2\right], \tag{1}$$

where $\rho > 0$ and $\bar{\mu} \geqslant 0$. It is worth mentioning that a practical comparison among 65 distinct augmented Lagrangian variants was presented in [20], and the PHR was observed to have the best performance in their experiments.

As usual for penalty-type methods, the core idea behind the PHR algorithm is to minimize $L_{\rho,\bar{\lambda},\bar{\mu}}(x)$ successively until some stopping criterion is satisfied, increasing ρ whenever needed, and updating $\bar{\lambda}$ and $\bar{\mu}$ in a suitable way. There are many possible implementations for it, but in this paper our analyses revolve around a consolidated implementation known as ALGENCAN [2, 22], which has a robust practical implementation provided by the TANGO project (www.ime.usp.br/~egbirgin/tango) and a good track record of applications (for a description of applications in several fields, we refer to [22]).

The algorithm is usually presented as in Algorithm 1.

Algorithm 1 ALGENCAN.

Input: Parameters $\tau \in (0, 1)$ and $\gamma > 1$, a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ with $\lim_{k\to\infty} \varepsilon_k = 0$, an initial penalty parameter $\rho_1 > 0$, a compact set $\mathcal{B} \subseteq \mathbb{R}^m \times \mathbb{R}^p_+$, and some $(\bar{\lambda}^1, \bar{\mu}^1) \in \mathcal{B}$.

Initialize $k \leftarrow 1$. Then:

Step 1 *(Solving the subproblem):* Compute an approximate stationary point x^k of $L_{\rho_k,\bar{\lambda}^k,\bar{\mu}^k}(x)$, that is, a point that satisfies

$$\left\| \nabla L_{\rho_k,\bar{\lambda}^k,\bar{\mu}^k}(x^k) \right\|_\infty \leqslant \varepsilon_k.$$

Step 2 *(Updating the penalty parameter):* Calculate

$$V^k \doteq \left(h(x^k), \min\left\{ -g(x^k), \frac{\bar{\mu}^k}{\rho_k} \right\} \right).$$

Then,

a. If $k = 1$ or $\left\| V^k \right\|_\infty \leqslant \tau \left\| V^{k-1} \right\|_\infty$, set $\rho_{k+1} \doteq \rho_k$;

b. Otherwise, take ρ_{k+1} such that $\rho_{k+1} \geqslant \gamma \rho_k$.

Step 3 *(Estimating new projected multipliers):* Choose some $(\bar{\lambda}^{k+1}, \bar{\mu}^{k+1}) \in \mathcal{B}$, set $k \leftarrow k + 1$ and go to Step 1.

Remark 1 Strictly speaking, ALGENCAN allows for additional box-constraints in its subproblems, which are solved by an active-set strategy. However, this is not relevant to our analysis, so we abuse the notation by naming Algorithm 1 as ALGENCAN.

In many practical situations, $\mathcal{B}$ is defined as a box in the form $[\lambda_{\min}, \lambda_{\max}]^m \times [0, \mu_{\max}]^p$, where $\lambda_{\min} \leqslant \lambda_{\max}$ and $\mu_{\max} \geqslant 0$ are given. Also, the usual choice of $(\bar{\lambda}^{k+1}, \bar{\mu}^{k+1})$ is the projection of $(\lambda^k \doteq \bar{\lambda}^k + \rho_k h(x^k), \mu^k \doteq [\bar{\mu}^k + \rho_k g(x^k)]_+)$ onto $\mathcal{B}$, but a priori it can be any other element of $\mathcal{B}$. This technique is often called *safeguarding* and, in this context, $\bar{\lambda}^k$ and $\bar{\mu}^k$ are called *safeguarded multipliers*. Some advantages of safeguarding are discussed in Section 3.3. The vector V^k defined by (3) is a joint measure of feasibility and complementarity that is meant to control the growth of the penalty parameter through *Step 2*. In particular, note that ALGENCAN even allows $\{\rho_k\}_{k\in\mathbb{N}}$ to be bounded, which is a specially meaningfull feature when (NLP) is convex [46]. As a matter of fact, there is no universal rule for updating ρ_k, but it is important to keep in mind that, from the numerical point of view, the difficulty of solving the subproblems grows when ρ_k increases. Also, in situations where the augmented Lagrangian method is performing well, the penalty parameter can even be allowed to decrease [21], mainly because moderate values of ρ_k usually mean a better behavior of the box-constraint solver for the subproblem. Another work that deals with the possibility of decreasing the penalty parameter is [25]. This topic is out of the scope of this paper, but nevertheless we choose to give some degree of freedom for ρ_k for the sake of generality.

2.1 The sequences generated by the augmented Lagrangian method

In order to build a rigorous analysis of ALGENCAN, it is fundamental to introduce a consistent mathematical characterization of its output sequences and their limit points. We get inspiration from the structure of sequential optimality conditions to encapsulate every feasible outcome of ALGENCAN in the following definition:

Definition 1 We say that x^* is an *Augmented Lagrangian AKKT* (AL-AKKT) point if there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to x^*$, $\{\rho_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$, and bounded sequences $\{\bar{\lambda}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m$, $\{\bar{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, such that

$$\nabla_x L(x^k, \lambda^k, \mu^k) \to 0 \quad \text{and} \quad V^k \to 0 \tag{2}$$

where $\lambda^k \doteq \bar{\lambda}^k + \rho_k h(x^k)$, $\mu^k \doteq [\bar{\mu}^k + \rho_k g(x^k)]_+$, and

$$V^k \doteq \left(h(x^k), \min\left\{ -g(x^k), \frac{\bar{\mu}^k}{\rho_k} \right\} \right). \tag{3}$$

In this context, we say that $\{x^k\}_{k\in\mathbb{N}}$ is a (primal) AL-AKKT sequence associated with the dual sequence $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$.

Although it is intuitively clear that Definition 1 is indeed a complete characterization of the output sequences of ALGENCAN, this is not obvious whatsoever. To make this relationship clear, we define an *instance* of ALGENCAN as a two-phase procedure consisting of:

1. Choosing the initial parameters: τ, γ, $\{\varepsilon^k\}_{k\in\mathbb{N}}$, ρ_1, $\mathcal{B}$, and $(\bar{\lambda}_1, \bar{\mu}_1)$;
2. Running ALGENCAN, producing sequences $\{x^k\}_{k\in\mathbb{N}}$ and $\{\lambda^k, \mu^k\}_{k\in\mathbb{N}}$.

Also, we say that an instance of ALGENCAN *generates* a given point x^* when x^* is an accumulation point of its output sequence $\{x^k\}_{k\in\mathbb{N}}$.

In the following lines, we present a formal argument that establishes the correspondence between the algorithmic representation of ALGENCAN and Definition 1.

Theorem 1 *For each AL-AKKT point x^*, there is an instance of* ALGENCAN *such that its output sequence $\{x^k\}_{k\in\mathbb{N}}$ has x^* as an accumulation point. Conversely, every feasible accumulation point generated by* ALGENCAN *is AL-AKKT.*

Proof Let $\{x^k\}_{k\in\mathbb{N}}$ be an AL-AKKT sequence convergent to x^*, with associated safeguarded multipliers $\{\bar{\lambda}^k\}_{k\in\mathbb{N}}$ and $\{\bar{\mu}^k\}_{k\in\mathbb{N}}$, and parameters $\{\rho_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$. We will split the proof in two cases: the first one is when $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$, which means that for any $\gamma > 1$, there is a subsequence of it such that $\rho_{k+1} \geqslant \gamma\rho_k$ for every k. We denote the index set of this subsequence by $K = \{i_1, i_2, i_3, \ldots\} \subset \mathbb{N}$. Then, fix $\tau \in (0, 1)$ and define the index set

$$K_= \doteq \left\{ i_k \geqslant 2 \mid \left\| V^{i_k} \right\| \leqslant \tau \left\| V^{i_{k-1}} \right\|_\infty \right\} \subset K.$$

This set may be finite or infinite. Despite of this, we define a set I and new sequences $\{\tilde{x}^j\}_{j\in I}$, $\{\tilde{\lambda}^j\}_{j\in I}$, $\{\tilde{\mu}^j\}_{j\in I}$, $\{\tilde{\rho}_j\}_{j\in I}$ in the following way:

1. Start with $\tilde{x}^{i_1} \doteq x^{i_1}$, $\tilde{\lambda}^{i_1} \doteq \bar{\lambda}^{i_1}$, $\tilde{\mu}^{i_1} \doteq \bar{\mu}^{i_1}$ and $\tilde{\rho}_{i_1} \doteq \rho_{i_1}$;
2. For each $i_k \in K\backslash\{i_1\}$,
   (a) if $i_k \in K_=$ then define the next elements $\tilde{x}^{i_k-\frac{1}{2}} \doteq x^{i_k}$, $\tilde{\lambda}^{i_k-\frac{1}{2}} \doteq \bar{\lambda}^{i_k}$, $\tilde{\mu}^{i_k-\frac{1}{2}} \doteq \bar{\mu}^{i_k}$ and $\tilde{\rho}_{i_k-\frac{1}{2}} \doteq \rho_{i_k}$;
   (b) independently if i_k is in $K_=$ or not, define the next elements $\tilde{x}^{i_k} \doteq x^{i_k}$, $\tilde{\lambda}^{i_k} \doteq \bar{\lambda}^{i_k}$, $\tilde{\mu}^{i_k} \doteq \bar{\mu}^{i_k}$ and $\tilde{\rho}_{i_k} \doteq \rho_{i_k}$,

and I is defined as K with duplicated elements whenever they belong to $K_=$. For example, if $K_= = \{i_2, i_3, i_6\}$ then

$$\{\tilde{x}^j\}_I = \{\, x^{i_1},\ x^{i_2-\frac{1}{2}} = x^{i_2},\ x^{i_2},\ x^{i_3-\frac{1}{2}} = x^{i_3},\ x^{i_3},\ x^{i_4},\ x^{i_5},\ x^{i_6-\frac{1}{2}} = x^{i_6},\ x^{i_6},\ \ldots\}$$

and $I = \{i_1, i_2, i_2, i_3, i_3, i_4, i_5, i_6, i_6, \ldots\}$ (for simplicity, we abuse the notation here by duplicating indexes). The other sequences are analogous. In other words, we duplicate an element of every sequence whenever its index is in $K_=$. In the next lines, the vectors $\tilde{V}^j$ are defined as in (3) for those new sequences.

There are the following possibilities for any two consecutive elements:

- $\tilde{x}^{i_{k-1}} \rightsquigarrow \tilde{x}^{i_k-\frac{1}{2}}$. In this case, $i_k \in K_=$ and then $\left\| \tilde{V}^{i_k-\frac{1}{2}} \right\|_\infty = \|\tilde{V}^{i_k}\|_\infty \leqslant \tau\|\tilde{V}^{i_{k-1}}\|_\infty$. Note that, by definition, $\tilde{\rho}_{i_k} = \tilde{\rho}_{i_k-\frac{1}{2}}$;
- $\tilde{x}^{i_k-\frac{1}{2}} \rightsquigarrow \tilde{x}^{i_k}$. Here, $\|\tilde{V}^{i_k}\|_\infty = \|\tilde{V}^{i_k-\frac{1}{2}}\|_\infty > \tau\|\tilde{V}^{i_k-\frac{1}{2}}\|_\infty$. The next element is $\tilde{x}^{i_{k+1}}$ or $\tilde{x}^{i_{k+1}-\frac{1}{2}}$. In both cases, $(\tilde{\rho}_{i_{k+1}-\frac{1}{2}} =)\tilde{\rho}_{i_{k+1}} > \tilde{\rho}_{i_k}$;
- $\tilde{x}^{i_{k-1}} \rightsquigarrow \tilde{x}^{i_k}$. In this case $i_k \notin K_=$ and thus $\|\tilde{V}^{i_k}\|_\infty > \tau \left\| \tilde{V}^{i_{k-1}} \right\|_\infty$. Again, the next element is $\tilde{x}^{i_{k+1}}$ or $\tilde{x}^{i_{k+1}-\frac{1}{2}}$, from which $(\tilde{\rho}_{i_{k+1}-\frac{1}{2}} =)\tilde{\rho}_{i_{k+1}} > \tilde{\rho}_{i_k}$.

Thus, we conclude that in this case the new sequences satisfy, consecutively, all requirements of ALGENCAN with safeguarded multipliers $\{\bar{\lambda}^j\}_{j\in I}$ and $\{\bar{\mu}^j\}_{j\in I}$, precision parameters $\varepsilon_j \doteq \left\|\nabla_x L(\tilde{x}^j, \lambda^j, \mu^j)\right\|_\infty$, where $\lambda^j = \tilde{\lambda}^j + \tilde{\rho}_j h(\tilde{x}^j)$ and $\mu^j = [\tilde{\mu}^j + \tilde{\rho}_j g(\tilde{x}^j)]_+$ for every $j \in I$, and any compact set $\mathcal{B}$ that contains every safeguarded multiplier. Then, since $\lim_{j\to\infty} \tilde{x}^j = x^*$, there is an instance of ALGENCAN that generates x^*. In order to see this, keep in mind that

$$\nabla_x L(x^k, \lambda^k, \mu^k) = \nabla L_{\rho_k, \bar{\lambda}^k, \bar{\mu}^k}(x^k),$$

for every $k \in \mathbb{N}$.

The second case is when $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded, which means x^* satisfies the KKT conditions since in this case $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$ is bounded and $V^k \to 0$ ensures the fulfilment of the complementarity conditions. Denote its associated Lagrange multipliers by $\tilde{\lambda}$ and $\tilde{\mu}$. Then, choose any compact set $\mathcal{B}$ such that $(\tilde{\lambda}, \tilde{\mu}) \in \mathcal{B}$ and define $(\bar{\lambda}^k, \bar{\mu}^k) \doteq (\tilde{\lambda}, \tilde{\mu})$, $\rho_1 \doteq 1$, and $x^k \doteq x^*$, for every $k \in \mathbb{N}$. Thus, there is an instance of ALGENCAN that generates x^* in this case, as well.

The converse is trivial, since $V^k \to 0$ independently on the boundedness of $\{\rho_k\}_{k\in\mathbb{N}}$. □

In light of Theorem 1, AL-AKKT provides an ideal comparison tool between ALGENCAN and some of its variants, as well as a proper language for analysing the effects of some particular choices of parameters on the output of the algorithm. This is the main focus of Section 3. Before that, we discuss from the sequential optimality conditions perspective, some other interesting properties of ALGENCAN that are very often difficult to perceive with the algorithmic language.

2.2 AL-AKKT is necessary for optimality

A common approach for building the convergence theory of an algorithm is to ground it on some universal necessary optimality condition, that is, a condition that is independent of the algorithm (whether it is of the sequential type or not). However, such independence does not add any value for the algorithm at all and it may even turn into a limitation for its convergence theory. The following theorem shows that AL-AKKT can be interpreted as a necessary optimality condition, but since it is also an exact description of the augmented Lagrangian method, it provides the best possible convergence theory for it.

Theorem 2 *Every local minimizer x^* of (NLP) is an AL-AKKT point, regardless of the choice of $\{(\bar{\lambda}^k, \bar{\mu}^k)\}_{k\in\mathbb{N}}$.*

Proof Let x^* be a local minimizer of (NLP). Then, there is a $\delta > 0$ such that x^* is the unique global solution of the localized problem

$$\begin{aligned} &\underset{x\in\mathbb{R}^n}{\text{minimize}}\ f(x) + 1/2\|x - x^*\|_2^2 \\ &\text{subject to}\ h(x) = 0,\quad g(x) \leqslant 0,\quad \|x - x^*\|_2 \leqslant \delta. \end{aligned} \tag{4}$$

The proof relies on arguments that are similar to the ones used in [22, Theorem 5.2]. In summary, the idea is to apply Algorithm 1 to (4), but assuming that we are able to compute a global solution of each subproblem in *Step 1* under the constraint $\|x - x^*\|_2 \leqslant \delta$, instead of only minimizing the correspondent augmented

Lagrangian (1) up to first-order stationarity. Then, for each $k \in \mathbb{N}$, let x^k be a global solution of the augmented subproblem

$$\begin{aligned} &\underset{x\in\mathbb{R}^n}{\text{minimize}}\ f(x) + \frac{1}{2}\|x - x^*\|_2^2 + \frac{\rho_k}{2}\sum_{i=1}^{m}\left[\frac{\bar{\lambda}_i^k}{\rho_k} + h_i(x)\right]^2 + \frac{\rho_k}{2}\sum_{j=1}^{p}\left[\frac{\bar{\mu}_j^k}{\rho_k} + g_j(x)\right]_+^2 \\ &\text{subject to } \|x - x^*\|_2 \leqslant \delta. \end{aligned} \tag{5}$$

Furthermore, the choice in *Step 3* can be made to match any sequence $\{(\bar{\lambda}^k, \bar{\mu}^k)\}_{k\in\mathbb{N}}$ chosen a priori, highlighting that our conclusions are independent of this choice.

Let us show that $\{x^k\}_{k\in\mathbb{N}}$ converges to x^*. By the optimality of x^k and the feasibility of x^* we have, for all k, that

$$\begin{aligned} &f(x^k) + \frac{1}{2}\left\|x^k - x^*\right\|_2^2 + \frac{\rho_k}{2}\sum_{i=1}^{m}\left[\frac{\bar{\lambda}_i^k}{\rho_k} + h_i(x^k)\right]^2 + \frac{\rho_k}{2}\sum_{j=1}^{p}\left[\frac{\bar{\mu}_j^k}{\rho_k} + g_j(x^k)\right]_+^2 \\ &\leqslant f(x^*) + \sum_{i=1}^{m}\frac{(\bar{\lambda}_i^k)^2}{2\rho_k} + \sum_{j=1}^{p}\frac{(\bar{\mu}_j^k)^2}{2\rho_k}. \end{aligned} \tag{6}$$

Then, for all k, we have

$$\left(\frac{\bar{\lambda}^k}{\rho_k} + h(x^k)\,,\ \left[\frac{\bar{\mu}^k}{\rho_k} + g(x^k)\right]_+\right) = \left(\frac{\bar{\lambda}^k}{\rho_k}\,,\ \frac{\bar{\mu}^k}{\rho_k}\right) + \tilde{V}^k,$$

where $\tilde{V}^k \doteq (h(x^k), \max\{g(x^k), -\bar{\mu}^k/\rho_k\})$. Observe that $|\tilde{V}_\ell^k| = |V_\ell^k|$ for all $k \in \mathbb{N}$ and $\ell = 1, \dots, m+p$. Thus, by (6), we get that

$$f(x^k) + \frac{1}{2}\left\|x^k - x^*\right\|_2^2 + \sum_{i=1}^{m}\bar{\lambda}_i^k\tilde{V}_i^k + \sum_{j=1}^{p}\bar{\mu}_j^k\tilde{V}_j^k + \frac{\rho_k}{2}\left\|\tilde{V}^k\right\|_2^2 \leqslant f(x^*). \tag{7}$$

Now, let us take an accumulation point $\bar{x}$ of $\{x^k\}_{k\in\mathbb{N}}$. There are two cases to consider:

- if $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded, then, by *Step 2* we have $\left\|\tilde{V}^k\right\|_2 = \left\|V^k\right\|_2 \to 0$ and therefore $\bar{x}$ is feasible. From (7), we have $f(\bar{x}) + (1/2)\|\bar{x} - x^*\|_2^2 \leqslant f(x^*)$. The optimality of x^* and the feasibility of $\bar{x}$ imply $\bar{x} = x^*$;
- if $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$ then dividing (7) by ρ_k and taking limits lead us to obtain $\left\|V^k\right\|_2 = \left\|\tilde{V}^k\right\|_2 \to 0$, and thus $\bar{x}$ is feasible. Again by (7), we have $f(\bar{x}) + (1/2)\|\bar{x} - x^*\|_2^2 \leqslant f(x^*)$ and hence $\bar{x} = x^*$.

Thus, $x^k \to \bar{x} = x^*$ and $\left\|x^k - x^*\right\|_2 < \delta$ for all k large enough. Then, the optimality conditions for the penalized problem (5) at x^k give us

$$\nabla f(x^k) + \nabla h(x^k)[\bar{\lambda}^k + \rho_k h(x^k)] + \nabla g(x^k)[\bar{\mu}^k + \rho_k g(x^k)]_+ = x^* - x^k \to 0.$$

We also proved that $V^k \to 0$ regardless of whether $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded or not, concluding the proof. □

By Theorem 1, it is possible to say that the kind of convergence analysis provided by AL-AKKT via Theorem 2 is stronger than all known previous results regarding both: optimality conditions in the form "KKT or not-CQ" and sequential conditions. Indeed, every necessary condition that supports ALGENCAN must be satisfied by all of its feasible limit points, and consequently implied by AL-AKKT. A detailed discussion on this topic can be found in Section 4, where we characterize the weakest CQ needed to establish convergence of ALGENCAN to KKT points, by means of AL-AKKT.

Remark 2 Similarly to the way Lagrange multipliers attest optimality of a KKT point in some sense, the sequences provided by an execution of ALGENCAN can be seen as optimality certificates of an AL-AKKT point. In order to see this, keep in mind that even though such certificates were generated by a regularized variant of the method in the proof of Theorem 2, they are also valid outputs of ALGENCAN with a suitable choice of parameters.

We highlight that Theorems 1 and 2 also mean that the mere fact of being a feasible limit point of ALGENCAN is itself a necessary optimality condition, which to the best of our knowledge, is novelty. While the usual convergence statement of ALGENCAN tells us that the method is expected to converge to a local minimizer under some conditions, Theorem 1 complements it by stating that every local minimizer is likely to be found. This attribute should not be ignored in nonconvex problems with multiple local minima, since some local minimizers may be more interesting than the others (for instance, the global minimizer, when it does exist). In fact, whether this is an advantage for the method may be a situational issue, but the existence of some solutions that can be avoided by the method without specifying it should always raise a red flag. For instance, this is the case of an augmented Lagrangian for Generalized Nash Equilibrium problems. See the discussion in [24].

Remark 3 Every KKT point is AL-AKKT as well. The reasoning is similar to the second case of the proof of Theorem 1: if x^* is KKT, let us say, with associated Lagrange multipliers $\tilde{\lambda}$ and $\tilde{\mu}$, then taking $x^k \doteq x^*$, $\rho_k \doteq k$, $\bar{\lambda}^k \doteq \tilde{\lambda}$, and $\bar{\mu}^k \doteq \tilde{\mu}$, for every $k \in \mathbb{N}$, is enough to conclude that x^* is also AL-AKKT.

Remark 4 If x^* is an AL-AKKT point and $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$ is bounded (in particular, if $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded), then it is a KKT point. In fact, in this case any accumulation point of $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$ serves as Lagrange multipliers, and $V^k \to 0$ ensures the complementarity condition.

A consequence of Theorem 2 and Remark 3 is that every possible point of interest for (NLP) is in the range of convergence of ALGENCAN.

2.3 About inexactly solving the subproblems

Note that *Step 1* only requires x^k to be an approximate stationary point of the Lagrangian function, $L_{\rho_k,\bar{\lambda}^k,\bar{\mu}^k}(x)$. This is done in order to maintain some degree of loyalty to what happens in practice, since either way x^k is likely to be obtained by some iterative method for solving an unconstrained subproblem, that declares convergence when the Lagrangian residue is small enough. Thus, we stress that we do not require any use of global methods for solving the subproblems, and similarly, no convexity assumptions are enforced.

However, it is natural to expect that the introduction of any kind of error tolerance might have negative influence over the outcome of the method. In [35], the authors study five *relaxation methods* for *mathematical problems with equilibrium constraints*, in particular, they compare the effects of forcing exactness on the subproblems (that is, an analogue of setting $\varepsilon_k = 0$ for every k in ALGENCAN) with the effects of allowing the subproblems to be solved inexactly. One of those methods was proven to keep the same convergence properties regardless of the exactness of the subproblems, whereas the others presented a strictly worse notion of convergence with inexactly solved subproblems.

Surprisingly, concerning ALGENCAN (via AL-AKKT), forcing exact stationarity by means of setting $\varepsilon_k = 0$ for every k may lead to unexpected and possibly undesired results, such as the exclusion of some local/global minimizers from the range of convergence of the method. The following example illustrates this fact:

Example 1 In $\mathbb{R}^2$, consider the minimization problem:

$$\text{minimize } -x_2 \text{ subject to } x_2^2 \leqslant 0, \quad x_1^2 x_2^2 \leqslant 0.$$

Clearly, $x^* \doteq (1, 0)$ is a global minimizer. Now, assume that there exists a sequence of stationary points of the Lagrangian $x^k \doteq (x_1^k, x_2^k)$ converging to x^* for some $\bar{\mu}_1^k$, $\bar{\mu}_2^k \in \mathbb{R}_+$. Thus, $\nabla L_{\rho^k, \bar{\mu}^k}(x^k)$ is equal to

$$\begin{bmatrix} 0 \\ -1 \end{bmatrix} + (\bar{\mu}_1^k + \rho_k (x_2^k)^2)_+ \begin{bmatrix} 0 \\ 2x_2^k \end{bmatrix} + (\bar{\mu}_2^k + \rho_k (x_1^k)^2 (x_2^k)^2)_+ \begin{bmatrix} 2x_1^k (x_2^k)^2 \\ 2x_2^k (x_1^k)^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

From the first line of the expression above, we see that $2x_1^k (x_2^k)^2 = 0$ or $(\bar{\mu}_2^k + \rho_k (x_1^k)^2 (x_2^k)^2) = 0$, for all $k \in \mathbb{N}$. In any case, since $x_1^k \to 1$, we get $x_2^k = 0$ for all k, contradicting the second expression.

Hence, the notion of convergence provided by Theorems 1 and 2 can be weakened under exactness, in contrast with our previous discussion. Moreover, Example 1 suggests that we should raise our caution for choosing the precision parameters ε_k, for the behavior of the method due to this choice is not trivially predictable. We refer to [18, 27] for some efficient strategies for choosing ε_k.

Similarly, it is also possible to prove that the augmented Lagrangian scheme can be incompatible with feasible sequences. In fact, forcing $V^k = 0$ for all k may not only remove minimizers from the range of convergence of ALGENCAN, but can also induce complete failure in the method. The following example illustrates this fact:

Example 2 In $\mathbb{R}$, consider the minimization problem:

$$\text{minimize } x_1 \text{ subject to } x_1^2 \leqslant 0.$$

Clearly $x^* \doteq 0$ is the only global minimizer. Now, assume that there exists a sequence of exact admissible points x^k converging to x^* and multipliers $\bar{\mu}^k \in \mathbb{R}_+$, such that $1 + [\bar{\mu}^k + \rho_k (x_1^k)^2]_+ x_1^k \to 0$. Forcing $V^k = 0$ we get that $x_1^k = 0$, which is a contradiction with $1 + [\bar{\mu}^k + \rho_k (x_1^k)^2]_+ x_1^k \to 0$.

Note that this issue does not depend on the choice of $\bar{\mu}^k$. In fact, it is well-known that the feasibility of the limit point is a paramount issue for numerical methods based on penalty approaches, which is a huge contrast with *interior-point* and *active-set* methods, where feasibility is always maintained. Example 2 indicates that any attempt of resolving this issue by forcing feasibility in ALGENCAN may actually hinder its theoretical convergence.

In some situations, one may be interested in computing the Lagrange multipliers associated with a KKT point. Even though the KKT point itself can always be found by ALGENCAN, the choice of parameters may have direct influence on whether its Lagrange multipliers are computed along with it or not. The example below illustrates a scenario where a KKT point is found by ALGENCAN, but its Lagrange multiplier is not.

Example 3 In $\mathbb{R}^2$, consider the minimization problem:

$$\text{minimize } -x_1^2 \text{ subject to } x_1^2 x_2 = 0.$$

Observe that $x^* \doteq (0, 0)$ is a KKT point whose set of multipliers is the whole $\mathbb{R}$. Consider $x_1^k \doteq 1/k$, $x_2^k \doteq x_1^k$, $\rho_k \doteq (2x_1^k x_2^k)^{-2}$ and $\bar{\lambda}^k \doteq 0$. For this choice, we see that

$$\begin{bmatrix} -2x_1^k \\ 0 \end{bmatrix} + \rho_k (x_1^k)^2 x_2^k \begin{bmatrix} 2x_1^k x_2^k \\ (x_1^k)^2 \end{bmatrix} \to \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

However, observe that the approximate multiplier diverges:

$$\lambda^k \doteq \rho_k (x_1^k)^2 x_2^k = \frac{1}{2x_2^k} \to \infty.$$

Thus, since $\{(x_1^k, x_2^k)\}_{k\in\mathbb{N}}$ is a valid AL-AKKT sequence certified by $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{\bar{\lambda}^k\}_{k\in\mathbb{N}}$, we can suppose without loss of generality, by the proof of Theorem 1, that the whole sequences $\{x^k\}_{k\in\mathbb{N}}$ and $\{\lambda^k\}_{k\in\mathbb{N}}$ are valid outputs of ALGENCAN where x^* is found, but its Lagrange multiplier is not.

The next section deepens the discussion about implementation details of ALGENCAN by formally comparing some modifications and parameter choices that are often considered in practice, by means of AL-AKKT.

3 Comparing some augmented Lagrangian variants

There are various traits that distinguish ALGENCAN from other similar penalty-based algorithms, such as the *classical external penalty method* [42, Framework 17.1] and the *shifted external penalty method* [38, Algorithm 3.1]. However, a theoretical comparison among them in terms of their limit points has not been made yet, to the best of our knowledge. Nevertheless, ALGENCAN is often preferred over such variants since it allows for ρ_k to remain bounded under some circumstances, which indicates that it is numerically more stable than the others. In this section, we address this question via AL-AKKT. We also investigate the use of safeguarding and the effects of employing a different penalty parameter for each constraint.

3.1 The shifted external penalty method

The shifted external penalty method is a variant of ALGENCAN without the admissibility control for controlling the growth of ρ_k, that is, where *Step 2* is replaced by *Step 2-b*. In this section, we call this variant SHIFTED-EP.

Algorithm 2 SHIFTED-EP.

Input: A parameter $\gamma > 1$, a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ with $\lim_{k\to\infty} \varepsilon_k = 0$, an initial penalty parameter $\rho_1 > 0$, a compact set $\mathcal{B}$, and some $(\bar{\lambda}^1, \bar{\mu}^1) \in \mathcal{B}$.
Initialize $k \leftarrow 1$. Then:
Step 1 *(Solving the subproblem):* Find an approximate stationary point x^k of $L_{\rho_k,\bar{\lambda}^k,\bar{\mu}^k}(x)$, that is, a point that satisfies

$$\left\| \tilde{V}^{i_k - \frac{1}{2}} \right\|_\infty \leqslant \varepsilon_k.$$

Step 2-b *(Updating the parameters):* Take ρ_{k+1} such that $\rho_{k+1} \geqslant \gamma\rho_k$;
Step 3 *(Estimating new projected multipliers):* Choose some $(\bar{\lambda}^{k+1}, \bar{\mu}^{k+1}) \in \mathcal{B}$, set $k \leftarrow k+1$, and go to Step 1.

In practice, ALGENCAN is expected to be numerically more stable than SHIFTED-EP due to more control on the growth of ρ_k. In theoretical terms, however, we see that there is no difference between them, which is somewhat surprising.

In order to make our argument clear, we present an alternative characterization of AL-AKKT:

Proposition 1 *A feasible point x^* is AL-AKKT if, and only if, there exist some sequences $\{x^k\}_{k\in\mathbb{N}} \to x^*$, $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$, and bounded sequences $\{\bar{\lambda}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m$, $\{\bar{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, such that $\nabla_x L(x^k, \lambda^k, \mu^k) \to 0$, where $\lambda^k \doteq \bar{\lambda}^k + \rho_k h(x^k)$ and $\mu^k \doteq [\bar{\mu}^k + \rho_k g(x^k)]_+$.*

Proof If x^* is AL-AKKT, then there are sequences $\{x_0^k\}_{k\in\mathbb{N}} \to x^*$, $\{\rho_{k,0}\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$, $\{\bar{\lambda}_0^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m$, and $\{\bar{\mu}_0^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, such that (2) holds. If $\rho_{k,0} \to \infty$, there is nothing left to prove. Otherwise x^* satisfies the KKT conditions (Remark 4) with, let us say, multipliers $\tilde{\lambda}$ and $\tilde{\mu}$. Consequently, due to Remark 3, x^* is AL-AKKT associated with the new sequences defined by $x^k \doteq x^*$, $\rho_k \doteq k(\to \infty)$, $\bar{\lambda}^k \doteq \tilde{\lambda}$, and $\bar{\mu}^k \doteq \tilde{\mu}$, for every $k \in \mathbb{N}$. The converse is straightforward from the feasibility of x^* along with the fact $V^k \to 0$ when $\rho_k \to \infty$. □

This characterization turns out to be simpler than Definition 1 since it does not rely on the computation of V^k. However, Definition 1 is not completely replaceable because when (NLP) is convex, the augmented Lagrangian is guaranteed to globally converge with a fixed value of $\rho > 0$, hence the admissibility criterion of *Step 2* of ALGENCAN is always satisfied when k is large enough and $\{\rho_k\}_{k\in\mathbb{N}}$ is always bounded [46]. Even so, the characterization of Proposition 1 is strongly related to SHIFTED-EP, which allows us to conclude the following:

Theorem 3 *For each AL-AKKT point x^*, there is an instance of* SHIFTED-EP *such that its output sequence $\{x^k\}_{k\in\mathbb{N}}$ has x^* as an accumulation point. Conversely, every feasible accumulation point generated by* SHIFTED-EP *is AL-AKKT.*

Proof Analogous to Theorem 1, in view of Proposition 1. □

However, the same conclusion is not necessarily valid for the classical external penalty method, which is the main focus of the next section.

3.2 The classical external penalty method

The classical external penalty method, which we call CLASSICAL-EP in this section, is precisely SHIFTED-EP with $\mathcal{B} = \{0\}$. In the previous section, we proved that ALGENCAN is equivalent to SHIFTED-EP in terms of their feasible limit points, so it is natural to question whether the same holds true for CLASSICAL-EP. The answer might seem obvious, for it is tempting to conclude that ALGENCAN and CLASSICAL-EP behave similarly, just because the shifts $\bar{\lambda}^k/\rho_k$ and $\bar{\mu}^k/\rho_k$ of ALGENCAN are small when ρ_k is large, but this is not entirely correct. The following example shows that the collection of all possible limit points of ALGENCAN contains, but may not be contained in the set of limit points of CLASSICAL-EP, and these extra points may be infeasible, even when $\rho_k \to \infty$. This is not necessarily a negative trait since we are mainly interested in feasible points. The purpose of the example is to emphasize that even when the shift is small, minimizing a shifted penalty function is not the same as minimizing the original penalty function.

Example 4 Consider the following problem

$$\text{minimize} \ -x \ \text{ subject to } \ \sin(x) = 0, \ \ \cos(x) = 0,$$

which is infeasible, which means that the admissibility test of *Step 2* of ALGENCAN will succeed only a finite number of times, so $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$. In this scenario, a point x^* can be reached by ALGENCAN when there is some sequence $\{x^k\}_{k\in\mathbb{N}} \to \bar{x}$ and a bounded $\{\bar{\lambda}^k\}_{k\in\mathbb{N}}$ such that

$$-1 + \cos(x^k)(\bar{\lambda}_1^k + \rho_k \sin(x^k)) - \sin(x^k)(\bar{\lambda}_2^k + \rho_k \cos(x^k)) \to 0, \tag{8}$$

which holds if, and only if, $\cos(x^k)\bar{\lambda}_1^k - \sin(x^k)\bar{\lambda}_2^k \to 1$. This is never satisfied when $\bar{\lambda}^k = 0$, so CLASSICAL-EP will never converge when applied to this problem. However, for every $x^* \in \mathbb{R}$, it is always possible to choose $\bar{\lambda}^k$ such that x^* is reached by ALGENCAN.

Knowing that the behaviors of ALGENCAN and CLASSICAL-EP may differ when taking infeasible points into consideration, what is left is to compare their feasible limit points, but this is also not trivial, and a careful analysis must be done. A direct consequence of Theorem 2 is that the set of common limit points between ALGENCAN and CLASSICAL-EP contains at least every local minimizer. Next, we show that the same conclusion is also true for KKT points.

Theorem 4 *If x^* is a KKT point, there is some instance of* CLASSICAL-EP *such that its output sequence $\{x^k\}_{k\in\mathbb{N}}$ has x^* as an accumulation point.*

Proof Let x^* be a KKT point of (NLP), and consider the sets

$$K(x) \doteq \{\nabla h(x)\lambda + \nabla g(x)\mu \mid \mu \geqslant 0,\ \mu_j = 0 \text{ if } g_j(x^*) < 0\} \tag{9}$$

and

$$K(x,\rho) \doteq \{\nabla h(x)[\rho h(x)] + \nabla g(x)[\rho g(x)]_+\}, \tag{10}$$

where $\rho > 0$. Clearly, the fact that x^* is a KKT point is equivalent to $-\nabla f(x^*) \in K(x^*)$. Note that x^* is a limit of a sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by CLASSICAL-EP if $x^k \to x^*$ and

$$\nabla f(x^k) + \nabla h(x^k)[\rho_k h(x^k)] + \nabla g(x^k)[\rho_k g(x^k)]_+ \to 0$$

for a certain sequence $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$. In other words, the classical external penalty is capable of reaching x^* if

$$-\nabla f(x^*) \in \limsup_{x\to x^*,\ \rho\to\infty} K(x,\rho).$$

So, in order to prove that x^* can be reached by CLASSICAL-EP, it is sufficient to prove that

$$K(x^*) \subset \limsup_{x\to x^*,\ \rho\to\infty} K(x,\rho).$$

Let $\mathcal{T}(x^*)$ be the tangent cone to the feasible set of (NLP) at x^*, and let $\mathcal{L}(x^*)$ be its linearized cone. It is always true that $\mathcal{T}(x^*) \subset \mathcal{L}(x^*)$, which implies $K(x^*) = \mathcal{L}(x^*)^\circ \subset \mathcal{T}(x^*)^\circ$. Now, given $w^* \in K(x^*)$, Lemma 4.3 of [15] ensures that there are sequences $\{w^k\}_{k\in\mathbb{N}} \to w^*$ and $\{\tilde{x}^k\}_{k\in\mathbb{N}} \to x^*$ such that

$$w^k = \nabla h(\tilde{x}^k)[kh(\tilde{x}^k)] + \nabla g(\tilde{x}^k)[kg(\tilde{x}^k)]_+.$$

As $w^k \in K(\tilde{x}^k, k)$ for all k, we have $w^* \in \limsup_{x\to x^*,\ \rho\to\infty} K(x,\rho)$, concluding the proof. □

To the best of our knowledge, the fact that CLASSICAL-EP may converge to any KKT point (Theorem 4) is new in the literature. As an immediate consequence of this fact, we see that if $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$ remains bounded in Definition 1 (in particular, if $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded), then ALGENCAN and CLASSICAL-EP are indistinguishable in terms of the quality of their limit points. Formally:

Corollary 1 *If x^* is an AL-AKKT point such that $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$ is bounded, then there is an associated AL-AKKT sequence $\{x^k\}_{k\in\mathbb{N}}$ with $(\bar{\lambda}^k, \bar{\mu}^k) \doteq (0,0)$ for all k. That is, in this case every feasible limit point of* ALGENCAN *can be reached by* CLASSICAL-EP *and vice-versa.*

Proof Since $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$ is bounded, x^* is a KKT point (Remark 4). So, the conclusion follows from Theorem 4. □

It is still not clear, for an arbitrary choice of $\bar{\lambda}^k$ and $\bar{\mu}^k$, whether every limit point of ALGENCAN is reachable by an instance of CLASSICAL-EP or not. Nevertheless, we conjecture they are equivalent regarding their feasible limit points. Some facts that support our belief are discussed in Section 3.4 and in Section 4, within Theorem 6 and Corollary 2.

3.3 On multiplier updates and safeguards

We presented ALGENCAN with a safeguarding procedure, where $\bar{\lambda}^k$ and $\bar{\mu}^k$ are taken from a compact set, but the original augmented Lagrangian algorithm proposed by Hestenes and Powell [30, 43] employs a different strategy. In their work, they use

C1. $\bar{\lambda}^{k+1} \doteq \bar{\lambda}^k + \rho_k h(x^k)$ and $\bar{\mu}^{k+1} \doteq [\bar{\mu}^k + \rho_k g(x^k)]_+$ for every $k \in \mathbb{N}$,

that is, $(\bar{\lambda}^{k+1}, \bar{\mu}^{k+1}) \doteq (\lambda^k, \mu^k)$ for every k, regardless of $\mathcal{B}$, which means $\{\bar{\lambda}^k\}_{k\in\mathbb{N}}$ and $\{\bar{\mu}^k\}_{k\in\mathbb{N}}$ are not necessarily bounded.

Clearly, there is no difference between employing **C1** and safeguarding when $\mathcal{B}$ is large enough and the sequences $\{\bar{\mu}^k\}_{k\in\mathbb{N}}$ and $\{\bar{\mu}^k\}_{k\in\mathbb{N}}$ converge to some Lagrange multipliers $\tilde{\lambda}$ and $\tilde{\mu}$, respectively. This holds, for instance, when $(\bar{x}, \tilde{\lambda}, \tilde{\mu})$ is a KKT point that satisfies the *second-order sufficient optimality condition* (SOSC) and the augmented Lagrangian scheme starts sufficiently close to $(x^*, \tilde{\lambda}, \tilde{\mu})$, with a penalty parameter ρ_1 large enough and ε_k appropriately controlled (see [27, Theorem 3.4]). On the other hand, without such assumptions, the behaviour of ALGENCAN with **C1** can be very different from ALGENCAN with safeguards, as it was shown in [36] by means of a simple example, with a unique minimizer that was also a KKT point. In their example, the safeguarded method was proven to generate sequences whose feasible limit points are exactly the minimizer of the problem, whilst the method that employs **C1** was proven to be unable to converge to it. Hence, in terms of reliability, using safeguarded multipliers is better than using **C1**, at least theoretically.

Continuing the discussion from [36], we are led to investigate the effects of removing safeguards from ALGENCAN, but without limiting the choices of $\bar{\lambda}^k$ and $\bar{\mu}^k$ to **C1**, that is, we consider $\mathcal{B} = \mathbb{R}^m \times \mathbb{R}^p_+$. First, note that a modest control over the parameters, such as

$$\frac{\left\| x^k - x^* \right\|_2}{\rho_k} \to 0 \tag{11}$$

is enough to establish a reasonable convergence theory for ALGENCAN with $\mathcal{B} = \mathbb{R}^m \times \mathbb{R}^p_+$, which goes in the exact same lines as the proofs of Theorems 1 and 2. In order to highlight this fact, we state it in a theorem environment as follows:

Theorem 5 *Let x^* be a local minimizer of (NLP). Then, x^* is reachable by* ALGENCAN *with $\mathcal{B} = \mathbb{R}^m \times \mathbb{R}^p_+$, regardless of the choices of the sequences $\{\bar{\lambda}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m$, and $\{\bar{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, as long as (11) holds.*

Note that (11) holds whenever $\{\bar{\lambda}^k\}_{k\in\mathbb{N}}$, $\{\bar{\mu}^k\}_{k\in\mathbb{N}}$ are bounded and $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$, or when $\|\bar{\lambda}^k, \bar{\mu}^k\|_2 = O(\rho_k^\beta)$, for every $\beta \in (0, 1)$. Thus, it allows a certain freedom in the growth of the approximate multipliers. However, the following example shows that it is not possible to relax (11) even further, for instance assuming $\left\|x^k - x^*\right\|_2 = O(\rho_k)$, without losing the property described in Theorem 5.

Example 5 In $\mathbb{R}^2$, consider the minimization problem:

$$\text{minimize } -x_2 \text{ subject to } x_1^2 x_2 \leqslant 0.$$

Note that $x^* \doteq (1, 0)$ is a local minimizer. Assume that there exists a sequence $x^k \doteq (x_1^k, x_2^k)$ with $x^k \to x^*$ and some approximate multiplier $\bar{\mu}^k$ with $\bar{\mu}^k/\rho_k \geqslant \alpha$, where $\alpha > 0$, for all k large enough, with $\rho_k \to \infty$, such that

$$\nabla L_{\rho,\bar{\mu}}(x) = \begin{bmatrix} 0 \\ -1 \end{bmatrix} + [\bar{\mu}^k + \rho_k (x_1^k)^2 (x_2^k)]_+ \begin{bmatrix} 2x_1^k x_2^k \\ (x_1^k)^2 \end{bmatrix} \to \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Thus, $[\bar{\mu}^k + \rho_k (x_1^k)^2 (x_2^k)]_+ \to 1$. Then, $[(\bar{\mu}^k/\rho_k) + (x_1^k)^2 (x_2^k)]_+ \to 0$ and this implies $\bar{\mu}^k/\rho_k \to 0$, which is a contradiction.

3.4 Independent penalty parameters

Here, we consider the possibility of different penalty parameters ρ's, one for each constraint. At first glance, this modification appears to make ALGENCAN more flexible, but it is of common sense that the use of different ρ's leads to sightly worse computational results (see [2]). In this section, we show in a formal way that using a common ρ is indeed better than using different ρ's. We refer to the variant of ALGENCAN with independent penalty parameters by the name SEP-ALGENCAN.

Algorithm 3 SEP-ALGENCAN.

Input: Some arrays of parameters $\tau \in (0, 1)^{m+p}$ and $\gamma \in (1, \infty)^{m+p}$, a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ with $\lim_{k\to\infty} \varepsilon_k = 0$, an initial array of penalty parameters $\rho^1 \in \mathbb{R}_{++}^{m+p}$, a compact set $\mathcal{B} \subseteq \mathbb{R}^m \times \mathbb{R}_+^p$, and some $(\bar{\lambda}^1, \bar{\mu}^1) \in \mathcal{B}$.
Initialize $k \leftarrow 1$. Then:
Step 1 *(Solving the subproblem):* Compute an approximate stationary point x^k of $L_{\rho^k,\bar{\lambda}^k,\bar{\mu}^k}(x)$, that is, a point that satisfies

$$\left\| \tilde{V}^{i_k - \frac{1}{2}} \right\|_\infty \leqslant \varepsilon_k.$$

Step 2 *(Updating the penalty parameter):* Calculate

$$V^k \doteq \left(h(x^k), \min\left\{ -g_1(x^k), \frac{\bar{\mu}_1^k}{\rho_{m+1}^k} \right\}, \ldots, \min\left\{ -g_p(x^k), \frac{\bar{\mu}_p^k}{\rho_{m+p}^k} \right\} \right). \tag{12}$$

Then, for each $\ell \in \{1, \ldots, m+p\}$,

a. If $k = 1$ or $|V_\ell^k| \leqslant \tau_\ell |V_\ell^{k-1}|$, set $\rho_\ell^{k+1} \doteq \rho_\ell^k$;
b. Otherwise, take some ρ_ℓ^{k+1} such that $\rho_\ell^{k+1} \geqslant \gamma_\ell \rho_\ell^k$.

Step 3 *(Estimating new projected multipliers):* Choose some $(\bar{\lambda}^{k+1}, \bar{\mu}^{k+1}) \in \mathcal{B}$, set $k \leftarrow k + 1$ and go to Step 1.

In order to make a proper analysis, we deal with the following version of AL-AKKT with separate ρ's:

Definition 2 We say that x^* is a *Sep-AL-AKKT* point if there are sequences $\{x^k\}_{k\in\mathbb{N}} \to x^*$, $\{\rho^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^{m+p}_{++}$, and bounded sequences $\{\bar{\lambda}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m$, $\{\bar{\mu}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p_+$, such that

$$\nabla_x L(x^k, \lambda^k, \mu^k) \to 0, \quad \text{and} \quad V^k \to 0$$

where $\lambda_i^k \doteq \bar{\lambda}_i^k + \rho_i^k h_i(x^k)$ for all $i = 1, \ldots, m$, $\mu_j^k \doteq [\bar{\mu}_j^k + \rho_{m+j}^k g_j(x^k)]_+$ for all $j = 1, \ldots, p$, and V^k is defined as in (12).

A procedure analogous to the proof of Theorem 1 is enough to conclude that every feasible limit point of an instance of SEP-ALGENCAN must be Sep-AL-AKKT, and that every Sep-AL-AKKT point is an accumulation point of an instance of SEP-ALGENCAN. Thus, the task of comparing ALGENCAN with SEP-ALGENCAN reduces to comparing Definitions 1 and 2.

Evidently, every AL-AKKT sequence is a Sep-AL-AKKT one, since the former is a particular case of the latter. However, the next example shows that the converse is not necessarily true.

Example 6 Let us consider the problem

$$\text{minimize } x_2 \text{ subject to } x_1^2 x_2 = 0, \ \ x_1 = 0,$$

adapted from [4, Example 3]. Using ρ_1 and ρ_2 for first and second constraints, respectively, we have

$$\nabla L_{\rho,\bar{\lambda}}(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + [\bar{\lambda}_1 + \rho_1 x_1^2 x_2] \begin{bmatrix} 2x_1x_2 \\ x_1^2 \end{bmatrix} + [\bar{\lambda}_2 + \rho_2 x_1] \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{13}$$

It is straightforward to verify that the feasible point $x^* \doteq (0, 0)$ is Sep-AL-AKKT with

$$\tilde{x}^k \doteq (1/k^2, -1/k^3), \quad \bar{\lambda}^k \doteq (0, 0), \quad \tilde{\rho}^k \doteq (k^{11}, k).$$

Now, we will show that x^* is not AL-AKKT. As the gradient of the first constraint vanishes at x^*, the bounded multiplier sequence $\{\bar{\lambda}_1^k\}_{k\in\mathbb{N}}$ does not matter in our analysis. Thus, we omit it. Suppose that $\{x^k\}_{k\in\mathbb{N}} \to x^*$ is an AL-AKKT sequence with the associated sequence $\{\rho^k\}_{k\in\mathbb{N}}$ and $\{\bar{\lambda}_2^k\}_{k\in\mathbb{N}}$. From (13) with $\rho_1^k = \rho_2^k \doteq \rho^k$, we must have

$$2\rho^k (x_1^k)^3 (x_2^k)^2 + \rho^k x_1^k + \bar{\lambda}_2^k \to 0 \ \text{ and } \ 1 + \rho^k (x_1^k)^4 x_2^k \to 0. \tag{14}$$

From the second expression of (14), we have $|\rho^k x_1^k| \to \infty$, and then using the first expression, $\rho^k x_1^k \, [2(x_1^k)^2 (x_2^k)^2 + 1] + \bar{\lambda}_2^k \to 0$, which leads to a contradiction.

That is, methods that employ a common ρ for all constraints are better, at least regarding the quality of the limit points. Also, note that the counterpart of CLASSICAL-EP with different penalty parameters is equivalent to SEP-ALGENCAN, which complements the discussion of Section 3.2. Indeed, if x^* satisfies

Sep-AL-AKKT with sequences $\{x^k\}_{k\in\mathbb{N}} \to x^*$, $\{\bar{\mu}^k\}_{k\in\mathbb{N}}$, $\{\bar{\lambda}^k\}_{k\in\mathbb{N}}$, and $\{\rho^k\}_{k\in\mathbb{N}}$, we can define new parameters

$$\bar{\rho}_i^k \doteq \begin{cases} \frac{\bar{\lambda}_i^k}{h_i(x^k)} + \rho_i^k & \text{if } h_i(x^k) \neq 0 \\ \rho_i^k & \text{otherwise} \end{cases}, \quad \bar{\rho}_{m+j}^k \doteq \begin{cases} \frac{\bar{\mu}_j^k}{[g_j(x^k)]_+} + \rho_{m+j}^k & \text{if } [g_j(x^k)]_+ \neq 0 \\ \rho_{m+j}^k & \text{otherwise} \end{cases},$$

for all $i \in \{1, \dots, m\}$ and all $j \in \{1, \dots, p\}$, so that x^* is also a limit point of CLASSICAL-EP with different ρ's. An immediate consequence of this fact is that the property of being Sep-AL-AKKT is invariant to the choice of $\{\bar{\lambda}^k\}_{k\in\mathbb{N}}$ and $\{\bar{\mu}^k\}_{k\in\mathbb{N}}$.

4 Tightest global convergence theory of the safeguarded augmented Lagrangian method to KKT points

One of the signature features of sequential optimality conditions is the presence of a companion *strict constraint qualification* (SCQ), which is basically the weakest CQ that guarantees equivalence between a given sequential optimality condition and the KKT conditions (see [16] for details). They are useful when compared with classical CQs, such as MFCQ [37] and Abadie's CQ [1], for acting like a measurement tool for the strength of the sequential optimality condition associated with them.

Since AL-AKKT can be viewed as a sequential optimality condition, it also has a companion SCQ, which is the main focus of this section. Let us recall the cones (9) and (10), related to the KKT conditions and CLASSICAL-EP, respectively. For a feasible x^* and $\rho > 0$, we have

$$K(x) = \{\nabla h(x)\lambda + \nabla g(x)\mu \mid \mu \geqslant 0,\ \mu_j = 0 \text{ if } g_j(x^*) < 0\}$$

and

$$K(x, \rho) = \{\nabla h(x)[\rho h(x)] + \nabla g(x)[\rho g(x)]_+\}.$$

From the proof of Theorem 4, we have $K(x^*) \subset \limsup_{x\to x^*,\, \rho\to\infty} K(x, \rho)$. The reverse inclusion is not always true (since we will show in Section 4.1 that this property strictly implies Abadie's CQ). This gap is exactly what characterizes the SCQ associated with AL-AKKT.

Definition 3 We say that a feasible point x^* satisfies the *AL-regularity* condition if

$$\limsup_{x\to x^*,\, \rho\to\infty} K(x, \rho) \subset K(x^*).$$

We use the name "AL-regularity" in the same fashion as [16]. The following result is a formal statement that AL-regularity is indeed the SCQ associated with AL-AKKT.

Theorem 6 *Every AL-AKKT point satisfying the AL-regularity condition is KKT. Conversely, if an AL-AKKT point x^* is also KKT, for every objective function f, then x^* satisfies the AL-regularity condition.*

Proof Let x^* be an AL-AKKT point with associated sequences $\{x^k\}_{k\in\mathbb{N}}$, $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{(\bar{\lambda}^k, \bar{\mu}^k)\}_{k\in\mathbb{N}}$. If $\{\rho_k\}_{k\in\mathbb{N}}$ is bounded, x^* is KKT independently of the presence of AL-regularity (see Remark 4). Thus, from now on suppose that $\rho_k \to \infty$. For simplicity, we denote

$$r(z, a, b) \doteq \nabla h(z)a + \nabla g(z)b.$$

Observe that

$$r(x^k, \bar{\lambda}^k + \rho_k h(x^k), [\bar{\mu}^k + \rho_k g(x^k)]_+) = r(x^k, \rho_k h(x^k), [\rho_k g(x^k)]_+) + r(x^k, \bar{\lambda}^k, \tilde{\mu}^k) \tag{15}$$

where $\tilde{\mu}^k := [\bar{\mu}^k + \rho_k g(x^k)]_+ - \rho_k g(x^k)_+ \geqslant 0$. Clearly, $r(x^k, \rho_k h(x^k), [\rho_k g(x^k)]_+) \in K(x^k, \rho_k)$ and, from the boundedness of $\{(\bar{\lambda}^k, \bar{\mu}^k)\}_{k\in\mathbb{N}}$, we also have, taking a subsequence if necessary, that $r(x^k, \bar{\lambda}^k, \tilde{\mu}^k)$ converges to some element in $K(x^*)$. Furthermore, observe that the left-hand side of (15) converges to $-\nabla f(x^k)$. Thus, by the validity of AL-regularity at x^*, we conclude that

$$-\nabla f(x^*) \in \left[\limsup_{x\to x^*,\, \rho\to\infty} K(x, \rho) \right] + K(x^*) \subset K(x^*) + K(x^*) = K(x^*),$$

that is, x^* is a KKT point.

Conversely, let $w^* \in \limsup_{x\to x^*,\, \rho\to\infty} K(x, \rho)$. Then, there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to x^*$, $\{w^k\}_{k\in\mathbb{N}} \to w^*$ and $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$ such that

$$w^k = r(x^k, \rho_k h(x^k), \rho_k g(x^k)_+) \in K(x^k, \rho_k)$$

for all k. Defining $f(x) = -(w^*)^T x$, the above expression gives

$$\nabla f(x^k) + \nabla h(x^k)[\rho_k h(x^k)] + \nabla g(x^k)[\rho_k g(x^k)_+] = -w^* + w^k \to 0.$$

Thus, we conclude that x^* is an AL-AKKT point with $(\bar{\lambda}^k, \bar{\mu}^k) = 0$ for all k. By hypothesis, it is also KKT. This implies $w^* = -\nabla f(x^*) \in K(x^*)$, and thus AL-regularity holds at x^*. □

An immediate use of AL-regularity is to give a partial answer to the question about the equivalence between CLASSICAL-EP and ALGENCAN. Recall that the sequences generated by CLASSICAL-EP are exactly AL-AKKT sequences with $(\bar{\lambda}^k, \bar{\mu}^k) = 0$, for all k, and $\rho_k \to \infty$. Thus, in view of Theorem 6, we conclude that AL-regularity is also the weakest CQ that guarantees convergence of CLASSICAL-EP to KKT points.

Corollary 2 *Every AL-AKKT point with $(\bar{\lambda}^k, \bar{\mu}^k) \doteq 0$ for all k and $\rho_k \to \infty$ satisfying the AL-regularity condition is KKT. Conversely, if an AL-AKKT point x^* with $(\bar{\lambda}^k, \bar{\mu}^k) \doteq 0$ for all k and $\rho_k \to \infty$ is also KKT, for every objective function f, then x^* satisfies the AL-regularity condition.*

Proof The first statement follows from Theorem 6. Conversely, the reader may notice that, in the proof of Theorem 6, we take $(\bar{\lambda}^k, \bar{\mu}^k) = 0$ for all k. □

The above discussion strongly suggests that the feasible accumulation points of sequences generated by ALGENCAN and those generated by CLASSICAL-EP are exactly the same; however, this should be investigated thoroughly in a future work.

We strongly emphasize that AL-regularity has a deeper meaning concerning the augmented Lagrangian because it describes the weakest CQ qualification required to prove convergence of ALGENCAN to KKT points. What follows is a formal statement of this fact.

Corollary 3 *Let x^* be a feasible accumulation point of a sequence generated by* ALGENCAN. *If x^* satisfies AL-regularity then it also satisfies the KKT conditions.*

Proof This is a direct consequence of Theorems 1 and 6. □

Thus, the best possible global convergence theory for ALGENCAN is the one built around Definition 1, Theorems 1 and 2, and Corollary 3. That means the only way of obtaining a stronger theory is imposing additional conditions over *Step 1* of the method. For example, requiring x^k to satisfy some approximate second-order necessary condition, in the sense of [8], is likely to lead to better results.

4.1 Relations of AL-regularity with other constraint qualifications

From the perspective of sequential optimality conditions, a natural question that arises is about the relation between other SCQs and AL-regularity outside the context of ALGENCAN. To the best of our knowledge, among the strongest sequential optimality conditions existent in the literature there are the *Positive AKKT* (PAKKT) condition and the *Approximate Gradient Projection* (AGP) condition, presented in [4] and [40], respectively. This means their associated SCQs are among the weakest possible. We are interested in both of them because they are independent of each other [4].

The AGP condition from [40] holds at a feasible point x^* when there exists some sequence $\{x^k\}_{k\in\mathbb{N}} \to x^*$ such that

$$\left\| \tilde{V}^{i_k - \frac{1}{2}} \right\|_{\infty} \to 0, \tag{16}$$

where

$$\Omega(x) \doteq \{z \in R^n \mid \nabla h(x)^T(z-x) = 0,\ \min\{0, g(x)\} + \nabla g(x)^T(z-x) \leqslant 0\},$$

which can be viewed as a linearization of the feasible set of (NLP) at $x \in \mathbb{R}^n$. Note that AGP does not explicitly require a Lagrange multiplier approximation to be verified. For this reason, it is suitable for proving convergence of numerical methods that do not provide multiplier approximations, such as *inexact restoration* methods [39]. Recently, however, it was proved that there is an equivalent version of AGP with Lagrange multipliers [3, Theorem 2.7].

According to [4, Definition 2.1], the PAKKT condition holds at a feasible point x^* of (NLP) when there are sequences $\{x^k\}_{k\in\mathbb{N}} \to x^*$, $\{\lambda^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^p$ and $\{\mu^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^m_+$ such that $\nabla_x L(x^k, \lambda^k, \mu^k) \to 0$, $\min\{-g(x^k), \mu^k\} \to 0$, and additionally,

$$\lim_{k\to\infty} \frac{|\lambda_i^k|}{\delta^k} > 0 \Rightarrow \lambda_i^k h_i(x^k) > 0, \quad \text{and} \quad \lim_{k\to\infty} \frac{\mu_i^k}{\delta^k} > 0 \Rightarrow \mu_i^k g_i(x^k) > 0,$$

for all k sufficiently large, where $\delta^k \doteq \|(1, \lambda^k, \mu^k)\|_\infty$, for every $k \in \mathbb{N}$. It was born as an improvement for the AKKT condition where the sign of the multipliers are controlled, even the ones associated with equality constraints. The difference with AKKT becomes very evident when considering *complementarity constraints* in (NLP). With regard to ALGENCAN, [4, Theorem 4.1] states that all of its feasible limit points must satisfy PAKKT with no additional assumption.

Now, since every AL-AKKT point is a feasible limit point of ALGENCAN (Theorem 1), which in turn satisfies both AGP and PAKKT, it follows that:

Theorem 7 *AL-AKKT implies both AGP and PAKKT.*

Following [4, Definition 2.3], the SCQ associated with PAKKT, namely *PAKKT-regularity*, consists in an upper semicontinuity-like of the mapping

$$K^P(x, \alpha, \beta) \doteq \left\{ r(x, \lambda, \mu) \in K(x) \middle| \begin{array}{l} \lambda_i h_i(x) \geqslant \alpha \text{ if } |\lambda_i| > \beta \left\|x^k - x^*\right\|_2 \\ \mu_j g_j(x) \geqslant \alpha \text{ if } \mu_j > \beta \left\|x^k - x^*\right\|_2 \end{array} \right\},$$

that is, PAKKT-regularity holds at x^* when

$$\limsup_{x \to x^*,\, \alpha \downarrow 0,\, \beta \downarrow 0} K^P(x, \alpha, \beta) \subset K(x^*),$$

which is similar to Definition 3 in some sense. Furthermore, the SCQ associated with the AGP condition from [16, Definition 1], namely *AGP-regularity*, is characterized by the upper semicontinuity of the normal cone of $\Omega(x)$. That is, we say that a feasible point x^* is AGP-regular when

$$\limsup_{x \to x^*,\, \varepsilon \to 0} N_{\Omega(x)}(x + \varepsilon) \subset N_{\Omega(x^*)}(x^*).$$

In order to compare AL-regularity with the strict constraint qualifications associated with PAKKT and AGP, it suffices to keep in mind the following characterization, presented in [4, Theorem 2.4] and [16, Theorem 1]:

Theorem 8 *A feasible point x^* of (NLP) satisfies PAKKT-regularity (respectively, AGP-regularity) if, and only if, for every continuously differentiable objective function, the fact of x^* being PAKKT (respectively, AGP) implies that x^* is KKT.*

Hence, their relation with AL-regularity follows from Theorem 7:

Corollary 4 *Both AGP-regularity and PAKKT-regularity imply AL-regularity.*

Proof Let x^* be a feasible point of (NLP) satisfying AGP-regularity. Then, for every continuously differentiable objective function we have that: if x^* is AGP, then it is also KKT. Hence, we also have that for every continuously differentiable objective function, if x is AL-AKKT, it is also AGP (due to Theorem 7) and, consequently, KKT. Now, it follows from Theorem 6 that x^* satisfies AL-regularity. Thus, AGP-regularity implies AL-regularity. Analogously, it is possible to prove that PAKKT-regularity also implies AL-regularity. □

It is important to note that the implications given by Corollary 4 are strict, since PAKKT and AGP are independent [4].

Finally, we present the relation between Abadie's CQ and AL-regularity. Let us denote the tangent cone to the feasible set of (NLP) at x^* by $\mathcal{T}(x^*)$, and its linearized cone by $\mathcal{L}(x^*)$. We say that Abadie's CQ holds at x^* if $\mathcal{T}(x^*) = \mathcal{L}(x^*)$. See [1].

Theorem 9 *AL-regularity implies Abadie's CQ.*

Proof The proof is analogous to that of [15, Theorem 4.4], since for the sequences $\{x^{k,l(k)}\}_{k\in\mathbb{N}}$ and $\{w^{k,l(k)}\}_{k\in\mathbb{N}}$ considered in that proof, $w^{k,l(k)} \in K(x^{k,l(k)}, l(k))$ holds true for all k, and $\lim_{k\to\infty} l(k) = \infty$. □

The implication in Theorem 9 is strict, as the next example shows.

Example 7 (Abadie's CQ does not imply AL-regularity) As in [19, Example 7.3], let us consider the constraints in $\mathbb{R}^2$

$$x_1^6 + x_2^3 \leqslant 0, \quad x_2 \leqslant 0$$

and the feasible point $x^* \doteq (0, 0)$. It is straightforward to verify that $\mathcal{T}(x^*) = \mathcal{L}(x^*) = \mathbb{R} \times \mathbb{R}_-$, and thus x^* conforms to Abadie's CQ. On the other hand, AL-regularity does not hold at x^* since $K(x^*) = \{0\} \times \mathbb{R}_+$ and, defining the sequences $\rho_k \doteq k^{11}$ and $x^k \doteq (1/k, 0)$ for all $k \in \mathbb{N}$, we see that $K(x^*) \not\ni (6, 0) \in \limsup_{x\to x^*,\, \rho\to\infty} K(x, \rho)$.

We summarize the relation among several known CQs from the literature and AL-regularity in Fig. 1.

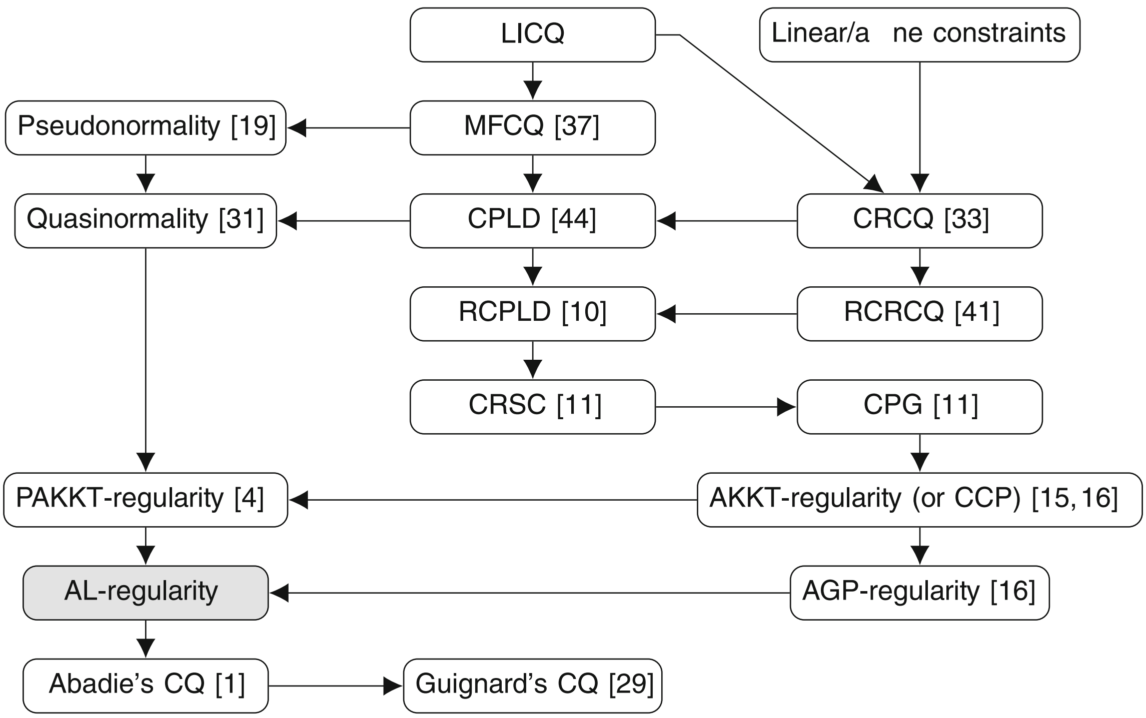


Fig. 1 Updated landscape of constraint qualifications for standard nonlinear programming. AL-regularity and every more stringent CQ are associated with the global convergence of ALGENCAN

5 Conclusions

Over the last years, the convergence of ALGENCAN has been enhanced by means of the so-called sequential optimality conditions (see [3, 4, 16, 17]), and the best results so far are related to the independent sequential conditions PAKKT [4] and AGP [6]. As a rule, sequential optimality conditions must be indeed necessary for optimality, but not only that: they must imply the KKT conditions under some mild constraint qualifications, and there must be at least one relevant algorithm whose feasible limit points satisfy them, in order to illustrate its usability. In this paper, we defined a new sequential optimality condition, called AL-AKKT, which not only satisfies the three properties we just mentioned (Theorems 2, 6, and 1, respectively), but also completely characterizes all feasible limit points of ALGENCAN. Consequently, its associated strict constraint qualification characterizes the weakest possible constraint qualification under which ALGENCAN is guaranteed to converge to KKT points. In particular, since AL-AKKT strictly implies both conditions PAKKT and AGP (Theorem 7), Theorem 6 improves the global convergence result from [4, Theorem 4.1], as well as guarantees that ALGENCAN converges to AGP points.

From a practical point of view, we recall that there are many distinct variants of the augmented Lagrangian method, with potentially different performances depending on the problem they are applied to. In a real world application, one may find suitable to use different implementations of the method to solve the same problem and then select the best solution among their outputs, afterwards. But should this be impossible or inconvenient, the results of this paper can be taken into consideration for deciding, beforehand, which implementation might be the best for a general problem, regarding the quality of the solutions that it may return. Based on our findings, we believe that the most reasonable implementation of the augmented Lagrangian, that balances theory and practice, is characterized by:

- The use of projected (bounded, safeguarded) Lagrange multipliers in the underlying problems of the method, since this leads to solutions with the same quality as the pure external penalty method (see Corollary 1 and the discussion in Section 3.3);
- The use of a penalty parameter growth control, since it has a positive effect over the numerical stability of the method (see, for instance, [2]) without any drawback on its convergence theory (Theorem 3);
- The use of a single penalty parameter ρ for all constraints, as suggested in Section 3.4.

Besides, it is reasonable that the set $\mathcal{B}$ of projected multipliers be large, for this increases the likelihood of $\bar{\lambda}$ and $\bar{\mu}$ to converge to actual Lagrange multipliers (when they exist), hence avoiding unnecessary increments of ρ. See also [2, Section 5] and the book [22]. As a matter of fact, the most similar variant to what we just described is the ALGENCAN implementation.

As for the optimality condition AL-AKKT, we remark that it contrasts with other sequential conditions, which are meant for unifying convergence theories of different algorithms, since it is intrinsic to AL strategies. Nevertheless, we believe our approach, which consists of characterizing a specialized sequential condition from an algorithm, may be useful for building convergence theories of other algorithms, for comparing some of their variants, and ultimately, for comparing different algorithms, from a theoretical point of view. For instance, in [14], the authors prove that the Newton–Lagrange method may not generate even AKKT sequences, which is the weakest form of sequential optimality condition in the literature. Moreover, some promising results were presented recently in [9] assessing the use of a stopping criterion for augmented Lagrangian methods based on a scaled KKT residual. This indicates that a convergence analysis of these algorithms based on an analogue of AL-AKKT should bring to light some interesting aspects of them. We also expect other algorithms to be analysed with the ideas introduced in this paper, together with extensions to other types of optimization frameworks [5, 6, 12, 13, 23, 24, 26, 28, 34, 45].

Acknowledgements We would like to thank the reviewers of this paper for their many insightful suggestions.

Funding This work has been partially supported by CEPID-CeMEAI (FAPESP 2013/07375-0), FAPES (grant 116/2019), FAPESP (grants 2018/24293-0, 2017/18308-2 and 2017/17840-2), CNPq (grants 301888/2017-5, 303427/2018-3, 404656/2018-8, 438185/2018-8 and 307270/2019-0), and PRONEX - CNPq/FAPERJ (grant E-26/010.001247/2016).

References

1. Abadie, J.: On the Kuhn-Tucker Theorem. In: Abadie, J. (ed.) Programming, Nonlinear, pp. 21–36. Wiley, New York (1967)
2. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: On augmented Lagrangian methods with general lower-level constraints. SIAM J. Optim. **18**(4), 1286–1309 (2007). https://doi.org/10.1137/060654797
3. Andreani, R., Cárdenas, A.R.V., Ramos, A., Ribeiro, A., Secchin, L.D.: On the convergence of augmented Lagrangian strategies for nonlinear programming. To appear in IMA Journal of Numerical Analysis (2020)
4. Andreani, R., Fazzio, N., Schuverdt, M., Secchin, L.: A sequential optimality condition related to the quasi-normality constraint qualification and its algorithmic consequences. SIAM J. Optim. **29**(1), 743–766 (2019). https://doi.org/10.1137/17M1147330
5. 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. Tech. rep. http://www.optimization-online.org/DB_HTML/2019/10/7436.html (2019)
6. Andreani, R., Gómez, W., Haeser, G., Mito, L.M., Ramos, A.: On optimality conditions for nonlinear conic programming. Tech. rep. http://www.optimization-online.org/DB_HTML/2020/03/7660.html (2020)
7. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**(5), 627–641 (2011). https://doi.org/10.1080/02331930903578700
8. Andreani, R., Haeser, G., Ramos, A., Silva, P.J.S.: A second-order sequential optimality condition associated to the convergence of optimization algorithms. IMA J. Numer. Anal. **37**(4), 1902–1929 (2017). https://doi.org/10.1093/imanum/drw064
9. 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. To appear in Mathematical Programming Computation. https://doi.org/10.1007/s12532-021-00207-9 (2021)
10. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: A relaxed constant positive linear dependence constraint qualification and applications. Math. Program. **135**(1), 255–273 (2012). https://doi.org/10.1007/s10107-011-0456-0

11. 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). https://doi.org/10.1137/110843939
12. Andreani, R., Haeser, G., Secchin, L.D., Silva, P.J.S.: New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences. SIAM J. Optim. **29**(4), 3201–3230 (2019). https://doi.org/10.1137/18M121040X
13. Andreani, R., Haeser, G., Viana, D.S.: Optimality conditions and global convergence for nonlinear semidefinite programming. Mathematical Programming. https://doi.org/10.1007/s10107-018-1354-5 (2018)
14. Andreani, R., Martínez, J.M., Santos, L.T.: Newton's method may fail to recognize proximity to optimal points in constrained optimization. Math. Program. **160**(1), 547–555 (2016). https://doi.org/10.1007/s10107-016-0994-6
15. Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: A cone-continuity constraint qualification and algorithmic consequences. SIAM J. Optim. **26**(1), 96–110 (2016). https://doi.org/10.1137/15M1008488
16. Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: Strict constraint qualifications and sequential optimality conditions for constrained optimization. Math. Oper. Res. **43**(3), 693–717 (2018). https://doi.org/10.1287/moor.2017.0879
17. Andreani, R., Martínez, J.M., Svaiter, B.F.: A new sequential optimality condition for constrained optimization and algorithmic consequences. SIAM J. Optim. **20**(6), 3533–3554 (2010). https://doi.org/10.1137/090777189
18. Arora, J.S., Chahande, A.I., Paeng, J.K.: Multiplier methods for engineering optimization. Int. J. Numer. Methods Eng. **32**(7), 1485–1525 (1991). https://doi.org/10.1002/nme.1620320706
19. Bertsekas, D.P., Ozdaglar, A.E.: Pseudonormality and a Lagrange multiplier theory for constrained optimization. Journal of Optimization Theory and Applications **114**(2), 287–343 (2002). https://doi.org/10.1023/A:1016083601322
20. Birgin, E.G., Castillo, R.A., MartÍnez, J.M.: Numerical comparison of augmented Lagrangian algorithms for nonconvex problems. Comput. Optim. Appl. **31**(1), 31–55 (2005). https://doi.org/10.1007/s10589-005-1066-7
21. Birgin, E.G., Martínez, J.M.: Augmented Lagrangian method with nonmonotone penalty parameters for constrained optimization. Comput. Optim. Appl. **51**(3), 941–965 (2012). https://doi.org/10.1007/s10589-011-9396-0
22. Birgin, E.G., Martínez, J.M.: Practical augmented lagrangian methods for constrained optimization. Society for Industrial and Applied Mathematics, Philadelphia, PA. https://doi.org/10.1137/1.9781611973365 (2014)
23. Bueno, L., Haeser, G., Lara, F., Rojas, F.: An augmented Lagrangian method for quasi-equilibrium problems. Comput. Optim. Appl. **76**, 737–766 (2020). https://doi.org/10.1007/s10589-020-00180-4
24. Bueno, L., Haeser, G., Rojas, F.: Optimality conditions and constraint qualifications for generalized Nash equilibrium problems and their practical implications. SIAM J. Optim. **29**(1), 31–54 (2019). https://doi.org/10.1137/17M1162524
25. Curtis, F.E., Jiang, H., Robinson, D.P.: An adaptive augmented Lagrangian method for large-scale constrained optimization. Math. Program. **152**, 201–245 (2015). https://doi.org/10.1007/s10107-014-0784-y
26. Feng, M., Li, S.: An approximate strong KKT condition for multiobjective optimization. TOP **26**(3), 489–509 (2018). https://doi.org/10.1007/s11750-018-0491-6
27. Fernandez, D., Solodov, M.V.: Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition. SIAM J. Optim. **22**(2), 384–407 (2012). https://doi.org/10.1137/10081085X
28. Gill, P.E., Kungurtsev, V., Robinson, D.P.: A shifted primal-dual penalty-barrier method for nonlinear optimization. SIAM J. Optim. **30**(2), 1067–1093 (2020). https://doi.org/10.1137/19M1247425
29. Guignard, M.: Generalized kuhn-Tucker conditions for mathematical programming problems in a Banach space. SIAM Journal on Control **7**(2), 232–241 (1969). https://doi.org/10.1137/0307016
30. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**(5), 303–320 (1969). https://doi.org/10.1007/BF00927673
31. Hestenes, M.R.: Optimization Theory: the Finite Dimensional Case. Wiley, New York (1975)
32. Izmailov, A.F., Solodov, M.V., Uskov, E.I.: Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. SIAM J. Optim. **22**(4), 1579–1606 (2012). https://doi.org/10.1137/120868359

33. Janin, R.: Directional derivative of the marginal function in nonlinear programming. In: Fiacco, A.V. (ed.) Sensitivity, Stability and Parametric Analysis, Mathematical Programming Studies, vol. 21, pp. 110–126. Springer, Berlin (1984). https://doi.org/10.1007/BFb0121214
34. Kanzow, C., Raharja, A.B., Schwartz, A.: Sequential optimality conditions for cardinality-constrained optimization problems with applications. Tech. rep. https://www.mathematik.uni-wuerzburg.de/fileadmin/10040700/paper/CC-AM-Stat-feb-12.pdf (2021)
35. Kanzow, C., Schwartz, A.: The price of inexactness: convergence properties of relaxation methods for mathematical programs with complementarity constraints revisited. Math. Oper. Res. **40**(2), 253–275 (2015). https://doi.org/10.1287/moor.2014.0667
36. Kanzow, C., Steck, D.: An example comparing the standard and safeguarded augmented Lagrangian methods. Oper. Res. Lett. **45**(6), 598–603 (2017). https://doi.org/10.1016/j.orl.2017.09.005
37. Mangasarian, O.L., Fromovitz, S.: The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. J. Math. Anal. Appl. **17**(1), 37–47 (1967). https://doi.org/10.1016/0022-247X(67)90163-1
38. Martínez, J.M.: Otimização prática usando o lagrangiano aumentado (in portuguese). Tech. rep. https://www.ime.unicamp.br/martinez/opusculo.html (2006)
39. Martínez, J.M., Pilotta, E.A.: Inexact-restoration algorithm for constrained optimization. Journal of Optimization Theory and Applications **104**(1), 135–163 (2000). https://doi.org/10.1023/A:1004632923654
40. Martínez, J.M., Svaiter, B.F.: A practical optimality condition without constraint qualifications for nonlinear programming. Journal of Optimization Theory and Applications **118**(1), 117–133 (2003). https://doi.org/10.1023/A:1024791525441
41. Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optimization **60**(4), 429–440 (2011). https://doi.org/10.1080/02331930902971377
42. Nocedal, J., Wright, S.J. Numerical Optimization, 2nd edn. Springer, Berlin (2006)
43. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) Optimization, pp. 283–298. Academic Press, New York (1969)
44. Qi, L., Wei, Z.: On the constant positive linear dependence condition and its application to SQP methods. SIAM J. Optim. **10**(4), 963–981 (2000). https://doi.org/10.1137/S1052623497326629
45. Ramos, A.: Mathematical programs with equilibrium constraints: a sequential optimality condition, new constraint qualifications and algorithmic consequences. Optimization Methods and Software, pp. 1–37. https://doi.org/10.1080/10556788.2019.1702661 (2019)
46. Rockafellar, R.T.: The multiplier method of Hestenes and Powell applied to convex programming. Journal on Optimization Theory and Applications **12**, 555–562 (1973). https://doi.org/10.1007/BF00934777
47. Rockafellar, R.T.: Augmented Lagrange multiplier functions and duality in nonconvex programming. SIAM Journal on Control **12**(2), 268–285 (1974). https://doi.org/10.1137/0312021

Numerical Algorithms
https://doi.org/10.1007/s11075-021-01241-3

CORRECTION

Correction to: On the best achievable quality of limit points of augmented Lagrangian schemes

Roberto Andreani[1] · Gabriel Haeser[2] · Leonardo M. Mito[2] · Alberto Ramos[3] · Leonardo D. Secchin[4]

Published online: 20 December 2021

The published article "R. Andreani, G. Haeser, L. M. Mito, A. Ramos, L. D. Secchin. On the best achievable quality of limit points of augmented Lagrangian schemes. *Numer. Algor.*, 2021. https://doi.org/10.1007/s11075-021-01212-8" has the errors listed below. The authors apologize for that.

- Step 1, Algorithms 2 and 3: $\|\tilde{V}^{i_k - \frac{1}{2}}\|_\infty \leqslant \varepsilon_k$ should be $\|\nabla L_{\rho_k, \bar{\lambda}^k, \bar{\mu}^k}(x^k)\|_\infty \leqslant \varepsilon_k$
- Expression (11): $\frac{\|x^k - x^*\|_2}{\rho_k} \to 0$ should be $\frac{\|(\bar{\lambda}^k, \bar{\mu}^k)\|_2}{\rho_k} \to 0$

The original article can be found online at https://doi.org/10.1007/s11075-021-01212-8.

✉ Roberto Andreani
andreani@unicamp.br

Gabriel Haeser
ghaeser@ime.usp.br

Leonardo M. Mito
leokoto@ime.usp.br

Alberto Ramos
albertoramos@ufpr.br

Leonardo D. Secchin
leonardo.secchin@ufes.br

1 Department of Applied Mathematics, University of Campinas, Rua Sérgio Buarque de Holanda, 651, 13083-859, Campinas, SP, Brazil

2 Department of Applied Mathematics, University of São Paulo, Rua do Matão, 1010, Cidade Universitária, 05508-090, São Paulo, SP, Brazil

3 Department of Mathematics, Federal University of Paraná, 81531-980, Curitiba, PR, Brazil

4 Department of Applied Mathematics, Federal University of Espírito Santo, Rodovia BR 101, Km 60, 29932-540, São Mateus, ES, Brazil

- Paragraph after Theorem 5: $\|\bar{\lambda}^k, \bar{\mu}^k\|_2 = O(\rho_k^\beta)$, should be $\|(\bar{\lambda}^k, \bar{\mu}^k)\|_2 = O(\rho_k^\beta)$ and $\|x^k - x^*\|_2 = O(\rho_k)$ should be $\|(\bar{\lambda}^k, \bar{\mu}^k)\|_2 = O(\rho_k)$
- Expression (16): $\|\tilde{V}^{i_k - \frac{1}{2}}\|_\infty \to 0$ should be $\|\text{proj}_{\Omega(x^k)}(x^k - \nabla f(x^k)) - x^k\|_\infty \to 0$
- Definition of $K^P(x, \alpha, \beta)$, after Theorem 7: $\|x^k - x^*\|_2$ should be $\|(1, \lambda, \mu)\|_\infty$
- Item 2. after Definition 1: $\{\lambda^k, \mu^k\}_{k\in\mathbb{N}}$ should be $\{(\lambda^k, \mu^k)\}_{k\in\mathbb{N}}$
- Example 1 and step 2 of Algorithm 3: $\bar{\mu_1^k}$, $\bar{\mu_2^k}$ and $\bar{\mu_p^k}$ should be $\bar{\mu}_1^k$, $\bar{\mu}_2^k$ and $\bar{\mu}_p^k$, respectively

Appendix C

External reference III

Title: First- and second-order optimality conditions for second-order cone and semidefinite programming under a constant rank condition.

Status: Under review.

What is attached: A preprint that has *not* been peer reviewed yet.

First- and second-order optimality conditions for second-order cone and semidefinite programming under a constant rank condition

R. Andreani* G. Haeser† L. M. Mito† H. Ramírez C.‡ T. P. Silveira†

February 3, 2022

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.

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 5, or [53]),

*Department of Applied Mathematics, University of Campinas, Campinas-SP, Brazil. Email: andreani@ime.unicamp.br
†Department of Applied Mathematics, University of São Paulo, São Paulo-SP, Brazil. Email: {ghaeser,leokoto,thiagops}@ime.usp.br
‡Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático (CNRS IRL 2807), Universidad de Chile, Santiago, Chile. Email: hramirez@dim.uchile.cl

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 5), 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: Section 2 consists of a nonlinear conic programming review emphasizing some aspects of the theory that are not commonly discussed in the literature; in Section 3, we analyze CRCQ for NLP and we show how it can be interpreted in terms of the faces of the nonnegative orthant. In Sections 4 and 5, we propose extensions of CRCQ for NSOCP and NSDP, respectively, and we prove some of its properties. Finally, in Section 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

$$S^\circ := \{z \in \mathbb{E} \mid \langle z, y \rangle \leq 0, \ \forall y \in S\}$$

and the *orthogonal complement* of S will be denoted by $S^\perp$. The notations $\text{cl}(S)$, $\text{int}(S)$, $\text{bd}(S)$, and $\text{bd}^+(S)$ 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 $\text{cone}(S)$, and the *smallest linear space* that contains S will be denoted by $\text{span}(S)$. 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 $D^2g(x)$ 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 $D^2g(x)$ over $d_1, d_2 \in \mathbb{R}^n$ will be denoted by $D^2g(x)[d_1, d_2]$.

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

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

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 $\Xi \colon \mathcal{N} \to \mathbb{F}$ (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*

$$\mathcal{L}_\Omega(\bar{x}) := \{d \in \mathbb{R}^n \mid Dg(\bar{x})d \in \mathcal{T}_\mathcal{K}(g(\bar{x}))\},$$

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 2.1. *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

$$\mathcal{N}_\mathcal{K}(g(\bar{x})) = \{ z \in \mathcal{K}^\circ \mid \langle g(\bar{x}), z \rangle = 0 \}.$$

Then, combining the first-order geometric necessary condition and Lemma 2.1 yields the following theorem, also by Guignard:

Theorem 2.1 (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 2.1 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 2.1,

$$\mathcal{T}_\Omega(\bar{x})^\circ = \mathcal{L}_\Omega(\bar{x})^\circ \quad \text{and} \quad H(\bar{x}) \text{ is closed}, \tag{4}$$

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 2.1. 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 2.1. *Consider the following problem, presented in [2, Subsection 2.1]:*

$$\begin{array}{ll} \text{Minimize} & f(x) := -x_2, \\ s.t. & g(x) := (x_1, x_1, x_2) \in K_3, \end{array}$$

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

$$\mathcal{T}_\Omega(\bar{x}) = \mathcal{L}_\Omega(\bar{x}) \quad \text{and} \quad H(\bar{x}) \text{ is closed}, \tag{6}$$

which implies Guignard's CQ, is known as *Abadie's CQ* (see also Börgens et al. [25, Definition 5.5]), and Example 2.1 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

$$g(\bar{x}) + Dg(\bar{x})d \in \operatorname{int}(\mathcal{K}).$$

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

$$\operatorname{Im}(Dg(\bar{x})) + \operatorname{lin}(\mathcal{T}_\mathcal{K}(g(\bar{x}))) = \mathbb{E}, \tag{8}$$

where $\operatorname{lin}(\mathcal{T}_\mathcal{K}(g(\bar{x}))) = \mathcal{T}_\mathcal{K}(g(\bar{x})) \cap -\mathcal{T}_\mathcal{K}(g(\bar{x}))$ 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 $\operatorname{lin}(\mathcal{T}_\mathcal{K}(g(\bar{x}))) = \operatorname{Ker}(D\Xi(g(\bar{x})))$; 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

$$C(\bar{x}) := \{d \in \mathbb{R}^n \mid d \in \mathcal{T}_\Omega(\bar{x}),\ \langle \nabla f(\bar{x}), d\rangle = 0\},$$

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 2.2 (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 nonnegative; 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 2.2 can be derived without constraint qualifications, using Fritz John (generalized) multipliers [24, Theorem 3.50].

Although the condition of Theorem 2.2 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 2.1. *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 2.2 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 2.2.

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.

[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].

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

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

which is a particular case of (NCP) with $\mathbb{E} = \mathbb{R}^{m+p}$, $\mathcal{K} = \mathbb{R}^m_+ \times \{0\}^p$, and $g(x) := (g_1(x), \ldots, g_{m+p}(x))$. 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 $\mathcal{A}(\bar{x}) := \{j \in \{1, \ldots, m\} \mid g_j(\bar{x}) = 0\}$.

Now, let us recall Janin's constant rank constraint qualification as it was first presented in [36].

Definition 3.1 (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 3.1; 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 3.2 (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 $\Xi \colon \mathcal{N} \to \mathbb{R}^{s+p}$ 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.

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.

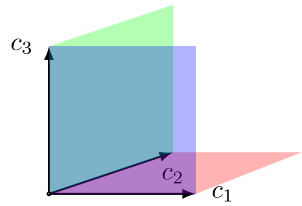


Figure 1: Faces of $\mathbb{R}^3_+$

For example, in Figure 1, the vertex of $\mathbb{R}^3_+$ corresponds to $J = \{1, 2, 3\}$; the one-dimensional faces $\mathrm{cone}(c_1)$, $\mathrm{cone}(c_2)$, and $\mathrm{cone}(c_3)$ 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 3.1. *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 $\mathcal{A}(\bar{x}) = \{1, \dots, s\}$. Moreover, let $c_1, \dots, c_{s+p}$ 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 $J \subseteq \{1, \dots, s\}$ such that

$$F = \left(\mathbb{R}^s_+ \bigcap_{j \in J} \{c_i\}^\perp \right) \times \{0\}^p,$$

which implies

$$F^\perp = R^\perp \times \mathbb{R}^p = \mathrm{span}\left(\{c_j \mid j \in J \cup \{s+1, \dots, s+p\}\} \right),$$

so

$$\begin{aligned} D\mathcal{G}(x)^T[F^\perp] &= \mathrm{span}(\{D\mathcal{G}(x)^T c_j\}_{j \in J \cup \{s+1, \dots, s+p\}}) \\ &= \mathrm{span}(\{\nabla g_j(x)\}_{j \in J \cup \{m+1, \dots, m+p\}}). \end{aligned} \tag{16}$$

Consequently,

$$\dim(Dg(x)^T[F^\perp]) = \mathrm{rank}(\{\nabla g_j(x)\}_{j \in J \cup \{m+1, \dots, m+p\}}).$$

The conclusion follows from the one-to-one correspondence between F and J.

□

The equivalent form of RCRCQ presented in Proposition 3.1 allows us to visualize what it actually describes, geometrically. Indeed, recall that $\mathbb{R}^n = D\mathcal{G}(x)^{-1}(\mathrm{span}(F)) + (D\mathcal{G}(x)^{-1}(\mathrm{span}(F)))^\perp$ and it is elementary to see that

$$(D\mathcal{G}(x)^{-1}(\mathrm{span}(F)))^\perp = D\mathcal{G}(x)^T[F^\perp].$$

This implies the following relation:

$$\dim(D\mathcal{G}(x)^{-1}(\mathrm{span}(F))) + \dim(D\mathcal{G}(x)^T[F^\perp]) = n.$$

Thus, RCRCQ can be equivalently stated as the constant dimension of $D\mathcal{G}(x)^{-1}(\mathrm{span}(F))$ for every $x \in \mathcal{V}$ at each $F \trianglelefteq \mathcal{C} = \mathbb{R}^{|\mathcal{A}(\bar{x})|}_+ \times \{0\}^p$. The set $D\mathcal{G}(x)^{-1}(\mathrm{span}(F))$, 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 $\text{Ker}(Dg_J(x))$ 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 3.1 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 3.1. *It is possible to obtain a characterization of CRCQ in the same style of Proposition 3.1. 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 3.1.*

In view of Remark 3.1, 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 4.2 and 5.3. 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:

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

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 4.2 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 $\Xi_j : \mathcal{N}_j \to \mathbb{R}^{m_j}$ is the identity function for every $j \in I_0(\bar{x})$, and $\Xi_j : \mathcal{N}_j \to \mathbb{R}$ 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 $\bar{\lambda}_j \in K^\circ_{m_j}$, $j \in I_0(\bar{x}) \cup I_B(\bar{x})$, 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 3.1.

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 4.1. *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 3.1 and note that Definition 4.1 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 4.2, 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 4.1. *([4, Proposition 3.1]) Let $\{\zeta_i(x)\}_{i\in\mathcal{I}}$ be a finite family of twice continuously differentiable functions $\zeta_i\colon \mathbb{R}^n \to \mathbb{R}$, $i \in \mathcal{I}$, 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*

$$\mathcal{S} := \{y \in \mathbb{R}^n \mid \langle \nabla\zeta_i(\bar{x}), y\rangle = 0,\ i \in \mathcal{I}\}.$$

Then, there exist 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 4.1. *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

$$F_j := \begin{cases} K_{m_j} & \text{if } Dg_j(\bar{x})d \in \text{int}(K_{m_j}),\\ \text{cone}(Dg_j(\bar{x})d), & \text{if } Dg_j(\bar{x})d \in \text{bd}^+(K_{m_j}),\\ \{0\}, & \text{if } Dg_j(\bar{x})d = 0.\end{cases} \tag{26}$$

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 $Dg_j(\bar{x})d \in \text{bd}^+(K_{m_j})$, let $A_j \in \mathbb{R}^{m_j\times m_j-1}$ be any matrix with full column rank such that $\text{Im}(A_j) = \{Dg_j(\bar{x})d\}^\perp$, and observe that

$$Dg_j(x)^T[F_j^\perp] = \text{span}\left(\left\{Dg_j(x)^T A_j^i\right\}_{i=1,\dots,m_j-1}\right)$$

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

$$Dg_j(x)^T[F_j^\perp] = \text{span}(\{\nabla g_{j,i}(x)\}_{i=0,\dots,m_j-1}),$$

where $\nabla g_{j,i}(x)$ denotes the i-th column of $Dg_j(x)^T$. And for every j such that $Dg_j(\bar{x})d \in \text{int}(K_{m_j})$, we have $Dg_j(x)^T[F_j^\perp] = \{0\}$. Finally, observe that $R^\perp = \text{span}(\{c_j\}_{j\in J})$ and then

$$D\phi(x)^T[R^\perp] = \text{span}\left(\{\nabla\phi_j(x)\}_{j\in J}\right).$$

Therefore, for every $x \in \mathcal{V}$, where $\mathcal{V}$ is the neighborhood of $\bar{x}$ given by Definition 4.1, 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:

$$\bigcup_{\substack{j\in I_0(\bar{x}) \\ Dg_j(\bar{x})d\in \mathrm{bd}^+(K_{m_j}) \\ i=1,\dots,m_j-1}} \left\{Dg_j(x)^T A_j^i\right\} \bigcup_{\substack{j\in I_0(\bar{x}) \\ Dg_j(\bar{x})d=0 \\ i=0,\dots,m_j-1}} \{\nabla g_{j,i}(x)\} \bigcup_{j\in J}\{\nabla\phi_j(x)\}, \tag{29}$$

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 4.1 to the family of functions

$$\zeta_{i,j}(x) := \begin{cases} \langle A_j^i, g_j(x)\rangle, & \text{if } j \in I_0(\bar{x}),\ Dg_j(\bar{x})d \in \mathrm{bd}^+(K_{m_j}),\ i = 1,\dots,m_j-1, \\ g_{j,i}(x), & \text{if } j \in I_0(\bar{x}),\ Dg_j(\bar{x})d = 0,\ i = 0,\dots,m_j-1, \\ \phi_j(x), & \text{if } j \in J, \end{cases} \tag{30}$$

where $g_{j,i}(x)$ denotes the i-th entry of $g_j(x)$ for $j \in J$. Then, consider the following subspace:

$$\mathcal{S} := \left\{ y \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} A_j^T Dg_j(\bar{x})y = 0, & \text{if } j \in I_0(\bar{x}),\ Dg_j(\bar{x})d \in \mathrm{bd}^+(K_{m_j}) \\ Dg_j(\bar{x})y = 0, & \text{if } j \in I_0(\bar{x}),\ Dg_j(\bar{x})d = 0 \\ \nabla\phi_j(\bar{x})^T y = 0, & \text{if } j \in J\ , \end{array} \right\},$$

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 4.1 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 4.1. 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 4.1 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

$$\mathcal{G}(\xi(t)) \in \mathrm{span}(F)$$

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 $Dg_j(\bar{x})d \in \mathrm{int}(K_{m_j})$, 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 $Dg_j(\bar{x})d \in \mathrm{bd}^+(K_{m_j})$, then $g_j(\xi(t)) \in \mathrm{span}(Dg_j(\bar{x})d)$ due to (26), and it follows from (31) that $g_j(\xi(t)) \in \mathrm{cone}(Dg_j(\bar{x})d)$ 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 4.1) without faces:

Corollary 4.1. *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 $\operatorname{Im}(A_j) = \{\eta_j\}^\perp$, 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 4.1 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 4.1 and its implications:

Remark 4.1. *Note that if all constraints are affine, then every feasible point satisfies the FCR property. Then, it follows from Theorem 4.2 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 2.1, 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 4.1, together with Theorem 4.2, motivates our extension of CRCQ (and RCRCQ) to NSOCP:

Definition 4.2. *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 4.1) reduces to CRCQ in this case, and so does Definition 4.2. Moreover, as mentioned before, it follows directly from Theorem 4.1, that:

Theorem 4.2. *The CRCQ condition of Definition 4.2 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 4.2. However, the converse is not true, as it can be seen in the following example:

Example 4.1. *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 4.2 holds at $\bar{x}$. However, Robinson's CQ is not satisfied at $\bar{x} = 0$, since

$$Dg(\bar{x})d = d(1, 1) \notin \operatorname{int}(K_2)$$

for every $d \in \mathbb{R}$. Consequently, nondegeneracy is not satisfied, either.

Observe that Example 4.1 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 4.2. *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 4.2. *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 3.1. 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 3.1, then reducing, and applying Definition 4.2 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 2.1 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 4.1.

Theorem 4.3. *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 2.1; that is,*

$$d^T \nabla^2 f(\bar{x}) d + \sum_{j=1}^{q} \langle D^2 g_j(\bar{x})[d, d], \bar{\lambda}_j \rangle - \sigma(d, \bar{x}, \bar{\lambda}) \geq 0, \tag{33}$$

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 4.1, 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

$$\mathcal{G}(\xi(t)) \in \operatorname{span}(F) \tag{34}$$

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, $\bar{\lambda}_j \in K^\circ_{m_j}$ 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 $\bar{\lambda}_j \in K^\circ_{m_j}$;
   - 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

$$\begin{aligned} R''(t) = & \sum_{j \in I_0(\bar{x})} \left\langle D^2 g_j(\xi(t))[\xi'(t), \xi'(t)], \bar{\lambda}_j \right\rangle + \sum_{j \in I_0(\bar{x})} \left\langle Dg_j(\xi(t))^T \bar{\lambda}_j, \xi''(t) \right\rangle \\ & + \sum_{j \in I_B(\bar{x})} [\bar{\lambda}_j]_0 \left(\left\langle D^2 \phi_j(\xi(t)) \xi'(t), \xi'(t) \right\rangle + \left\langle \nabla \phi_j(\xi(t)), \xi''(t) \right\rangle \right). \end{aligned}$$

Due to the fact $R''(t)$ is continuous, taking the limit $t \to 0$, we obtain

$$\begin{aligned} R''(0) = & \sum_{j \in I_0(\bar{x})} \left\langle D^2 g_j(\bar{x})[d, d], \bar{\lambda}_j \right\rangle + \sum_{j \in I_0(\bar{x})} \left\langle Dg_j(\bar{x})^T \bar{\lambda}_j, \xi''(0) \right\rangle \\ & + \sum_{j \in I_B(\bar{x})} [\bar{\lambda}_j]_0 \left(\left\langle D^2 \phi_j(\bar{x}) d, d \right\rangle + \frac{1}{[g_j(\bar{x})]_0} \left\langle Dg_j(\bar{x})^T R_{m_j} g_j(\bar{x}), \xi''(0) \right\rangle \right). \end{aligned}$$

The above expression can be simplified using the relation

$$\begin{aligned} \left\langle D^2 \phi_j(\bar{x}) d, d \right\rangle &= \frac{\langle \widehat{Dg_j(\bar{x}) d}, \widehat{g_j(\bar{x})} \rangle^2}{\|\widehat{g_j(\bar{x})}\|^3} - \frac{\|\widehat{Dg_j(\bar{x}) d}\|^2}{\|\widehat{g_j(\bar{x})}\|} + \left\langle D^2 g_j(\bar{x})[d, d], \frac{R_{m_j} g_j(\bar{x})}{\|\widehat{g_j(\bar{x})}\|} \right\rangle \\ &= \frac{1}{[g_j(\bar{x})]_0} \left(\left\langle R_{m_j} Dg_j(\bar{x}) d, \ Dg_j(\bar{x}) d \right\rangle + \left\langle D^2 g_j(\bar{x})[d, d], R_{m_j} g_j(\bar{x}) \right\rangle \right), \end{aligned}$$

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

$$\begin{aligned} R''(0) \quad &= \sum_{j\in I_0(\bar{x})\cup I_B(\bar{x})} \langle D^2 g_j(\bar{x})[d,d], \bar{\lambda}_j\rangle + \sum_{j\in I_0(\bar{x})\cup I_B(\bar{x})} \left\langle Dg_j(\bar{x})^T\bar{\lambda}_j, \xi''(0)\right\rangle \\ &+ \sum_{j\in I_B(\bar{x})} \frac{[\bar{\lambda}_j]_0}{[g_j(\bar{x})]_0} \langle R_{m_j} Dg_j(\bar{x})d,\ Dg_j(\bar{x})d\rangle = 0. \end{aligned} \tag{38}$$

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:

$$\nabla f(\bar{x})^T \xi''(0) = \sum_{j\in I_0(\bar{x})\cup I_B(\bar{x})} \langle D^2 g_j(\bar{x})[d,d], \bar{\lambda}_j\rangle + \sum_{j\in I_B(\bar{x})} \frac{[\bar{\lambda}_j]_0}{[g_j(\bar{x})]_0} d^T Dg_j(\bar{x})^T R_{m_j} Dg_j(\bar{x})d. \tag{39}$$

Therefore, since $\bar{\lambda}_j = 0$ for every $j \in I_{\text{int}}(\bar{x})$, plugging (39) into (35) yields

$$d^T\nabla^2 f(\bar{x})d + \sum_{j=1}^{q} \langle D^2 g_j(\bar{x})[d,d], \bar{\lambda}_j\rangle - \sigma(d,\bar{x},\bar{\lambda}) \geq 0.$$

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 4.3 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 4.3 implies the classical condition of Theorem 2.2, then it also induces a sufficient (no-gap) second-order optimality condition after replacing $\geq$ by $>$ in inequality (33).

Remark 4.3. *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]:*

$$\begin{array}{ll} \text{Minimize} & f(x) := -x, \\ s.t. & g(x) := (x, x + x^2) \in K_2, \end{array}$$

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 4.2 does not hold at $\bar{x}$, take $F := \operatorname{cone}((1,1)) \trianglelefteq K_2$ and note that

$$Dg(x)^T[F^{\perp}] = \operatorname{span}(-2x)$$

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 4.2 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 4.3 (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*

$$\mathcal{D}(x,w) := \left\{ Dg_j(x)^T (1, -w_j) \right\}_{j \in J_1} \bigcup \left\{ Dg_j(x)^T (1, w_j) \right\}_{j \in J_2} \bigcup \left\{ Dg_j(x)^T \left(1, -\frac{\widehat{g_j(x)}}{\|\widehat{g_j(x)}\|} \right) \right\}_{j \in J_3}$$

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 4.2 does not imply Seq-CRCQ.

Example 4.3. *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 4.1). 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 4.2 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 4.3 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:

$$\begin{aligned} \text{Minimize} \quad & f(x), \\ \text{s.t.} \quad & G(x) \succeq 0. \end{aligned} \tag{NSDP}$$

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 Section 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 $\mathcal{E}_{\bar{E}} \colon \mathcal{N} \to \mathbb{R}^{m \times m-r}$ 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 5.1. *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 $\Pi(\bar{Y})\bar{E} = \bar{E}$, meaning that the $m - r$ columns of $\Pi(Y)\bar{E}$ are linearly independent for every $Y \in \mathcal{N}$; as a consequence, the function $Y \mapsto \mathcal{E}_{\bar{E}}(Y) := \operatorname{gramschmidt}\left(\Pi(Y)\bar{E}\right)$ is well-defined and also analytic in (a possibly smaller) $\mathcal{N}$.

Regarding item 2, note that $\mathcal{E}_{\bar{E}}(Y)^T \mathcal{E}_{\bar{E}}(Y) = \mathbb{I}_{m-r}$ due to the Gram-Schmidt procedure, and it follows from the linear independence of the columns of $\Pi(Y)\bar{E}$ that $\operatorname{Im}(\mathcal{E}_{\bar{E}}(Y)) = \operatorname{span}(\{u_{r+1}(Y), \ldots, u_m(Y)\})$ whenever $Y \in \mathcal{N}$. Finally, observe that $\mathcal{E}_{\bar{E}}(\bar{Y}) = \Pi(\bar{Y})\bar{E} = \bar{E}$ which proves item 3. $\square$

Notice, however, that the columns of $\mathcal{E}_{\bar{E}}(Y)$ are not necessarily eigenvectors of $Y \in \mathcal{N}$.

Remark 5.1. *If $\bar{Y} = 0$, then for every orthogonal matrix $\bar{E} \in \mathbb{R}^{m \times m}$ it holds that $\mathcal{E}_{\bar{E}}(Y) = \bar{E}$ for every $Y \in \mathbb{S}^m$. Indeed, in this case we have $\operatorname{span}(\{u_1(Y), \ldots, u_m(Y)\}) = \mathbb{R}^m$ for every $Y \in \mathbb{S}^m$, and since $\bar{E}$ is itself orthogonal, it follows that*

$$\mathcal{E}_{\bar{E}}(Y) = \operatorname{gramschmidt}\left(\Pi(Y)\bar{E}\right) = \operatorname{gramschmidt}\left(\bar{E}\right) = \bar{E}$$

for every $Y \in \mathbb{S}^m$.

Now, let $\bar{x} \in \Omega$, denote the rank of $G(\bar{x})$ by r, and let $\bar{E} \in \mathbb{R}^{m \times m-r}$ be an arbitrary matrix with orthonormal columns that span $\operatorname{Ker}(G(\bar{x}))$. Let $\mathcal{E}_{\bar{E}}$ be constructed as in (40) around $\bar{Y} = G(\bar{x})$ and observe that $\mathbb{S}^m_+$ is reducible to

$$\mathcal{C} := \mathbb{S}^{m-r}_+$$

in a neighborhood $\mathcal{N}$ of $G(\bar{x})$ by the mapping $\Xi \colon \mathcal{N} \to \mathbb{S}^{m-r}$ given by

$$\Xi(Y) := \mathcal{E}_{\bar{E}}(Y)^T Y \mathcal{E}_{\bar{E}}(Y),$$

for every $Y \in \mathcal{N}$ close enough to $G(\bar{x})$ so that $\mu_i(Y) > 0$ for every $i = 1, \ldots, r$. For simplicity of notation we shall omit the subscript $\bar{E}$ from this point forth, unless $\bar{E}$ is not clear from the context. That said, define the function $E := \mathcal{E}_{\bar{E}} \circ G$, consider the reduced constraint function

$$\mathcal{G}(x) := E(x)^T G(x) E(x),$$

and for every x sufficiently close to $\bar{x}$, we have that $G(x) \in \mathbb{S}^m_+$ if, and only if, $\mathcal{G}(x) \in \mathbb{S}^{m-r}_+$. Moreover, it is worth recalling that, since the function $\mathcal{E}_{\bar{E}}$ is analytic in $\mathcal{N}$, the degree of differentiability of $\mathcal{G}$ is the same as of G.

Following Bonnans and Shapiro [24, Equation 5.161], we see that the linearized cone of the original constraints of (NSDP) at $\bar{x} \in \Omega$ can be written as

$$\mathcal{L}_\Omega(\bar{x}) = \left\{ d \in \mathbb{R}^n \mid \bar{E}^T DG(\bar{x}) d \bar{E} \succeq 0 \right\},$$

which also coincides with the linearized cone of the reduced constraint at $\bar{x}$, because $E(\bar{x}) = \bar{E}$ and for each x close enough to $\bar{x}$, we have

$$D\mathcal{G}(x)[\,\cdot\,] = DE(x)[\,\cdot\,]^T G(x)E(x) + E(x)^T DG(x)[\,\cdot\,]E(x) + E(x)^T G(x)DE(x)[\,\cdot\,],$$

so $D\mathcal{G}(\bar{x})[\,\cdot\,] = \bar{E}^T DG(\bar{x})[\,\cdot\,]\bar{E}$. For more details on this reduction approach, see [21, 23].

In the next section, we will introduce a constant rank-type condition for NSDP in terms of the faces of the reduced cone.

5.1 A facial constant rank constraint qualification for NSDP

Following the exposition of Pataki [50], the faces of $\mathcal{C} = \mathbb{S}^{m-r}_+$ can be represented in a very simple way: F is a face of $\mathbb{S}^{m-r}_+$ if, and only if, there exists an orthogonal matrix $U \in \mathbb{R}^{m-r\times m-r}$ and some $s \in \{1, \ldots, m-r\}$ such that

$$F = \left\{ U \begin{bmatrix} A_{11} & 0 \\ 0 & 0 \end{bmatrix} U^T \;\middle|\; A_{11} \in \mathbb{S}^s_+ \right\}.$$

With this in mind, let us define the analogue of Definition 4.2 for NSDP:

Definition 5.1. *Let $\bar{x} \in \Omega$ and let $\bar{E} \in \mathbb{R}^{m\times m-r}$ be a matrix with orthonormal columns that span* $\operatorname{Ker}(G(\bar{x}))$*, where r denotes the rank of $G(\bar{x})$. We say that the* facial constant rank *(FCR) property holds at $\bar{x}$* with respect to $\bar{E}$ *if there exists a neighborhood $\mathcal{V}$ of $\bar{x}$ such that, for each $F \trianglelefteq \mathbb{S}^{m-r}_+$, the dimension of $D\mathcal{G}(x)^T[F^\perp]$ remains constant for every $x \in \mathcal{V}$.*

Following the discussion after Proposition 3.1 and also after Definition 4.1, to better visualize the meaning of Definition 5.1, recall that Ω is locally equivalent to $\mathcal{G}^{-1}(\mathbb{S}^{m-r}_+)$ and that the faces of $\mathbb{S}^{m-r}_+$ can be regarded as linear approximations of $\mathbb{S}^{m-r}_+$, in some sense. Then, for every $F \trianglelefteq \mathbb{S}^{m-r}$, the set $D\mathcal{G}(x)^{-1}(\operatorname{span}(F))$ defines a possible linear approximation of Ω around $\bar{x}$. The reasoning after Proposition 3.1 still holds in the context of NSDP and it follows that the FCR property holds at $\bar{x} \in \Omega$ (with respect to $\bar{E}$) if, and only if, the dimension of $D\mathcal{G}(x)^{-1}(\operatorname{span}(F))$ remains constant for all x in a neighborhood of $\bar{x}$, at every $F \trianglelefteq \mathbb{S}^{m-r}_+$. From this point of view, the FCR property demands all linear approximations of the feasible set to remain with constant dimension in the vicinity of $\bar{x}$.

Now, we proceed to the main result of this section.

Theorem 5.1. *Let $\bar{x} \in \Omega$ and $r :=$ $\operatorname{rank}(G(\bar{x}))$. If $\bar{x}$ satisfies the FCR property with respect to some $\bar{E} \in \mathbb{R}^{m\times m-r}$ as described in Definition 5.1, then $\mathcal{T}_\Omega(\bar{x}) = \mathcal{L}_\Omega(\bar{x})$.*

Proof. Let $d \in \mathcal{L}_\Omega(\bar{x})$, denote by s the rank of $\bar{E}^T DG(\bar{x})d\bar{E}$, and let $\bar{Q} \in \mathbb{R}^{m-r\times m-r}$ be an orthogonal matrix such that

$$\bar{Q}^T \bar{E}^T DG(\bar{x})d\bar{E}\bar{Q} = \begin{bmatrix} R & 0 \\ 0 & 0 \end{bmatrix},$$

where $R \succ 0$ is an $s \times s$ diagonal matrix. Let $\bar{W}$ be the matrix formed by the columns of $\bar{Q}$ corresponding to the positive eigenvalues of $\bar{E}^T DG(\bar{x})d\bar{E}$; that is, $\bar{W}^T\bar{E}^T DG(\bar{x})d\bar{E}\bar{W} = R$. Then, consider the face of $\mathbb{S}^{m-r}_+$ given by

$$F := \left\{ \bar{Q} \begin{bmatrix} A_{11} & 0 \\ 0 & 0 \end{bmatrix} \bar{Q}^T \;\middle|\; A_{11} \in \mathbb{S}^s_+ \right\} \tag{41}$$

and note that $\bar{E}^T DG(\bar{x})d\bar{E} \in F$. Let $\eta_1, \ldots, \eta_N$ be a basis of $F^\perp$, where $N := \dim(F^\perp)$, and note that

$$D\mathcal{G}(x)^T[F^\perp] = \operatorname{span}\left(\left\{ D\mathcal{G}(x)^T[\eta_i] \right\}_{i\in\{1,\ldots,N\}} \right). \tag{42}$$

Therefore, the FCR property can be equivalently stated as the constant rank of the family

$$\left\{ D\mathcal{G}(x)^T[\eta_i] \right\}_{i\in\{1,\ldots,N\}}$$

in a neighborhood of $\bar{x}$. Furthermore, let $\zeta_i(x) := \langle \mathcal{G}(x), \eta_i \rangle$ and note that

$$\nabla\zeta_i(x) = D\mathcal{G}(x)^T[\eta_i]$$

for every $i \in \{1, \ldots, N\}$.

Then, by Proposition 4.1, there exist two neighborhoods V_1 and V_2 of $\bar{x}$, and a curve $\psi\colon V_1 \to V_2$ such that $\psi(\bar{x}) = \bar{x}$, $D\psi(\bar{x}) = \mathbb{I}_n$, and $\zeta_i(\psi^{-1}(\bar{x}+y)) = \zeta_i(\bar{x})$ for every $i \in \{1,\dots,N\}$ and every y in the subspace

$$\mathcal{S} := \{y \in \mathbb{R}^n \mid \langle \nabla\zeta_i(\bar{x}), y\rangle = 0,\ \forall i \in \{1,\dots,N\}\}.$$

Since $D\mathcal{G}(\bar{x})d \in F$, we see that $\langle d, D\mathcal{G}(\bar{x})^T[\eta_i]\rangle = \langle D\mathcal{G}(\bar{x})d, \eta_i\rangle = 0$ for every $i \in \{1,\dots,N\}$, so $d \in \mathcal{S}$. Then, let $\varepsilon > 0$ be such that $\bar{x} + td \in V_2$ for every $t \in (-\varepsilon,\varepsilon)$, and define $\xi(t) := \psi^{-1}(\bar{x}+td)$ for every such t. Moreover, note that $\xi'(t) = d$ and $\xi(0) = \bar{x}$.

Now, for every $t \in (-\varepsilon,\varepsilon)$, we have $\zeta_i(\xi(t)) = \langle \mathcal{G}(\xi(t)), \eta_i\rangle = \zeta_i(\bar{x}) = 0$ for every $i \in \{1,\dots,N\}$ because $\mathcal{G}(\bar{x}) = 0$, whence follows that

$$\mathcal{G}(\xi(t)) \in \operatorname{span}(F)$$

for every such t, meaning also

$$\bar{Q}^T\mathcal{G}(\xi(t))\bar{Q} = \begin{bmatrix} \bar{W}^T\mathcal{G}(\xi(t))\bar{W} & 0 \\ 0 & 0 \end{bmatrix}.$$

On the other hand, considering the Taylor expansion of $\mathcal{G}(\xi(t))$ around $t = 0$,

$$\mathcal{G}(\xi(t)) = \bar{E}^T G(\bar{x})\bar{E} + t\bar{E}^T DG(\bar{x})d\bar{E} + o(t) = t\bar{E}^T DG(\bar{x})d\bar{E} + o(t),$$

we observe that

$$\bar{W}^T\mathcal{G}(\xi(t))\bar{W} = t\bar{W}^T\bar{E}^T DG(\bar{x})d\bar{E}\bar{W} + o(t) = tR + o(t) \succ 0,$$

for $t \in (0,\varepsilon)$, shrinking ε if necessary. Thus,

$$\mathcal{G}(\xi(t)) \in F \subseteq \mathbb{S}^{m-r}_+$$

for every $t \in [0,\varepsilon)$, and then $G(\xi(t)) \succeq 0$ for all such t. Therefore, it follows that $d \in \mathcal{T}_\Omega(\bar{x})$.

□

Remark 5.2. *Similarly to Remark 4.1, we observe that if G is affine, then every $\bar{x} \in \Omega$ satisfies the FCR property with respect to any $\bar{E}$, which implies Definition 5.1 is not a CQ on its own, unless $H(\bar{x})$ as defined in (2) is closed. We remark this fact because it implies that the weakest CQ that guarantees zero duality gap in linear SDP problems is the closedness of $H(\bar{x})$.*

With this in mind, we present our extension of CRCQ (and RCRCQ) for NSDP:

Definition 5.2 (CRCQ)**.** *Let $\bar{x} \in \Omega$ and $r := \operatorname{rank}(G(\bar{x}))$. We say that $\bar{x}$ satisfies the* constant rank constraint qualification condition for NSDP *(CRCQ) if it satisfies the FCR property with respect to some matrix $\bar{E} \in \mathbb{R}^{m\times m-r}$ with orthonormal columns spanning* $\operatorname{Ker}(G(\bar{x}))$ *and, in addition, $H(\bar{x})$ is closed.*

And, as an immediate consequence of Theorem 5.1, we obtain the following:

Theorem 5.2. *Let $\bar{x} \in \Omega$. If $\bar{x}$ satisfies CRCQ, then it also satisfies Abadie's CQ.*

Now, we are led to compare our CRCQ condition with other CQs from the literature. First, let us show that it is, in general, independent of Robinson's CQ.

Example 5.1. *Consider the following constraint:*

$$G(x) := \begin{bmatrix} -x & 0 \\ 0 & x \end{bmatrix}.$$

The only feasible point is $\bar{x} = 0$, for which one has $G(\bar{x}) = 0 \in \mathbb{S}^2_+$. Given that G is linear, the FCR property is automatically satisfied at $\bar{x}$ with respect to any $\bar{E}$ (see Remark 5.2), so in this case CRCQ is equivalent to the closedness of the set

$$H(\bar{x}) = \left\{ DG(\bar{x})^T A \;\middle|\; A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \in \mathbb{S}^2_+ \right\}.$$

But since $DG(\bar{x})^T A = \langle DG(\bar{x}), A\rangle = a_{22} - a_{11}$, and a_{11} and a_{22} are both nonnegative, it follows that $H(\bar{x}) = \mathbb{R}$, which is closed. On the other hand, Robinson's CQ does not hold at $\bar{x}$. In fact, given a real number $d \in \mathbb{R}$, we have that

$$G(\bar{x}) + DG(\bar{x})d = \begin{bmatrix} -d & 0 \\ 0 & d \end{bmatrix},$$

which is not in $\operatorname{int}(\mathbb{S}^2_+)$ *regardless of $d \in \mathbb{R}$.*

The above example shows that CRCQ does not imply Robinson's CQ. Conversely, we will show in the next example, that Robinson's CQ does not imply CRCQ either.

Example 5.2. *Consider the following constraint given by*

$$G(x) := \begin{bmatrix} x_2 & x_1^2 \\ x_1^2 & x_2 \end{bmatrix}$$

at the point $\bar{x} = (0, 0)$*. Then, for any direction* $d = (d_1, d_2) \in \mathbb{R}^2$*, it follows that*

$$DG(x)d = \begin{bmatrix} 0 & 2x_1 \\ 2x_1 & 0 \end{bmatrix} d_1 + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} d_2.$$

It is enough to consider $d = (0, 1)$ *to see that*

$$G(\bar{x}) + DG(\bar{x})d = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \in \operatorname{int}(\mathbb{S}^2_+)$$

and Robinson's CQ holds at $\bar{x}$*. However, since* $G(\bar{x}) = 0$*, for any orthogonal matrix* $\bar{E} \in \mathbb{R}^{2\times 2}$ *we have* $\mathcal{E}_{\bar{E}}(G(x)) = \bar{E}$ *for every* $x \in \mathbb{R}^n$ *(see Remark 5.1) and* $\mathcal{G}(x) = \bar{E}^T G(x)\bar{E}$*. Then, take*

$$F := \left\{ \bar{E}^T \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} \bar{E} \;\middle|\; a \geq 0 \right\} \trianglelefteq \mathbb{S}^2_+$$

and note that

$$D\mathcal{G}(x)^T[F^\perp] = DG(x)^T[\bar{E}F^\perp\bar{E}^T] = \operatorname{span}\left(\{\nabla G_{12}(x), \nabla G_{22}(x)\}\right).$$

Then, since $\nabla G_{12}(x) = [2x_1, 0]^T$ *and* $\nabla G_{22}(x) = [0, 1]^T$*, for every* $x \in \mathbb{R}^2$*, the dimension of the subspace above is 1 at* $\bar{x}$*, but it is equal to 2 for every* x *close enough to* $\bar{x}$ *such that* $x_1 \neq 0$*. Therefore, CRCQ does not hold at* $\bar{x}$*.*

We call the reader's attention to an interesting aspect of Example 5.2, which is the fact $D\mathcal{G}(x)^T[F^\perp]$ is invariant to $\bar{E}$ in that problem. In particular, this means that CRCQ in this case could be equivalently stated as the closedness of $H(\bar{x})$ plus the fulfilment of the FCR property at $\bar{x}$ with respect to *all* $\bar{E}$, instead of *some* $\bar{E}$. We could not prove or disprove that this holds in general and, in fact, we conjecture that this is not always true. However, this holds true provided $G(\bar{x}) = 0$, or, more generally, when $\mathcal{E}_{\bar{E}}$ is a constant function, as it is proved below:

Proposition 5.1. *Let* $\bar{x} \in \Omega$*, denote* $r \doteq \operatorname{rank}(G(\bar{x}))$*, and suppose that* $\mathcal{E}_{\bar{E}}(G(x)) = \bar{E}$ *for every* x *near* $\bar{x}$ *and every matrix* $\bar{E}$ *with orthonormal columns that span* $\operatorname{Ker}(G(\bar{x}))$*. Then, the FCR property is fulfilled at* $\bar{x}$ *with respect to some* $\bar{E}$ *if, and only if, it is fulfilled with respect to all* $\bar{E}$*.*

Proof. Suppose that $\bar{x}$ satisfies the FCR property with respect to a certain $\bar{E}$. Then, let $\tilde{E} \in \mathbb{R}^{m\times m-r}$ be any matrix with orthonormal columns that span $\operatorname{Ker}G(\bar{x})$, and let $U \in \mathbb{R}^{m-r\times m-r}$ be an orthogonal (change of basis) matrix such that $\tilde{E} = \bar{E}U$. Set

$$\mathcal{G}_1(x) := \bar{E}^T G(x)\bar{E} \quad \text{and} \quad \mathcal{G}_2(x) := \tilde{E}^T G(x)\tilde{E},$$

for every $x \in \mathbb{R}^n$, and observe that, for any face $F \trianglelefteq \mathbb{S}^{m-r}_+$ and every $x \in \mathbb{R}^n$, the following holds:

$$\begin{aligned} D\mathcal{G}_2(x)^T[F^\perp] &= DG(x)^T\left[\tilde{E}F^\perp\tilde{E}^T\right] \\ &= DG(x)^T\left[\bar{E}UF^\perp U^T\bar{E}^T\right] \\ &= D\mathcal{G}_1(x)^T\left[UF^\perp U^T\right]. \end{aligned}$$

Moreover, define the linear mapping $Y \mapsto L[Y] := UYU^T$ and note that $S := UFU^T$ is a face of $L(\mathbb{S}^{m-r}_+) = \mathbb{S}^{m-r}_+$ because L is injective (which follows directly from the definition of "face of a convex set") and, moreover,

$$S^\perp = (UFU^T)^\perp = \{M \in \mathbb{S}^{m-r}_+ \mid \forall N \in F,\ \langle M, UNU^T\rangle = \langle U^T MU, N\rangle = 0\} = UF^\perp U^T.$$

Summing up the above facts, we see that for every $F \trianglelefteq \mathcal{C}$ there exists another face $S \trianglelefteq \mathcal{C}$ such that

$$D\mathcal{G}_2(x)^T[F^\perp] = D\mathcal{G}_1(x)^T[S^\perp],$$

and since the dimension of $D\mathcal{G}_1(x)^T[S^\perp]$ is assumed to be constant in a neighborhood of $\bar{x}$ by hypothesis, so is the dimension of $D\mathcal{G}_2(x)^T[F^\perp]$, further implying that if the FCR property holds with respect to $\bar{E}$, then it must also hold with respect to $\tilde{E}$ which was chosen arbitrarily and remained fixed from the beginning. The converse statement is trivial. □

Combining Remark 5.1 which states that if $G(\bar{x}) = 0$ then $\mathcal{E}_{\bar{E}}(G(x)) = \bar{E}$ for every x and every $\bar{E}$, and Proposition 5.1, we conclude that if $G(\bar{x}) = 0$ then the FCR property is invariant to the choice of $\bar{E}$. This is not surprising, for it is possible to say that this representation issue regarding $\bar{E}$ is roughly a consequence of the many possible ways of "dragging" the problem to the vertex of a reduced cone $\mathcal{C}$ before defining CRCQ and, in particular, if $G(\bar{x})$ is already at the vertex of $\mathbb{S}^m_+$ these "many possible ways" are essentially rotations, which are expected not to interfere with rank-based conditions.

Now, to resume our comparison between CRCQ and other constraint qualifications from the literature, recall that the nondegeneracy condition holds at $\bar{x}$ if, and only if, $D\mathcal{G}(\bar{x})^T$ is injective (and this holds regardless of $\bar{E}$). By the continuity of $D\mathcal{G}$, we have that if $\bar{x}$ satisfies nondegeneracy, then $D\mathcal{G}(x)^T$ remains injective for every x sufficiently close to $\bar{x}$. Then, the dimension of $D\mathcal{G}(x)^T[F^\perp]$ remains constant for every such x with respect to any $\bar{E}$, at every $F \trianglelefteq \mathbb{S}^{m-r}_+$, and it follows that nondegeneracy strictly implies CRCQ as in Definition 5.2.

Remark 5.3. *Note that our approach can be trivially extended to an NSDP problem with multiple constraints. Moreover, to deal with a separate equality constraint $h(x) = 0$, where $h\colon \mathbb{R}^n \to \mathbb{R}^p$, in the same spirit of Remark 4.2, one should consider a constraint in the form*

$$g(x) := (G(x), h(x)) \quad \text{and} \quad \mathcal{K} := \mathbb{S}^m_+ \times \{0\}^p.$$

This yields an extension of RCRCQ after applying Definition 5.2 to the reduced form of this new problem, because $F \trianglelefteq \mathcal{K}$ if, and only if, $F = R \times \{0\}^p$, where $R \trianglelefteq \mathbb{S}^m_+$ in this case. To extend CRCQ one should write the equality constraint as a pair of inequality constraints in the form $h(x) \in \mathbb{R}^p_+$ and $-h(x) \in \mathbb{R}^p_+$, giving rise to a multifold NSDP problem where $\mathbb{R}_+$ is seen as a copy of $\mathbb{S}^1_+$.

For a last comparison, we should mention a constraint qualification presented in one of our previous works [8], which was called *Sequential CRCQ* (Seq-CRCQ) therein. As a matter of fact, Seq-CRCQ differs from Definition 5.2 in many aspects. For instance, Example 5.1 shows that Seq-CRCQ is not implied by CRCQ. This example has already appeared in [8, Example 4.1], where we show that Seq-CRCQ is not satisfied at $\bar{x} = 0$; on the other hand, we showed in Example 5.1, that CRCQ holds at $\bar{x}$. Thus, CRCQ is either strictly weaker than Seq-CRCQ, or completely independent of it. Despite our efforts to clarify the converse statement, we were not able to prove nor find a counterexample for it, so this is left as an open problem.

5.2 Strong second-order optimality conditions for NSDP

The earliest work that provides a practical characterization of the sigma-term in NSDP is Shapiro's [54], using second-order directional derivatives of the least eigenvalue function, $\mu_{\min}\colon \mathbb{S}^m \to \mathbb{R}$. Shapiro proved that

$$\sigma(d, \bar{x}, \bar{\lambda}) = d^T \mathcal{H}(\bar{x}, \bar{\lambda}) d,$$

for any $d \in C(\bar{x})$ and $\bar{\lambda} \in \Lambda(\bar{x})$, where

$$\mathcal{H}(\bar{x}, \bar{\lambda}) := \left[2 \left\langle D_{x_i} G(\bar{x}) G(\bar{x})^\dagger D_{x_j} G(\bar{x}),\ \bar{\lambda} \right\rangle \right]_{i,j=1,\ldots,n}$$

and $G(\bar{x})^\dagger$ denotes the Moore-Penrose pseudoinverse of $G(\bar{x})$. Shapiro also proved that if a local minimizer $\bar{x} \in \Omega$ satisfies nondegeneracy and its associated Lagrange multiplier $\bar{\lambda} \in \Lambda(\bar{x})$ is such that $\operatorname{rank}(\bar{\lambda}) + \operatorname{rank}(G(\bar{x})) = m$ (strict complementarity), then $\bar{x}$ satisfies SOC with respect to $\bar{\lambda}$.

Later, other authors provided creative ways of obtaining SOC via some local reformulation of (NSDP) with no curvature. For instance, Lourenço et al. [42] wrote $G(x) \succeq 0$ in the form $G(x) - Z^2 = 0$ with an additional variable $Z \in \mathbb{S}^m$, and then obtained SOC for NSDP out of SOC for NLP – under nondegeneracy and strict complementarity. Forsgren [27] rediscovered Shapiro's characterization of the sigma-term and obtained SOC (under nondegeneracy, but without strict complementarity) using a special reformulation of the problem. Jarre [37] provided an elementary construction of SOC via a certain Schur complement, under nondegeneracy, strict complementarity, and assuming that the tangent cone of the linearized constraint $G(\bar{x}) + DG(\bar{x})d \in \mathcal{K}$ coincides with $\mathcal{T}_\Omega(\bar{x})$. Fukuda et al. [28] used the characterization

$$\mathbb{S}^m_+ = \{ Z \in \mathbb{S}^m \mid \|\Pi_{\mathbb{S}^m_+}(-Z)\|^2 = 0 \}$$

combined with an external penalty method and the Clarke subdifferential of $\Pi_{\mathbb{S}^m_+}$, to achieve a weaker second-order condition, which is stated only in terms of the lineality space of $C(\bar{x})$. However, their results were obtained assuming only Robinson's CQ together with the so-called *weak constant rank* (WCR) property, which is not a CQ on its own.

Following this line of research, the main contribution of this section consists of proving that every local minimizer $\bar{x}$ of (NSDP) satisfies SOC with respect to any Lagrange multiplier $\bar{\lambda} \in \Lambda(\bar{x})$ under the FCR property. In particular, when in addition $H(\bar{x})$ is closed (which leads to CRCQ), then the FCR property implies $\Lambda(\bar{x}) \neq \emptyset$. *A priori*, we make no special requirement on $\Lambda(\bar{x})$.

Theorem 5.3. *Let $\bar{x} \in \Omega$ be a local minimizer of (NSDP) and let $r := \operatorname{rank}(G(\bar{x}))$. Suppose that $\bar{x}$ satisfies the FCR property with respect to some matrix $\bar{E} \in \mathbb{R}^{m\times m-r}$ with orthonormal columns that span $\operatorname{Ker}(G(\bar{x}))$. Then, for every $\bar{\lambda} \in \Lambda(\bar{x})$, the inequality*

$$d^T \nabla^2 f(\bar{x}) d + \left\langle D^2 G(\bar{x})[d,d], \bar{\lambda} \right\rangle - \sigma(d, \bar{x}, \bar{\lambda}) \geq 0$$

holds for every $d \in C(\bar{x}) = \mathcal{L}_\Omega(\bar{x}) \cap \{\nabla f(\bar{x})\}^\perp$.

Proof. If $\Lambda(\bar{x}) = \emptyset$, then the result holds trivially; otherwise, let $\bar{\lambda} \in \Lambda(\bar{x})$ be arbitrary and fixed. Let $\bar{P} \in \mathbb{R}^{m\times r}$ be a matrix with orthonormal eigenvector columns associated with the r positive eigenvalues of $G(\bar{x})$ and define $\bar{U} := [\bar{E}, \bar{P}]$.

Now, let $d \in C(\bar{x})$ be arbitrary; so $\nabla f(\bar{x})^T d = 0$ and $\bar{E}^T DG(\bar{x}) d \bar{E} \succeq 0$. Following the proof of Theorem 5.1, let $\bar{Q} := [\bar{W}, \bar{Z}] \in \mathbb{R}^{m-r\times m-r}$ be an orthogonal matrix such that $\bar{Z}^T \bar{E}^T DG(\bar{x}) d \bar{E} \bar{Z} = 0$ and $\bar{W}^T \bar{E}^T DG(\bar{x}) d \bar{E} \bar{W} \succ 0$, and let s denote the rank of $\bar{E}^T DG(\bar{x}) d \bar{E}$. Moreover, let $F \trianglelefteq \mathbb{S}^{m-r}_+$ be defined as in (41); that is:

$$F := \left\{ \bar{Q} \begin{bmatrix} A_{11} & 0 \\ 0 & 0 \end{bmatrix} \bar{Q}^T \;\middle|\; A_{11} \in \mathbb{S}^s_+ \right\},$$

and note that $\bar{E}^T DG(\bar{x}) d \bar{E} \in F$. Similarly to the proof of Theorem 5.1, since the FCR property holds at $\bar{x}$, there exists some $\varepsilon > 0$ and a twice continuously differentiable curve $\xi \colon (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ such that $\xi(0) = \bar{x}$, $\xi'(0) = d$, and

$$\mathcal{G}(\xi(t)) \in \operatorname{span}(F)$$

for all $t \in (-\varepsilon, \varepsilon)$. Moreover, $\mathcal{G}(\xi(t)) \in F$ for every $t \in [0, \varepsilon)$. Since $\bar{x}$ is a local minimizer of (NSDP) and $\xi(t)$ is feasible for every small $t \geq 0$, then $t = 0$ is a local minimizer of the function $\phi(t) := f(\xi(t))$ subject to $t \geq 0$. Consequently, it is easy to see that

$$\phi''(0) = d^T \nabla^2 f(\bar{x}) d + \nabla f(\bar{x})^T \xi''(0) \geqslant 0. \tag{43}$$

The rest of the proof consists of computing the term $\nabla f(\bar{x})^T \xi''(0)$. By construction, we have $\mathcal{G}(\xi(t)) \in \operatorname{span}(F)$ for every $t \in (-\varepsilon, \varepsilon)$, so $\bar{Z}^T \mathcal{G}(\xi(t)) \bar{Z} = 0$ and the reduced complementarity function

$$R(t) := \left\langle \bar{Z}^T \mathcal{G}(\xi(t)) \bar{Z},\ \bar{Z}^T \bar{E}^T \bar{\lambda} \bar{E} \bar{Z} \right\rangle$$

has value zero, for all $t \in (-\varepsilon, \varepsilon)$. Therefore,

$$\begin{aligned} R'(t) \;=\; & \left\langle \bar{Z}^T (DE(\xi(t))\xi'(t))^T G(\xi(t)) E(\xi(t)) \bar{Z},\ \bar{Z}^T \bar{E}^T \bar{\lambda} \bar{E} \bar{Z} \right\rangle \\ & + \left\langle \bar{Z}^T E(\xi(t))^T DG(\xi(t))\xi'(t) E(\xi(t)) \bar{Z},\ \bar{Z}^T \bar{E}^T \bar{\lambda} \bar{E} \bar{Z} \right\rangle \\ & + \left\langle \bar{Z}^T E(\xi(t))^T G(\xi(t)) DE(\xi(t))\xi'(t) \bar{Z},\ \bar{Z}^T \bar{E}^T \bar{\lambda} \bar{E} \bar{Z} \right\rangle \end{aligned}$$

also has value zero for every small t. Differentiating once more, and taking the limit $t \to 0$, we obtain

$$\begin{aligned} R''(0) \;=\; & \left\langle \bar{Z}^T \bar{E}^T D^2 G(\bar{x})[d,d] \bar{E} \bar{Z} \;+\; \bar{Z}^T \bar{E}^T DG(\bar{x}) \xi''(0) \bar{E} \bar{Z},\ \bar{Z}^T \bar{E}^T \bar{\lambda} \bar{E} \bar{Z} \right\rangle \\ & + 2 \left\langle \bar{Z}^T (DE(\bar{x})d)^T DG(\bar{x}) d \bar{E} \bar{Z} \;+\; \bar{Z}^T \bar{E}^T DG(\bar{x}) d DE(\bar{x}) d \bar{Z},\ \bar{Z}^T \bar{E}^T \bar{\lambda} \bar{E} \bar{Z} \right\rangle \\ & + 2 \left\langle \bar{Z}^T (DE(\bar{x})d)^T G(\bar{x}) DE(\bar{x}) d \bar{Z},\ \bar{Z}^T \bar{E}^T \bar{\lambda} \bar{E} \bar{Z} \right\rangle \;=\; 0. \end{aligned} \tag{44}$$

However, following Shapiro and Fan [56, Equation 3.8], and Bonnans and Ramírez [23, Equation 67], we see that

$$DE(\bar{x}) d = D\mathcal{E}_{\bar{E}}(G(\bar{x})) DG(\bar{x}) d = -G(\bar{x})^\dagger DG(\bar{x}) d \bar{E}. \tag{45}$$

Substituting (45) into (44), the two last lines of expression (44) can be greatly simplified, which leads to the following:

$$\begin{aligned} R''(0) \;&=\; \left\langle \bar{Z}^T\bar{E}^T\left(D^2G(\bar{x})[d,d]+DG(\bar{x})\xi''(0)\right)\bar{E}\bar{Z},\ \bar{Z}^T\bar{E}^T\bar{\lambda}\bar{E}\bar{Z}\right\rangle \\ &\quad -\ 2\left\langle \bar{Z}^T\bar{E}^T(DG(\bar{x})d)^TG(\bar{x})^{\dagger}DG(\bar{x})d\bar{E}\bar{Z},\ \bar{Z}^T\bar{E}^T\bar{\lambda}\bar{E}\bar{Z}\right\rangle \;=\; 0. \end{aligned} \tag{46}$$

However, by the complementarity condition we have $\bar{\lambda}\bar{P}=0$, and using the KKT condition together with $\nabla f(\bar{x})^Td=0$ we obtain

$$\begin{aligned} 0 &= \langle d,-\nabla f(\bar{x})\rangle = \langle d, DG(\bar{x})^T\bar{\lambda}\rangle = \langle DG(\bar{x})d,\bar{\lambda}\rangle \\ &= \left\langle \bar{U}^TDG(\bar{x})d\bar{U},\ \bar{U}^T\bar{\lambda}\bar{U}\right\rangle = \left\langle \bar{E}^TDG(\bar{x})d\bar{E},\ \bar{E}^T\bar{\lambda}\bar{E}\right\rangle \\ &= \left\langle \bar{Q}^T\bar{E}^TDG(\bar{x})d\bar{E}\bar{Q},\ \bar{Q}^T\bar{E}^T\bar{\lambda}\bar{E}\bar{Q}\right\rangle \\ &= \left\langle \bar{Z}^T\bar{E}^TDG(\bar{x})d\bar{E}\bar{Z},\ \bar{Z}^T\bar{E}^T\bar{\lambda}\bar{E}\bar{Z}\right\rangle, \end{aligned}$$

but since $\bar{Z}^T\bar{E}^TDG(\bar{x})d\bar{E}\bar{Z}\succ 0$ and $\bar{W}^T\bar{E}^T\bar{\lambda}\bar{E}\bar{W}\preceq 0$, this implies

$$\bar{Z}^T\bar{E}^T\bar{\lambda}\bar{E}\bar{Z}=0,$$

which in turn implies $\bar{\lambda}\bar{E}\bar{Z}=0$. With this at hand, we obtain

$$R''(0)=\left\langle D^2G(\bar{x})[d,d]+DG(\bar{x})\xi''(0)-2(DG(\bar{x})d)^TG(\bar{x})^{\dagger}DG(\bar{x})d,\ \bar{\lambda}\right\rangle = 0, \tag{47}$$

and, by the KKT conditions, this leads to

$$\nabla f(\bar{x})^T\xi''(0) = -\langle DG(\bar{x})\xi''(0),\bar{\lambda}\rangle = \left\langle D^2G(\bar{x})[d,d]-2(DG(\bar{x})d)^TG(\bar{x})^{\dagger}DG(\bar{x})d,\ \bar{\lambda}\right\rangle. \tag{48}$$

Substituting (48) into (43) yields

$$d^T\nabla^2 f(\bar{x})d+\langle D^2G(\bar{x})[d,d],\bar{\lambda}\rangle - d^T\mathcal{H}(\bar{x},\bar{\lambda})d\geq 0.$$

Since $d\in C(\bar{x})$ was chosen arbitrarily, and $\bar{\lambda}$ is fixed from the beginning, the proof is complete.

□

As a corollary of Theorem 5.3, using the fact $\Lambda(\bar{x})\neq\emptyset$ when CRCQ holds at $\bar{x}$, we see that CRCQ ensures the fulfilment of SOC at the pair $(\bar{x},\bar{\lambda})$ for every Lagrange multiplier $\bar{\lambda}\in\Lambda(\bar{x})$.

6 Final remarks

The constant rank constraint qualification (CRCQ) is one of the most important regularity conditions in nonlinear programming (NLP), with several relevant applications regarding global convergence of algorithms, second-order optimality conditions, and some topics of stability theory. However, one of the main reasons why most of these interesting results still remain exclusive to NLP is that CRCQ itself seems intrinsic to NLP. Until very recently, there was no extension or analogue of it in the conic programming context. In a recent pair of papers [8, 9], we presented an extension of CRCQ for nonlinear semidefinite and second-order cone programming using sequences and the eigenvector structure of their respective cones, which would allow us to adopt a strategy similar to the existing nonlinear programming literature. See also [7]. While this is interesting from the point of view of algorithms, it may not be an appropriate tool for other uses. Therefore, in this paper we adopted a more innovative approach: we first characterized CRCQ for NLP in a geometrical way, by means of the faces of a reduced cone, and then we showed this geometrical characterization could carry the essence of CRCQ to more general contexts. As far as we know, this is also the first time an intuitive interpretation of CRCQ was ever presented, and it is surprisingly simple: CRCQ describes the situations where every possible linear approximation of the feasible set (around a point of interest) preserves its dimension under small perturbations. As a side note, we should mention that this definition is either independent or strictly weaker than the ones presented in [8, 9].

As an application of our results, we obtained a strong second-order necessary optimality condition under CRCQ, in terms of any given Lagrange multiplier. This improves the classical result that is obtained under nondegeneracy, and serves as an alternative for the condition that can be obtained under Robinson's CQ, where for each direction in the critical cone, there is a Lagrange multiplier satisfying the second-order condition. We expect CRCQ to be an alternative to Robinson's CQ in other situations, especially those related with stability analysis of parametric nonlinear conic optimization programs, in view of the nonlinear programming literature – see, for instance, references [33, 36, 47]. Given that CRCQ is independent of Robinson's CQ, we believe that this work allows the development of a new parallel strand in the study of stability. In a recent work, Gfrerer and Mordukhovich [29], fully characterized *tilt stable* local minimizers of NLP problems under the so-called *bounded extreme point property*, which is implied by CRCQ (improving a previous work that assumed CRCQ and MFCQ [49]). Assessing the usability of CRCQ as presented in this paper in an eventual extension of their results to NSOCP and NSDP may be regarded as a future work idea. Moreover, we expect CRCQ to be useful for supporting the convergence theory of some iterative algorithms, and also to encourage the development of algorithms that rely on faces for solving nonlinear conic problems.

The techniques employed in this paper suggest that a further extension of CRCQ, for general reducible cones, is possible. In this paper we adopted a more pragmatic approach by working explicitly with NSOCP and NSDP for clarity, leaving the investigation of a more general result to future works. In fact, it would also be interesting to not rely on reducibility at all, which should be possible by taking into account the faces of the tangent cone to $\mathcal{K}$ at $g(\bar{x})$, or perturbations of it, instead of the faces of the reduced cone $\mathcal{C}$. Furthermore, this work may inspire extensions of weaker constant rank-type conditions from NLP (together with their applications) to the conic environment, with emphasis on the well-established *constant positive linear dependence* condition [10, 12, 51] and the *constant rank of the subspace component* condition [11].

Funding

This work has received financial support from CEPID-CeMEAI (FAPESP 2013/07375-0), FAPESP (grants 2018/24293-0, 2017/18308-2, 2017/17840-2, 2017/12187-9, and 2020/00130-5), 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 Centro de Modelamiento Matemático (CMM), ACE210010 and FB210005, BASAL funds for center of excellence, both from ANID (Chile).

References

[1] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming Series B*, 95:3–51, 2003. doi:10.1007/s10107-002-0339-5.

[2] E. D. Andersen, C. Roos, and T. Terlaky. Notes on duality in second order and p-order cone optimization. *Optimization*, 51(4):627–643, 2002. doi:10.1080/0233193021000030751.

[3] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. On augmented Lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, 18(4):1286–1309, 2008. doi:10.1137/060654797.

[4] R. Andreani, C. E. Echagüe, and M. L. Schuverdt. Constant-rank condition and second-order constraint qualification. *Journal of Optimization theory and Applications*, 146(2):255–266, 2010. doi:10.1007/s10957-010-9671-8.

[5] R. Andreani, E. H. Fukuda, G. Haeser, H. Ramírez C., D. O. Santos, P. J. S. Silva, and T. P. Silveira. Erratum to: New constraint qualifications and optimality conditions for second order cone programs. *To appear in Set-Valued and Variational Analysis*, 2021. doi:10.1007/s11228-021-00573-5.

[6] R. Andreani, G. Haeser, L. M. Mito, H. Ramırez, D. O. Santos, and T. P. Silveira. Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming. *To appear in Optimization Letters*, 2020. doi:10.1007/s11590-021-01737-w.

[7] R. Andreani, G. Haeser, L. M. Mito, and H. Ramírez C. Weak notions of nondegeneracy in nonlinear semidefinite programming. Technical report, 2020. Available at ArXiv. URL: https://arxiv.org/abs/2012.14810.

[8] R. Andreani, G. Haeser, L. M. Mito, and H. Ramírez C. Sequential constant rank constraint qualifications for nonlinear semidefinite programming with applications. Technical report, 2021. Available at ArXiv. URL: https://arxiv.org/abs/2106.00775v2.

[9] R. Andreani, G. Haeser, L M. Mito, H. Ramírez C., and T. P. Silveira. Sequential constant rank for nonlinear second-order cone programming problems. Technical report, 2021.

[10] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programming*, 135(1-2):255–273, 2012. doi:10.1007/s10107-011-0456-0.

[11] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. Two new weak constraint qualifications and applications. *SIAM Journal on Optimization*, 22(3):1109–1135, 2012. doi:10.1137/110843939.

[12] R. Andreani, J. M. Martínez, and 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. doi:10.1007/s10957-004-1861-9.

[13] R. Andreani, J. M. Martínez, and M. L. Schuverdt. On second-order optimality conditions for nonlinear programming. *Optimization*, 56:529–542, 2007. doi:10.1080/02331930701618617.

[14] M. Anitescu. Degenerate nonlinear programming with a quadratic growth condition. *SIAM Journal on Optimization*, 10(4):1116–1135, 2000. doi:10.1137/S1052623499359178.

[15] A. Arutyunov. Second-order conditions in extremal problems. The abnormal points. *Transactions of the American Mathematical Society*, 350(11):4341–4365, 1998. doi:10.1090/S0002-9947-98-01775-9.

[16] A. Auslender and H. Ramírez. Penalty and barrier methods for convex semidefinite programming. *Mathematical Methods of Operations Research*, 63:Article number: 195, 2006. doi:10.1007/s00186-005-0054-0.

[17] A. Baccari. On the classical necessary second-order optimality conditions. *Journal of Optimization Theory and Applications*, 123(1):213–221, 2004. doi:10.1023/B:JOTA.0000043998.04008.e6.

[18] R. Behling, G. Haeser, A. Ramos, and D. S. Viana. On a conjecture in second-order optimality conditions. *Journal of Optimization Theory and Applications*, 176(3):625–633, 2018. Extended version at arXiv:1706.07833. arXiv:arXiv:1706.07833v1, doi:10.1007/s10957-018-1229-1.

[19] A. Ben-Tal and J. Zowe. A unified theory of first and second order conditions for extremum problems in topological vector spaces. In *Optimality and stability in mathematical programming (Mathematical Programming Studies)*, pages 39–76. Springer, 1982. doi:10.1007/BFb0120982.

[20] J. F. Bonnans. A semi-strong sufficiency condition for optimality in non convex programming and its connection to the perturbation problem. *Journal of Optimization Theory and Applications*, 60:7–18, 1989.

[21] J. F. Bonnans, R. Cominetti, and A. Shapiro. Second order optimality conditions based on parabolic second order tangent sets. *SIAM Journal on Optimization*, 9(2):466–492, 1999. doi:10.1137/S1052623496306760.

[22] J. F. Bonnans and H. Ramírez. Perturbation analysis of second-order cone programming problems. *Mathematical Programming*, 104(2):205–227, 2005. doi:10.1007/s10107-005-0613-4.

[23] J. F. Bonnans and H. Ramírez. Strong regularity of semidefinite programming problems. Technical report, 2005. DIM-CMM N° B-05/06-137.

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

[25] E. Börgens, C. Kanzow, P. Mehlitz, and 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. doi:10.1137/19M1306804.

[26] R. Cominetti. Metric regularity, tangent sets, and second-order optimality conditions. *Applied Mathematics and Optimization*, 21(1):265–287, 1990. doi:10.1007/BF01445166.

[27] A. Forsgren. Optimality conditions for nonconvex semidefinite programming. *Mathematical Programming*, 88(1):105–128, 2000. doi:10.1007/PL00011370.

[28] E. H. Fukuda, G. Haeser, and L. M. Mito. Second-order analysis for semidefinite and second-order cone programming via sequential optimality conditions. Technical report, 2020. Available at Optimization Online. URL: http://www.optimization-online.org/DB_HTML/2020/08/7951.html.

[29] H. Gfrerer and B. S. Mordukhovich. Complete characterizations of tilt stability in nonlinear programming under weakest qualification conditions. *SIAM Journal on Optimization*, 25(4):2081–2119, 2015. doi:10.1137/15M1012608.

[30] F. J. Gould and J. W. Tolle. Optimality conditions and constraint qualifications in Banach space. *Journal of Optimization Theory and Applications*, 15(6):667–684, 1975. doi:10.1007/BF00935506.

[31] M. Guignard. Generalized Kuhn-Tucker conditions for mathematical programming in a Banach space. *SIAM Journal of Control*, 7:232–241, 1969. doi:10.1137/0307016.

[32] L. Guo, H.-H. Lin, and Ye. Second-order optimality conditions for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 158:33–64, 2013. doi:10.1007/s10957-012-0228-x.

[33] L. Guo, H.-H. Lin, J. J. Ye, and J. Zhang. Sensitivity analysis of the value function for parametric mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 24(3):1206–1237, 2014. doi:10.1137/130929783.

[34] R. Henrion, A. Y. Kruger, and J. V. Outrata. Some remarks on stability of generalized equations. *Journal of Optimization Theory and Applications*, 159:681–697, 2013. doi:10.1007/s10957-012-0147-x.

[35] T. Hoheisel, C. Kanzow, and A. Schwartz. Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. *Mathematical Programming*, 137:257–288, 2013. doi:10.1007/s10107-011-0488-5.

[36] R. Janin. Directional derivative of the marginal function in nonlinear programming. In *Sensitivity, Stability and Parametric Analysis (Mathematical Programming Studies)*, pages 110–126. Springer Berlin Heidelberg, 1984. doi:10.1007/BFb0121214.

[37] F. Jarre. Elementary optimality conditions for nonlinear SDPs. In *Handbook on Semidefinite, Conic and Polynomial Optimization*. International Series in Operations Research & Management Science, 2012. doi:10.1007/978-1-4614-0769-0_16.

[38] H. Jiang and D. Ralph. Smooth SQP methods for mathematical programs with nonlinear complementarity constraints. *SIAM Journal on Optimization*, 10(3):779–808, 2000. doi:10.1137/S1052623497332329.

[39] T. Kato. *Perturbation theory for linear operators*. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[40] H. Kawasaki. An envelope-like effect of infinitely many inequality constraints on second-order necessary conditions for minimization problems. *Mathematical Programming*, 41(1):73–96, 1988. doi:10.1007/BF01580754.

[41] T. W. Liu. A reduced hessian sqp method for inequality constrained optimization. *Computational Optimization and Applications*, 49:31–59, 2011. doi:10.1007/s10589-009-9285-y.

[42] B. F. Lourenço, E. H. Fukuda, and M. Fukushima. Optimality conditions for nonlinear semidefinite programming via squared slack variables. *Mathematical Programming*, 166:1–24, 2016. doi:10.1007/s10107-016-1040-4.

[43] M. C. Maciel, S. A. Santos, and G. N. Sottosanto. On second-order optimality conditions for vector optimization. *Journal of Optimization Theory and Applications*, 149:332–351, 2011. doi:10.1007/s10957-010-9793-z.

[44] P. Mehlitz and L. I. Minchenko. R-regularity of set-valued mappings under the relaxed constant positive linear dependence constraint qualification with applications to parametric and bilevel optimization. *To appear in Set-Valued and Variational Analysis*, 2021. doi:10.1007/s11228-021-00578-0.

[45] L. Minchenko and A. Leschov. On strong and weak second-order necessary optimality conditions for nonlinear programming. *Optimization*, 65(9):1693–1702, 2016. doi:10.1080/02331934.2016.1179300.

[46] L. Minchenko and S. Stakhovski. On relaxed constant rank regularity condition in mathematical programming. *Optimization*, 60(4):429–440, 2011. doi:10.1080/02331930902971377.

[47] L. Minchenko and S. Stakhovski. Parametric nonlinear programming problems under the relaxed constant rank condition. *SIAM Journal on Optimization*, 1(314–332), 2011. doi:10.1137/090761318.

[48] M. R. C. Monte and V. A. de Oliveira. A constant rank constraint qualification in continuous-time nonlinear programming. *Set-Valued and Variational Analysis*, 29:61–81, 2021. doi:10.1007/s11228-020-00537-1.

[49] B. S. Mordukhovich and T. T. A. Nghia. Second-order characterizations of tilt stability with applications to nonlinear programming. *Mathematical Programming*, 149:83–104, 2015. doi:10.1007/s10107-013-0739-8.

[50] G. Pataki. The geometry of semidefinite programming. In R. Saigal, L. Vandenberghe, and H. Wolkowicz, editors, *Handbook of Semidefinite Programming.*, pages 29–65. Kluwer Academic Publishers, Waterloo, Canada, 2000. doi:10.1007/978-1-4615-4381-7_3.

[51] L. Qi and Z. Wei. On the constant positive linear dependence conditions and its application to SQP methods. *SIAM Journal on Optimization*, 10(4):963–981, 2000. doi:10.1137/S1052623497326629.

[52] S. Qiu. A globally convergent regularized interior point method for constrained optimization. *To appear in Optimization Methods and Software*, 2021. doi:10.1080/10556788.2021.1908283.

[53] S. M. Robinson. First-order conditions for general nonlinear optimization. *SIAM Journal on Applied Mathematics*, 30(4):597–610, 1976. doi:10.1137/0130053.

[54] A. Shapiro. First and second order analysis of nonlinear semidefinite programs. *SIAM Journal on Optimization*, 77(1):301–320, 1997. doi:10.1007/BF02614439.

[55] A. Shapiro. On uniqueness of Lagrange multipliers in optimization problems subject to cone constraints. *SIAM Journal on Optimization*, 7:508–518, 1997. doi:10.1137/S1052623495279785.

[56] A. Shapiro and M. K. H. Fan. On eigenvalue optimization. *SIAM Journal on Optimization*, 5(3):552–569, 1995. doi:10.1137/0805028.

[57] S. Steffensen and M. Ulbrich. A new relaxation scheme for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 20(5):2504–2539, 2010. doi:10.1137/090748883.

[58] S. J. Wright. Modifying SQP for degenerate problems. *SIAM Journal on Optimization*, 13(2):470–497, 2002. doi:10.1137/S1052623498333731.

[59] M. Xu and J. J. Ye. Relaxed constant positive linear dependence constraint qualification and its application to bilevel programs. *Journal of Global Optimization*, 78:181–205, 2020. doi:10.1007/s10898-020-00907-x.

[60] Y. Zhang and L. Zhang. New constraint qualifications and optimality conditions for second order cone programs. *Set-Valued and Variational Analysis*, 27:693–712, 2019. doi:10.1007/s11228-018-0487-2.

Appendix D

External reference IV

Title: Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming.

Status: Accepted for publication in Optimization Letters.

What is attached: Our copy of the paper.

Optimization Letters
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

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

Published online: 13 April 2021

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 $K^\circ := \{v \in \mathbb{R}^m \mid \langle v, y\rangle \le 0, \forall y \in K\}$. 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):

$$\begin{aligned} \text{Minimize} \quad & f(x), \\ \text{s.t. } h_i(x) = 0, \quad & i = 1, \dots, p, \\ g_j(x) \le 0, \quad & j = 1, \dots, q, \end{aligned} \tag{1}$$

where $f, h_i, g_j \colon \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable functions. We denote by $A(x^*) := \{j \in \{1, \dots, q\} \mid g_j(x^*) = 0\}$, 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:

$$\mathscr{L}(x^*) := \left\{d \in \mathbb{R}^n \mid \nabla h_i(x^*)^T d = 0, i = 1, \dots, p;\ \nabla g_j(x^*)^T d \le 0, j \in A(x^*)\right\}.$$

Its polar may be computed via Farkas' Lemma, obtaining:

$$\mathscr{L}(x^*)^\circ = \left\{ v \in \mathbb{R}^n \;\middle|\; v = \sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*), \mu_j \ge 0, j \in A(x^*) \right\}.$$

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 $I \subseteq \{1, \dots, p\}$ 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:

$$\begin{aligned} \text{Minimize} \quad & f(x), \\ \text{s.t.} \quad & h_i(x) = 0, \quad i = 1, \dots, p, \\ & g_j(x) \in K_{m_j}, \quad j = 1, \dots, \ell, \end{aligned} \tag{2}$$

where the functions are continuously differentiable and the second-order cones are denoted by $K_{m_j} := \{(z_0, \bar{z}) \in \mathbb{R} \times \mathbb{R}^{m_j - 1} \mid z_0 \geq \|\bar{z}\|\}$ 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 $\lambda \in \mathbb{R}^p$, $\mu_j \in K_{m_j}$, $j = 1, \dots, \ell$, such that

$$\nabla_x L(x^*, \lambda, \mu) = \nabla f(x^*) + J_h(x^*)^T \lambda - \sum_{j=1}^{\ell} J_{g_j}(x^*)^T \mu_j = 0, \tag{3}$$

$$\langle \mu_j, g_j(x^*) \rangle = 0, \quad j = 1, \dots, \ell. \tag{4}$$

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

$$\begin{aligned} \text{int}(K_{m_j}) &:= \{(z_0, \bar{z}) \in \mathbb{R} \times \mathbb{R}^{m_j - 1} \mid z_0 > \|\bar{z}\|\}, \\ \text{bd}^+(K_{m_j}) &:= \{(z_0, \bar{z}) \in \mathbb{R} \times \mathbb{R}^{m_j - 1} \mid z_0 = \|\bar{z}\| > 0\}. \end{aligned}$$

Thus, given a feasible point x^*, we introduce the index sets:

$$\begin{aligned} I_{int}(x^*) &:= \{j \in \{1, \dots, \ell\} \mid g_j(x^*) \in \text{int}(K_{m_j})\}, \\ I_B(x^*) &:= \{j \in \{1, \dots, \ell\} \mid g_j(x^*) \in \text{bd}^+(K_{m_j})\}, \\ I_0(x^*) &:= \{j \in \{1, \dots, \ell\} \mid g_j(x^*) = 0\}. \end{aligned}$$

Moreover, the complementarity condition (4) can be equivalently written as

$$\mu_j \circ g_j(x^*) = 0, \quad j = 1, \dots, \ell, \tag{5}$$

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 $\lambda_i, i = 1, \ldots, p$; $\mu_j \geq 0,\ j \in A(x^*)$, 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 $v_1, \ldots, v_{p+q} \in \mathbb{R}^n$ be such that $\{v_i\}_{i=1}^p$ are linearly independent. Consider scalars $\beta_i, i = 1, \ldots, p+q$, and denote $y := \sum_{i=1}^{p+q} \beta_i v_i$. Then, there exist $J \subseteq \{p+1, \ldots, p+q\}$ and scalars $\hat{\beta}_i, i \in \{1, \ldots, p\} \cup J$, such that $\{v_i\}_{i \in \{1,\ldots,p\} \cup J}$ are linearly independent, $\beta_i > 0$ implies $\hat{\beta}_i > 0$, for all $i \in J$, and $y = \sum_{i \in \{1,\ldots,p\} \cup J} \hat{\beta}_i v_i$.*

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 $I^k \subseteq \{1, \ldots, p\}$, $J^k \subseteq A(x^*)$, $\tilde{\mu}_j^k \geq 0$, $j \in J^k$, 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 $I_k = \{1, \ldots, p\}$, 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

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

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

$$\left.\begin{aligned} J_h(x^*)^T \lambda + \sum_{j=1}^{\ell} J_{g_j}(x^*)^T \mu_j = 0, \\ g_j(x^*) \cdot \mu_j = 0, \;\; \forall j = 1, \ldots, \ell, \\ \mu_j \in \mathbb{S}^{m_j}_+, \;\; \forall j = 1, \ldots, \ell, \end{aligned}\right\} \quad \Rightarrow \quad \mu_j = 0, \;\; \forall j = 1, \ldots, \ell. \tag{16}$$

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

$$I_R(x^*) := \{j \in \{1, \ldots, \ell\} \mid 0 = \sigma_{\min}(g_j(x^*)) \text{ is simple}\}$$

and

$$I_N(x^*) := \{j \in \{1, \ldots, \ell\} \mid 0 = \sigma_{\min}(g_j(x^*)) \text{ is not simple}\}.$$

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*

$$\begin{aligned} & \nabla f(x^k) + J_h(x^k)^T \lambda^k - \sum_{j \in I_N(x^*)} J_{g_j}(x^k)^T \mu_j^k \\ & \quad - \sum_{j \in I_R(x^*)} \alpha_j^k \nabla \sigma_{\min}(g_j(x^k)) \to 0, \end{aligned} \tag{21}$$

$$\sigma_i(g_j(x^*)) > 0 \Rightarrow \sigma_i(\mu_j^k) \to 0, \quad i = 1, \ldots, m_j, \quad \forall j \in I_N(x^*), \tag{22}$$

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

$$J_-(x^*) := \Big\{ j_0 \in I_R(x^*) \Big| -\nabla\sigma_{min}(g_{j_0}(x^*)) = \sum_{i=1}^{p}\lambda_i\nabla h_i(x^*) + \sum_{j\in I_R(x^*)}\alpha_j\nabla\sigma_{\min}(g_j(x^*)), \text{ for some } \lambda_i \in \mathbb{R}, \alpha_j \geq 0 \Big\}, \tag{24}$$

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 $I \subseteq \{1, \ldots, p\}$ 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 E

External reference V

Title: On the weak second-order optimality condition for nonlinear semidefinite and second-order cone programming.

Status: Under review.

What is attached: A preprint that has *not* been peer reviewed yet.

On the weak second-order optimality condition for nonlinear semidefinite and second-order cone programming

Ellen H. Fukuda[1†], Gabriel Haeser[1†] and Leonardo M. Mito[1*†]

[1]Graduate School of Informatics, Kyoto University, Kyoto, Japan.
[2*]Department of Applied Mathematics, University of São Paulo, São Paulo, Brazil.

*Corresponding author(s). E-mail(s): leokoto@ime.usp.br;
Contributing authors: ellen@i.kyoto-u.ac.jp; ghaeser@ime.usp.br;
†These authors contributed equally to this work.

Abstract

Second-order necessary optimality conditions for nonlinear conic programming problems that depend on a single Lagrange multiplier are usually built under nondegeneracy and strict complementarity. In this paper we establish a condition of such type for two classes of nonlinear conic problems, namely semidefinite and second-order cone programming, assuming Robinson's constraint qualification and a weak constant rank-type property which are, together, strictly weaker than nondegeneracy. Our approach is done via a penalty-based strategy, which is aimed at providing strong global convergence results for first- and second-order algorithms. Since we are not assuming strict complementarity, the critical cone does not reduce to a subspace, thus, the second-order condition we arrive at is defined in terms of the lineality space of the critical cone. In the case of nonlinear programming, this condition reduces to the standard second-order condition widely used as second-order stationarity measure in the algorithmic practice.

The authors received financial support from Fundação de Amparo à Pesquisa do Estado de São Paulo (grants 2018/24293-0, 2017/18308-2, and 2017/17840-2), Conselho Nacional de Desenvolvimento Científico e Tecnológico (grants 303427/2018-3 and 404656/2018-8), Fundação Carlos Chagas de Amparo à Pesquisa do Estado do Rio de Janeiro (grant E-26/010.001247/2016), and a Grant-in-Aid for scientific research (C)(19K11840) from Japan Society for the Promotion of Science.

Keywords: Optimality conditions, Second-order conditions, Weak constant rank, Semidefinite programming, Second-order cone programming.

MSC Classification: 90C46 , 90C30 , 90C26 , 90C22.

1 Introduction

Consider the following *nonlinear conic programming* (NCP) problem in standard form:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & g(x) \in \mathcal{K}, \\ & h(x) = 0, \end{aligned} \tag{NCP}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$, $g\colon \mathbb{R}^n \to \mathbb{E}$ and $h\colon \mathbb{R}^n \to \mathbb{R}^p$ are twice continuously differentiable functions, $\mathbb{E}$ is a finite-dimensional linear space equipped with an inner product $\langle\cdot,\cdot\rangle$ and the norm $||\cdot||$ induced by it, and $\mathcal{K} \subseteq \mathbb{E}$ is a closed convex cone that is assumed to be *self-dual*, which means $\mathcal{K} = \mathcal{K}^* \doteq \{w \in \mathbb{E}\colon \forall y \in \mathcal{K}, \langle w, y\rangle \geqslant 0\}$.

We are primarily interested in second-order necessary optimality conditions for two well-established particular cases of (NCP):

- *Nonlinear second-order cone programming* (NSOCP), which is obtained when $\mathbb{E} = \mathbb{R}^m$ and $\mathcal{K}$ is the so-called (Lorentz) *second-order cone*, defined as $\mathbb{L}^m \doteq \{(w_0, \overline{w}) \in \mathbb{R}\times\mathbb{R}^{m-1}\colon w_0 \geqslant ||\overline{w}||_2\}$ when $m > 1$ and $\mathbb{L}^1 \doteq \{w \in \mathbb{R}\colon w \geqslant 0\}$, or the Cartesian product of r second-order cones in $\mathbb{R}^{m_i}$, with $i \in \{1,\dots,r\}$ and $m_1 + \cdots + m_r = m$;
- *Nonlinear semidefinite programming* (NSDP), which is obtained when $\mathbb{E} = \mathbb{S}^m$ is the space of all $m\times m$ real symmetric matrices and $\mathcal{K}$ is the cone $\mathbb{S}^m_+ \doteq \{W \in \mathbb{S}^m\colon \forall d \in \mathbb{R}^m, d^\top W d \geqslant 0\}$ of all positive semidefinite matrices, or a Cartesian product in the form $\mathcal{K} = \mathbb{S}^{m_i}_+ \times\cdots\times \mathbb{S}^{m_r}_+$, with $m_1+\cdots+m_r = m$.

Both fields have grown independently and accumulated a large set of applications over the years, for example, in robust control [32, 33], passive reduced-order modelling [36], structural optimization [49, 54], the sphere covering problem [24], and others (see [19, 55, 68] for a vast collection of examples). In conjunction, several algorithms have been developed for them, such as interior-point methods [23, 47, 69], sequential quadratic programming methods [51, 52], Newton-type methods [38, 50], and augmented Lagrangian methods [4, 15, 71], to name a few (see Yamashita and Yabe [70] for more details). Consequently, some theoretical aspects of NSOCP and NSDP, such as optimality conditions and regularity, have gained much relevance in the community as well. In particular, necessary optimality conditions are especially useful for giving theoretical global convergence support for iterative algorithms, in the sense that every feasible limit point of a given algorithm can be proven to satisfy some necessary optimality condition under a set of hypotheses. In fact, the reliability of an algorithm is deeply related with the strength of the optimality condition that supports its global convergence theory. From this point of view, second-order necessary optimality conditions improve the first-order ones by considering the curvature of the problem data over the set of directions where first-order information has little meaning, which is usually called *cone of critical directions* (or *critical cone*). Note that this kind of convergence theory is different from what is usually done for convex optimization problems, where *second-order sufficient conditions* are used as convergence hypotheses. In the nonconvex case, the latter results in a local convergence analysis. Since the results of this paper are meant to be used in the aid of global convergence, we focus on necessary optimality conditions.

It is worth mentioning that second-order analysis in non-polyhedral conic contexts, such as NSOCP and NSDP, is considerably more intricate than in polyhedral contexts, such as in *nonlinear programming* (NLP). This is justified by the fact that the curvature of $\mathcal{K}$ must be taken into account, besides the curvature of the functions defining the problem. The initial efforts to characterize this curvature were done by Kawasaki [53], whose results were generalized and refined by Cominetti [30], and later completed by Bonnans, Cominetti, and Shapiro [27] with the notion of *second-order regularity*. Then, Shapiro [65] obtained a specialized statement for it in the context of NSDP, that was later re-discovered by Forsgren [35], Jarre [48], and Lourenço, Fukuda, and Fukushima [56], using distinct nontrivial techniques that make each proof interesting on its own. For NSOCP, second-order necessary optimality conditions were first characterized by Bonnans and Ramírez [28], and later studied by Fukuda and Fukushima [37] who also presented sufficient conditions. Thus, it is possible to say that the motivation for studying alternative ways of deriving second-order conditions for conic problems, in particular NSDP and NSOCP, has gone far beyond practical usage, but nevertheless we believe practice should not be ignored.

With this in mind, some useful tools for proving new first- and second-order optimality conditions for optimization problems, which are deeply connected to the algorithmic approach, are the so-called *sequential optimality conditions*. They were introduced in NLP, and later extended to NSOCP and NSDP, as KKT variants designed for building convergence theory of iterative algorithms (for details, we refer to the work of Andreani et al. [4, 6, 7, 15, 17]) and they gained some attention for being able to sharpen most convergence results for them in a general and unified manner (see, for instance, [14, Sec. 5.2]). Also, a second-order sequential condition has recently appeared in the work of Andreani et al. [13] for NLP, which not only provided an ideal way of incorporating second-order information in numerical methods, but also an intuitive strategy for building second-order analysis under weaker hypotheses than the traditional *linear independence constraint qualification* (LICQ). These improvements were obtained by considering a somewhat "weak" second-order necessary optimality condition in the sense that only the lineality space of the critical cone is taken into account in their results. However, as it is well-known in NLP, this "weak" condition is the most suitable second-order condition for global convergence analysis of algorithms, since the stronger conditions that deal with the whole critical cone are not guaranteed to be fulfilled at the convergence points of a large class of algorithms, such as barrier-type methods [40] and augmented Lagrangian-type methods [18], even under very strong hypotheses. Besides, checking the validity of the "strong" second-order condition is an NP-hard class problem, whereas checking the "weak" condition is of polynomial class. Nevertheless, as far as we know, the latter condition has never received due attention in nonconvex conic contexts other than NLP.

Inspired by [13], we prove that every local minimizer satisfies the weaker version of the second-order necessary condition, for NSOCP and NSDP, but under weaker assumptions than all previous related works. In fact, the meaning of our results lies in the fact we assume neither *nondegeneracy* nor *strict complementarity*, since under these hypotheses the "weak" and "strong" second-order conditions are equivalent. Our approach is based on sequential conditions, which suggests that our results may be useful for proving convergence of algorithms to second-order stationary points of NSOCP and NSDP problems. We stress that even though NSOCP can be represented in terms of NSDP, it is interesting to discriminate them since the numerical methods designed to solve each problem might have different performances in practice [1]. Also, it is not straightforward to derive second-order results for NSOCP only based on the NSDP results.

This paper is structured as follows: we begin by reviewing some classical results on first- and second-order optimality conditions for (NCP) and its particular cases with some degree of details, in Section 2. Then, we present our second-order analysis for NSOCP in Section 3, and for NSDP in Section 4. At last, in Section 5 we give some final considerations about this paper and related works.

2 Technical background

In this section, we introduce our notation and present some results from the literature that are directly related to ours. We also review in details some classical results on first- and second-order optimality conditions for NSOCP and NSDP.

We consider the standard inner product in $\mathbb{R}^n$, given by $\langle a, b\rangle \doteq \sum_{i=1}^{n} a_i b_i$, and the Euclidean norm, given by $||a||_2 \doteq \sqrt{\langle a, a\rangle}$, for every $a, b \in \mathbb{R}^n$. The terms $\operatorname{int}(\mathcal{K})$, $\operatorname{bd}(\mathcal{K})$, and $\operatorname{bd}^+(\mathcal{K})$ stand for the *interior, boundary*, and *boundary excluding the origin* of $\mathcal{K}$, respectively. Also, for any closed convex cone C, $\operatorname{lin}(C) \doteq C \cap (-C)$ denotes its *lineality space*, which is the largest subspace contained in C.

For a given finite indexed set $\{a_i \colon i \in \{1, \dots, k\}\} \subset \mathbb{R}$, we denote the array that has a_i in its i-th position by $[a_i]_{i\in\{1,\dots,k\}} \in \mathbb{R}^k$ and, analogously, the matrix whose entries are the elements of $\{b_{ij} \colon i, j \in \{1, \dots, k\}\} \subset \mathbb{R}$ is denoted by $[b_{ij}]_{i,j\in\{1,\ldots,k\}} \in \mathbb{R}^{k\times k}$. The identity matrix of $\mathbb{R}^{n\times n}$ is denoted by $\mathbb{I}_n$. The *gradient* and the *Hessian* of a function $f\colon \mathbb{R}^n \to \mathbb{R}$ at an arbitrary point $x \in \mathbb{R}^n$ are represented by $\nabla f(x)$ and $\nabla^2 f(x)$, respectively, and the first derivative of $g\colon \mathbb{R}^n \to \mathbb{E}$ at x is the linear mapping $Dg(x)[\cdot]\colon \mathbb{R}^n \to \mathbb{E}$ defined by the action

$$Dg(x)[h] \doteq \sum_{i=1}^{n} \partial_i g(x) h_i$$

for every $h \in \mathbb{R}^n$, where $\partial_i g(x) \in \mathbb{E}$ is the partial derivative of g in the i-th variable, at x. In particular, if $\mathbb{E} = \mathbb{R}^m$ then $Dg(x)$ is exactly the Jacobian matrix of g at x, in the canonical basis of $\mathbb{R}^m$; for instance, in this case the i-th row of $Dg(x)$ is given by the *transpose* of $\nabla g_i(x)$, which is denoted by $\nabla g_i(x)^\top$, where $i \in \{1,\ldots,m\}$. The *adjoint* of $Dg(x)$ is the linear mapping $Dg(x)^*[\cdot]\colon \mathbb{E} \to \mathbb{R}^n$ such that $\langle Dg(x)[h], w\rangle = \langle h, Dg(x)^*[w]\rangle$ holds for every $h \in \mathbb{R}^n$ and every $w \in \mathbb{E}$, hence

$$Dg(x)^*[w] = [\langle \partial_i g(x), w\rangle]_{i\in\{1,\ldots,n\}},$$

for every $w \in \mathbb{E}$ and, if $\mathbb{E} = \mathbb{R}^m$ then $Dg(x)^* = Dg(x)^\top$. Similarly, we define the action of the linear mapping $D^2 g(x)^*[\cdot]\colon \mathbb{E} \to \mathbb{R}^{n\times n}$ by

$$D^2 g(x)^*[w] \doteq [\langle \partial_i \partial_j g(x), w\rangle]_{i,j\in\{1,\ldots,n\}},$$

for every $w \in \mathbb{E}$.

The *orthogonal projection* of $w \in \mathbb{E}$ onto $\mathcal{K}$ is the point $\Pi_{\mathcal{K}}(w) \in \mathcal{K}$ such that

$$||w - \Pi_{\mathcal{K}}(w)|| = \min\{||w - z||\colon z \in \mathcal{K}\}.$$

Note that $\Pi_{\mathcal{K}}(w)$ is well-defined as a convex function of w since $\mathcal{K}$ is closed and convex. Also, a very useful fact is that every $w \in \mathbb{E}$ can be written as

$$w = \Pi_{\mathcal{K}}(w) - \Pi_{\mathcal{K}}(-w),$$

with $\langle \Pi_{\mathcal{K}}(w), \Pi_{\mathcal{K}}(-w)\rangle = 0$. This is commonly called the *Moreau's decomposition* of w and one of its many consequences is that $w \in \mathcal{K}$ if, and only if, $\Pi_{\mathcal{K}}(-w) = 0$. Hence, the function

$$\Phi(x) \doteq \frac{1}{2}\left(||h(x)||_2^2 + ||\Pi_{\mathcal{K}}(-g(x))||^2\right)$$

can be used as a measure of violation of the constraints of (NCP), that is, a measure of infeasibility. A result by Fitzpatrick and Phelps [34, Thm. 2.2] can be employed to derive an expression for the gradient of Φ at x:

Theorem 1 *For every $x \in \mathbb{R}^n$, we have*

$$\nabla\Phi(x) = Dh(x)^\top h(x) - Dg(x)^*[\Pi_{\mathcal{K}}(-g(x))].$$

Also, we observe that $\nabla\Phi$ is a Lipschitz function, but it is not differentiable everywhere. In our analyses, we make use of its second derivative, which must be taken in the nonsmooth sense.

2.1 Some elements of nonsmooth analysis

Let X and Y be finite-dimensional normed linear spaces over $\mathbb{R}$. Let $F\colon X \to Y$ be a locally Lipschitz function and denote the set in which it is differentiable by $\mathcal{D}(F)$. The so-called *B-subdifferential* of F at a point $x \in X$, is the set of all limiting derivatives of F at x, denoted by

$$\partial_B F(x) \doteq \left\{ V \in \mathcal{L}(X,Y)\colon \exists \{x^k\}_{k\in\mathbb{N}} \subset \mathcal{D}(F),\ x^k \to x,\ DF(x^k) \to V \right\},$$

where $\mathcal{L}(X,Y)$ denotes the set of all linear mappings from X to Y, and similarly to the previous section, $DF(x^k) \in \mathcal{L}(X,Y)$ denotes the first derivative of F at x^k, for every $x^k \in \mathcal{D}(F)$. Evidently, for every x, the set $\partial_B F(x)$ is compact, and it is a singleton when $x \in \mathcal{D}(F)$, but it is not convex in general. Then, we also define the *Clarke subdifferential* of F at x, denoted by $\partial F(x)$, as the *convex hull* of $\partial_B F(x)$, that is,

$$\partial F(x) \doteq \operatorname{Conv}(\partial_B F(x)).$$

In particular, when $X = \mathbb{R}^n$ and $Y = \mathbb{R}$, the *generalized Hessian* of F at x is defined as

$$\partial^2 F(x) \doteq \partial \nabla F(x),$$

which is the convex hull of the set of all limiting Hessian matrices of F at x. Following Hiriart-Urruty et al. [45], the second-order necessary optimality condition for unconstrained minimizers of F when it is differentiable, is the following:

Theorem 2 *If $x^\star$ is a local minimizer of a differentiable function $F\colon \mathbb{R}^n \to \mathbb{R}$ such that ∇F is locally Lipschitz, then $\nabla F(x^\star) = 0$, and for each $d \in \mathbb{R}^n$, there exists some $M \in \partial^2 F(x^\star)$ such that $d^\top M d \geqslant 0$. In other words,*

$$\limsup_{d\in\mathbb{R}^n} d^\top \partial^2 F(x^\star) d \geqslant 0.$$

We refer to [45, Thm. 3.1] for a proof. As observed by Hiriart-Urruty et al., it is not true that $d^\top M d \geqslant 0$ for all $M \in \partial^2 F(x^\star)$, in general, and not even this holds for some fixed M and all d. We employ this result to analyse the second derivative of Φ, which is a *nonsmooth-smooth* composition, so a chain rule is also required. There are several different extensions of the chain rule for subdifferentials, but the following result, by Páles and Zeidan [59], is enough for our purposes:

Theorem 3 *Fix some $x \in X$ and let $G\colon X \to Y$ and $F\colon Y \to Y$ be functions such that G is continuously differentiable at x, and F is Lipschitz in a neighborhood of $G(x)$. Then, we have*

$$\partial(F \circ G)(x) \subseteq \partial F(G(x)) \circ DG(x),$$

where $\partial F(G(x)) \circ DG(x) \doteq \{V \circ DG(x) \colon V \in \partial F(G(x))\}$.

Proof The result follows from [59, Thm. 5.1] since it was originally proved for Banach spaces that satisfy the Radon-Nikodým property, which holds for every reflexive space, and every finite-dimensional space with a norm is reflexive. □

Specificities about the subdifferential of the orthogonal projection onto the second-order and the semidefinite cone will be given in their respective sections.

2.2 Necessary optimality conditions and constraint qualifications

A *constraint qualification* (CQ) is any assumption over the constraints at a feasible point x, that implies that the feasible set is similar to its first-order approximation around x. For instance, one of the most relevant ones is *Robinson's CQ* [63], that holds at a feasible point x when $Dh(x)$ has full row rank and there exists some $d \in \mathbb{R}^n$ such that[1]

$$g(x) + Dg(x)[d] \in \operatorname{int}(\mathcal{K})$$
and
$$Dh(x)d = 0.$$

It is widely known that Robinson's CQ is a generalization of the classical *Mangasarian-Fromovitz constraint qualification* (MFCQ) from NLP. Such regularity condition allows one to study the optimality of a point in terms of the first-order approximation of the problem around it, that is, it is possible to prove that for every local solution $x^\star$ of (NCP) that satisfies Robinson's CQ, there exists some $\omega^\star \in \mathcal{K}$ and some $\mu^\star \in \mathbb{R}^p$ such that

$$\nabla_x L(x^\star, \omega^\star, \mu^\star) = 0, \tag{1}$$

and

$$\langle g(x^\star), \omega^\star \rangle = 0, \tag{2}$$

where

$$L(x, \omega, \mu) \doteq f(x) - \langle g(x), \omega \rangle + \langle h(x), \mu \rangle$$

is the *Lagrangian* function of (NCP) and

$$\nabla_x L(x, \omega, \mu) \doteq \nabla f(x) - Dg(x)^*[\omega] + Dh(x)^\top \mu$$

is the gradient of $L(x, \omega, \mu)$ with respect to x. Equations (1) and (2) compose the so-called *Karush-Kuhn-Tucker* (KKT) conditions and, in this context, $\omega^\star$

---

[1] We use this characterization of Robinson's CQ as a definition because $\mathcal{K}$ is assumed to be self-dual, and consequently, to have nonempty interior.

and $\mu^\star$ are *Lagrange multipliers* associated with $x^\star$. Points that satisfy the KKT conditions are often called *first-order stationary* or *KKT points*. Condition (2) is often called *complementarity*, and when additionally $-\omega^\star$ belongs to the relative interior of $T_{\mathcal{K}}(g(x^\star))^\circ$, where

$$T_{\mathcal{K}}(g(x^\star)) \doteq \left\{ d \in \mathbb{E}\colon \begin{array}{l} \exists \{d^k\}_{k\in\mathbb{N}} \to d, \; \exists \{\alpha^k\}_{k\in\mathbb{N}} \to 0, \\ \forall k \in \mathbb{N}, \alpha^k > 0, \; g(x^\star) + \alpha^k d^k \in \mathcal{K} \end{array} \right\}$$

is the (*Bouligand*) *tangent cone* to $\mathcal{K}$ at $g(x^\star)$, we say that *strict complementarity* holds at the pair $(x^\star, \omega^\star)$ [29, Def. 4.74]. A relevant implication of Robinson's CQ is the boundedness of the set $\mathcal{M}(x^\star)$ of all Lagrange multipliers associated with a local solution $x^\star$.

Second-order optimality conditions give extra information over the set of directions where first-order information is not meaningful. That is, we are interested in the set

$$C(x^\star) \doteq \{d \in \mathbb{R}^n \colon \langle \nabla f(x^\star), d\rangle = 0, \; Dh(x^\star)^\top d = 0, \; Dg(x^\star)[d] \in T_{\mathcal{K}}(g(x^\star))\},$$

which is the critical cone of (NCP) at $x^\star$. If Robinson's CQ holds at a local minimizer $x^\star$ of (NCP), then besides KKT, it also satisfies the *basic second-order necessary condition* (BSOC), that is, for every $d \in C(x^\star)$ there are Lagrange multipliers $\omega_d^\star \in \mathcal{K}$ and $\mu_d^\star \in \mathbb{R}^p$ such that (1), (2), and

$$d^\top (\nabla_x^2 L(x^\star, \omega_d^\star, \mu_d^\star) + \sigma(x^\star, \omega_d^\star))d \geqslant 0 \tag{3}$$

hold, where

$$\nabla_x^2 L(x, \omega, \mu) \doteq \nabla^2 f(x) - D^2 g(x)^*[\omega] + \sum_{i=1}^{p} \mu_i \nabla^2 h_i(x)$$

and $\sigma(x, \omega)$ is the so-called "*sigma-term*", as presented by Cominetti [30, Thm. 4.1]. In that paper, the author builds second-order conditions for (NCP) based on the *second-order tangent set* of $\mathcal{K}$ at $g(x)$ along $Dg(x)[d]$, that may be denoted by $T_{\mathcal{K}}^2(g(x), Dg(x)[d])$, and then establishes a "dual form" for it using the support function of $T_{\mathcal{K}}^2(g(x), Dg(x)[d])$, which is precisely the sigma-term. Hence, the sigma-term represents a possible curvature of $\mathcal{K}$ at $g(x)$, to some extent, and it can be proved that $\sigma(x, \omega) = 0$ when $\mathcal{K}$ is *polyhedral*, such as in NLP (for details, see [30]). In fact, the difficulty of second-order analysis in contexts more general than NLP lies almost entirely on the characterization of the sigma-term, which can be a very challenging task.

One of the major practical drawbacks of BSOC is that in order to verify whether it holds or not at a given point x, one must know the whole set $\mathcal{M}(x)$, which is not always possible. The stronger optimality condition where inequality (3) holds for every $d \in C(x^\star)$, for some pair of multipliers $(\omega^\star, \mu^\star)$ (not depending on d), which is sometimes called the *semi-strong necessary* *optimality condition*, does not present such a drawback. However, deciding the positivity of a matrix over a cone is an NP-hard class problem [58], and so is checking the semi-strong condition.

A more practical alternative to BSOC and the semi-strong condition is the so-called *weak second-order necessary condition* (WSOC), which is defined as follows:

Definition 1 Let $x^\star$ be a KKT point associated with some Lagrange multipliers $\omega^\star \in \mathcal{K}$ and $\mu^\star \in \mathbb{R}^p$. We say that WSOC holds at $x^\star$ when

$$d^\top(\nabla_x^2 L(x^\star, \omega^\star, \mu^\star) + \sigma(x^\star, \omega^\star))d \geqslant 0, \tag{4}$$

for every $d \in S(x^\star) \doteq \mathrm{lin}(C(x^\star))$, which is the largest subspace contained in $C(x^\star)$.

Note that in Definition 2.1 we only take directions in the subspace $S(x^\star)$, called the *critical subspace* of (NCP) at $x^\star$, which coincides with $C(x^\star)$ under strict complementarity[2]. At first sight, a second-order condition that only covers $S(x^\star)$ instead of the whole $C(x^\star)$ may seem disadvantageous in comparison with the semi-strong condition. In fact, the semi-strong condition implies WSOC. However, there are strong evidences that suggest that it is unlikely that BSOC or the semi-strong condition can be used to support the global convergence theory of any practical algorithm, unless $C(x^\star) = S(x^\star)$. In fact, for the particular case of NLP, Gould and Toint [40] presented a simple counterexample, with a quadratic objective function and a constraint of the form $x \geqslant 0$, for which a large class of barrier-type methods may produce an output sequence whose limit points fail to satisfy both BSOC and the semi-strong condition, even when every iterate of such sequence satisfies the second-order sufficient condition for its respective penalized problem. Later, Andreani and Secchin [18] made a small modification in Gould and Toint's counterexample to obtain the same conclusion for augmented Lagrangian-type algorithms. WSOC, on the other hand, is guaranteed to be fulfilled under weak assumptions for some variants of the two methods we mentioned above [2, 57], and also for a regularized SQP method for NLP [39]. The negative conclusions regarding BSOC and the semi-strong condition have led some authors to doubt the existence of an algorithm that could be associated with a second-order condition that takes the whole critical cone into consideration. Following this discussion, Andreani et al. [13] managed to characterize the weakest second-order constraint qualification that could guarantee the fulfilment of the semi-strong condition at the limit points of a large class of penalization-type algorithms that encompasses, for instance, all the aforementioned ones. However, such a constraint qualification was proven not to imply nor to be implied by LICQ [13, Ex. 4.5 and 4.6], and to be violated even for box constraints.

[2]We will give a short proof for the fact $C(x^\star) = \mathrm{lin}(C(x^\star))$ under strict complementarity, for completeness: Let $d \in C(x^\star)$ and suppose that there exists a Lagrange multiplier $-\omega^\star$ in the relative interior of $T_{\mathcal{K}}(g(x^\star))^\circ$. Note that $\langle \nabla f(x^\star), d\rangle = \langle \omega^\star, Dg(x^\star)[d]\rangle = 0$ by the KKT conditions. Hence, $-\omega^\star \in T_{\mathcal{K}}(g(x^\star))^\circ \cap \{Dg(x^\star)[d]\}^\perp$, which implies $T_{\mathcal{K}}(g(x^\star))^\circ \subseteq \{Dg(x^\star)[d]\}^\perp$ and, consequently, $\mathrm{span}(Dg(x^\star)[d]) \subseteq T_{\mathcal{K}}(g(x^\star))$. Then, $Dg(x^\star)[d] \in \mathrm{lin}(T_{\mathcal{K}}(g(x^\star)))$, so $d \in \mathrm{lin}(C(x^\star))$.

Despite the good algorithmic advantages of WSOC, Robinson's CQ alone is not enough to guarantee its fulfilment at local minimizers – see, for instance, the counterexample by Baccari [21, Sec. 3] or the discussion in [22]. Instead, the existing results on WSOC usually require a stronger CQ called *nondegeneracy* (or *transversality*), which holds at a feasible point x when

$$\mathbb{E} \times \mathbb{R}^p = (\text{lin}(T_{\mathcal{K}}(g(x))) + \text{Im}(Dg(x))) \times \text{Im}(Dh(x)). \tag{5}$$

It was translated from differential equations to optimization by Shapiro and Fan [66] and it is well-known that, for every nondegenerate solution $x^\star$ of (NCP), the set $\mathcal{M}(x^\star)$ is a singleton, what resembles the effects of LICQ in NLP. Thus, nondegeneracy is analogous to LICQ, in this sense.

Theorem 4 *If $x^\star$ is a local minimizer of* (NCP) *that satisfies nondegeneracy, then the KKT conditions hold at $x^\star$ for some Lagrange multipliers $\omega^\star \in \mathcal{K}$ and $\mu^\star \in \mathbb{R}^p$ and, moreover, WSOC holds with respect to these multipliers.*

Note that Theorem 2.4 is simply a rephrasing of the necessity of BSOC after assuming uniqueness of the Lagrange multiplier (nondegeneracy), but we stated it as it is for comparison purposes since our main results consist of proving of Theorem 2.4 under less demanding conditions.

In the context of NLP, Andreani, Martínez, and Schuverdt [16] were able to prove Theorem 2.4 replacing nondegeneracy (LICQ) with only MFCQ together with the so-called *weak constant rank* (WCR) property, which holds at a feasible point $x^\star$ when there exists a neighborhood $\mathcal{N}$ of $x^\star$ such that

$$\{\nabla g_i(x)\}_{i\colon\, g_i(x^\star)=0} \cup \{\nabla h_j(x)\}_{j\in\{1,\dots,p\}} \tag{6}$$

has the same rank for every $x \in \mathcal{N}$. It is worth mentioning that WCR is not a CQ on its own [16, Ex. 5.1] and that the joint condition "MFCQ+WCR" was proven to be strictly weaker than LICQ [16, Ex. 5.2]. Later, a simpler proof of this result was presented by Andreani et al. [13, Crlr. 4.3 and Thm. 4.1], using sequential optimality conditions. In the following sections, we generalize the WCR property and the result of [16] for NSOCP and NSDP, using an approach similar to [13].

As a matter of fact, Andreani, Echagüe, and Schuverdt [3] presented a result similar to Theorem 2.4, but under Janin's *constant rank constraint qualification* (CRCQ) [46], which is also weaker than LICQ and independent of "MFCQ+WCR". However, extending constant rank-type CQs to conic contexts is not easy, and finding an extension that preserves all of its interesting properties is even more difficult. In fact, there is a series of papers by Andreani et al. [8, 10–12] presenting distinct extensions of CRCQ for NSDP and NSOCP that suit distinct applications. For instance, [10, 12] deal with convergence of algorithms to first-order stationary points but no second-order properties were proven, whereas [11] presents a more geometric approach with some interesting theoretical properties but no application towards algorithms was provided.

We should mention, nevertheless, that the extension of WCR presented this paper is not a particular case of any of the conditions from the aforementioned papers.

3 Second-order cone programming

The standard NSOCP problem can be seen as a particular case of (NCP) where $\mathbb{E} = \mathbb{R}^m$ and $\mathcal{K} \doteq \mathcal{K}_1 \times \cdots \times \mathcal{K}_r$ is a Cartesian product of Lorentz cones, that is, $\mathcal{K}_i \doteq \mathbb{L}^{m_i}$ for all $i \in \{1, \ldots, r\}$, where $m_1 + \cdots + m_r = m$ and $\mathcal{K}_i \subset \mathbb{R}^{m_i}$. In this section, we consider $\mathbb{R}^m$ with its standard inner product and the Euclidean norm. The notation $w = (w_0, \overline{w})$ refers to a partition of $w \in \mathbb{R}^{m_i}$ where $w_0 \in \mathbb{R}$ is its first entry and $\overline{w} \in \mathbb{R}^{m_i - 1}$ is the subvector with the remaining entries. To make the NSOCP problem explicit, define $g \doteq (g_1, \ldots, g_r)$ with $g_i \colon \mathbb{R}^n \to \mathbb{R}^{m_i}$ for every $i \in \{1, \ldots, r\}$, and obtain

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & g_i(x) \in \mathcal{K}_i, \forall i \in \{1, \ldots, r\} \\ & h(x) = 0. \end{aligned} \tag{NSOCP}$$

As usual in the study of (NSOCP), given a feasible point x, we define the following sets of indices, which constitute a partition of $\{1, \ldots, r\}$:

$$\begin{aligned} I_0(x) &\doteq \{i \in \{1, \ldots, r\} \colon g_i(x) = 0\}, \\ I_B(x) &\doteq \{i \in \{1, \ldots, r\} \colon g_i(x) \in \text{bd}^+(\mathcal{K}_i)\}, \\ I_I(x) &\doteq \{i \in \{1, \ldots, r\} \colon g_i(x) \in \text{int}(\mathcal{K}_i)\}. \end{aligned} \tag{7}$$

Moreover, when we are dealing with a KKT point $x^\star$ associated with Lagrange multipliers $\omega^\star \in \mathcal{K}$ and $\mu^\star \in \mathbb{R}^p$, we consider the subset of $I_B(x^\star)$ given by

$$I_{BB}(x^\star, \omega^\star) \doteq \{i \in I_B(x^\star) \colon \omega_i^\star \in \text{bd}^+(\mathcal{K}_i)\}$$

and the critical subspace of (NSOCP) at $x^\star$ can be written in terms of such indices, as follows:

$$S(x^\star) = \left\{ d \in \mathbb{R}^n \colon \begin{array}{l} Dh(x^\star)d = 0; \quad Dg_i(x^\star)d = 0, i \in I_0(x^\star); \\ g_i(x^\star)^\top \Gamma_i Dg_i(x^\star)d = 0, i \in I_B(x^\star) \end{array} \right\}, \tag{8}$$

where

$$\Gamma_i \doteq \begin{bmatrix} 1 & 0^\top \\ 0 & -\mathbb{I}_{m_i - 1} \end{bmatrix} \in \mathbb{R}^{m_i \times m_i} \quad \text{for all } i \in \{1, \ldots, r\}. \tag{9}$$

The sigma-term at $x^\star$, when specialized to (NSOCP), can be written as

$$\sigma(x^\star, \omega^\star) = \sum_{i \in I_{BB}(x^\star, \omega^\star)} \sigma_i(x^\star, \omega^\star),$$

where

$$\sigma_i(x^\star, \omega^\star) = -\frac{[\omega_i^\star]_0}{[g_i(x^\star)]_0} Dg_i(x^\star)^\top \Gamma_i Dg_i(x^\star), \quad \text{for all } i \in I_{BB}(x^\star, \omega^\star). \tag{10}$$

We refer to [37] for details.

Also, the specialized characterization of the nondegeneracy condition in NSOCP, following Bonnans and Ramírez [28, Prop. 19], can be written as follows:

Proposition 5 *Let $x^\star \in \mathbb{R}^n$ be a feasible point of (NSOCP). The nondegeneracy condition holds at $x^\star$ if, and only if, the set*

$$\{\nabla h_i(x^\star)\}_{i\in\{1,\ldots,p\}} \bigcup \{\nabla g_{ij}(x^\star)\}_{\substack{i\in I_0(x^\star) \\ j\in\{1,\ldots,m_i\}}} \bigcup \{Dg_i(x^\star)^\top \Gamma_i \tilde{g}_i(x^\star)\}_{i\in I_B(x^\star)} \tag{11}$$

is linearly independent, where $\nabla g_{ij}(x)$ denotes the transpose of the j-th row of $Dg_i(x)$ and

$$\tilde{g}_i(x) \doteq (||\overline{g_i(x)}||_2, \overline{g_i(x)}). \tag{12}$$

In [4, Def. 3.3], the authors extend a sequential optimality condition called *Approximate-KKT* (AKKT) from NLP [7] to the NSOCP context. In short, AKKT is a punctual necessary optimality condition that also incorporates a bit of local information. That is, every point $x^\star$ that satisfies AKKT (though not necessarily KKT) is accompanied by a sequence $\{x^k\}_{k\in\mathbb{N}} \to x^\star$ such that each x^k approximately satisfies the KKT conditions with some approximate Lagrange multipliers ω^k and μ^k. Since our analyses are based on AKKT, we now recall its definition and some of its properties.

Definition 2 (AKKT for NSOCP) A feasible point $x^\star$ of (NSOCP) satisfies the AKKT condition when there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to x^\star$, $\{\omega^k\}_{k\in\mathbb{N}} \subset \mathcal{K}$, and $\{\mu^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p$ such that

$$\nabla_x L(x^k, \omega^k, \mu^k) \to 0 \tag{13}$$

and

$$\begin{aligned} &i \in I_I(x^\star) \Rightarrow \omega_i^k \to 0, \\ i \in I_B(x^\star) \Rightarrow \omega_i^k \to 0 \text{ or } \omega_i^k \in \text{bd}^+(\mathcal{K}_i) \text{ with } &\frac{\overline{\omega}_i^k}{||\overline{\omega}_i^k||_2} \to \frac{\overline{g_i(x^\star)}}{||\overline{g_i(x^\star)}||_2}. \end{aligned} \tag{14}$$

It was proved in [4, Thm. 3.1] that AKKT is indeed a genuine necessary optimality condition independently of CQs, in contrast with KKT. Also, their proof is constructive, which means it tells us how to obtain the sequences of perturbed KKT points and multipliers. Next, we state their result with a slightly different phrasing, in order to highlight such construction.

Theorem 6 *Let $x^\star$ be a local minimizer of (NSOCP). Then, for any given sequence $\{\rho_k\}_{k\in\mathbb{N}} \to +\infty$, there exists a sequence $\{x^k\}_{k\in\mathbb{N}} \to x^\star$, such that each x^k is a local minimizer of the regularized penalty function*

$$F_k(x) \doteq f(x) + \frac{1}{4}||x - x^\star||_2^4 + \frac{\rho_k}{2}\left(\sum_{i=1}^{r} ||\Pi_{\mathcal{K}_i}(-g_i(x))||_2^2 + ||h(x)||_2^2\right).$$

Also, the multiplier sequences given by $\omega_i^k \doteq \rho_k \Pi_{\mathcal{K}_i}(-g_i(x^k))$ for all $i \in \{1, \dots, r\}$ and $\mu^k \doteq \rho_k h(x^k)$ satisfy (13) and (14) together with $\{x^k\}_{k\in\mathbb{N}}$. Consequently, $x^\star$ satisfies AKKT.

A key property of AKKT, as stated in [4, Thm. 3.1], is that the sequences of multipliers from Definition 3.1 must be bounded when $x^\star$ satisfies Robinson's CQ. Hence, AKKT implies KKT under Robinson's CQ. Also, in the same paper the authors present a variant of the classical *Powell-Hestenes-Rockafellar* (PHR) Augmented Lagrangian method (see [44, 61, 64]) and prove that its output sequences can be fully described by AKKT.

3.1 Second-order optimality conditions

Here, we build second-order analysis for (NCP) primarily under Robinson's CQ instead of nondegeneracy and strict complementarity, but since Robinson's CQ alone is not enough to complete that task [21], we also introduce a generalized version of the WCR property.

Definition 3 (WCR for NSOCP) We say that the *weak constant rank* property is satisfied at a feasible point $x^\star$ of (NSOCP) if there exists a neighborhood $\mathcal{N}$ of $x^\star$ such that the set

$$\{\nabla h_i(x)\}_{i\in\{1,\dots,p\}} \bigcup \{\nabla g_{ij}(x)\}_{\substack{i\in I_0(x^\star)\\ j\in\{1,\dots,m_i\}}} \bigcup \{Dg_i(x)^\top \Gamma_i \tilde{g}_i(x)\}_{i\in I_B(x^\star)} \tag{15}$$

has the same rank, for all $x \in \mathcal{N}$.

In view of the characterization of nondegeneracy for NSOCP provided by Proposition 3.1, we see that nondegeneracy implies both Robinson's CQ and WCR in this context, just as in the NLP case. On the other hand, [16, Ex. 5.2] exhibits a point that satisfies MFCQ and WCR, but not LICQ. Hence, the joint condition "Robinson's CQ+WCR" is strictly weaker than nondegeneracy.

The main feature of the WCR property in NLP is its effect on the continuity of perturbations of the critical subspace around a feasible point $x^\star$. Next, we prove that this property is maintained in (NSOCP).

Lemma 1 *Let $x^\star \in \mathbb{R}^n$ be a feasible point of (NSOCP). Then, the WCR property holds at $x^\star$ if, and only if, the set-valued mapping $x \mapsto S(x, x^\star)$ is inner semicontinuous at $x^\star$, where*

$$S(x, x^\star) \doteq \left\{ d \in \mathbb{R}^n : \begin{array}{l} Dh(x)d = 0; \quad \forall i \in I_0(x^\star), Dg_i(x)d = 0; \\ \forall i \in I_B(x^\star), \tilde{g}_i(x)^\top \Gamma_i Dg_i(x)d = 0 \end{array} \right\}, \tag{16}$$

and $\tilde{g}_i$ is defined in (12)*.*

Proof Following the steps of the proof of [42, Prop. 2], we see that [20, Thm. 1.1.8] tells us that $x \mapsto S(x, x^\star)$ is inner semicontinuous at $x^\star$ if, and only if, the set-valued mapping $x \mapsto S(x, x^\star)^{\circ}$ is outer semicontinuous at $x^\star$, where $S(x, x^\star)^{\circ} \doteq -S(x, x^\star)^*$ denotes the *polar* of $S(x, x^\star)$. Since in this case we have

$$S(x, x^\star)^{\circ} = \left\{ \begin{array}{c} \sum_{i \in I_0(x^\star)} Dg_i(x)^\top a_i + \sum_{i=1}^{p} \nabla h_i(x) b_i + \\ + \sum_{i \in I_B(x^\star)} Dg_i(x)^\top \Gamma_i \tilde{g}_i(x) c_i \end{array} : \ a_i \in \mathbb{R}^{m_i}, \ b_i, c_i \in \mathbb{R} \right\},$$

the result follows directly from [31, Prop. 3.2.9]. □

As in NLP, the subspace $S(x, x^\star)$ may be called *perturbed critical subspace* of (NSOCP) at x, around $x^\star$. The last ingredient we need for the main theorem of this section is an explicit characterization of the subdifferential of the projection onto $\mathcal{K}_i$. In order to present that, for each $i \in \{1, \ldots, r\}$, let $M_i \colon \mathbb{R} \times \mathbb{R}^{m_i - 1} \to \mathbb{R}^{m_i \times m_i}$ be defined as

$$M_i(\xi, w) \doteq \frac{1}{2} \begin{bmatrix} 1 & w^\top \\ w & (1+\xi)\mathbb{I}_{m_i - 1} - \xi w w^\top \end{bmatrix}$$

and observe that the matrix $M_i(\xi, u)$ is symmetric positive semidefinite whenever $|\xi| \leq 1$ and $||w||_2 = 1$ [50, Lem. 2.8].

The following lemma, that can be found in [60, Lem. 14] and [43, Prop. 4.8], provides a description of the B-subdifferential of the projection onto $\mathcal{K}_i$, in terms of $M_i(\xi, w)$.

Lemma 2 *The B-subdifferential $\partial_B \Pi_{\mathcal{K}_i}(z)$ of the orthogonal projection onto $\mathcal{K}_i$ at $z \in \mathbb{R}^{m_i}$ is given as follows:*

(a) If $z \in \operatorname{int}(-\mathcal{K}_i)$, then $\partial_B \Pi_{\mathcal{K}_i}(z) = \{0\}$;
(b) If $z \in \operatorname{int}(\mathcal{K}_i)$, then $\partial_B \Pi_{\mathcal{K}_i}(z) = \{\mathbb{I}_{m_i}\}$;
(c) If $z \notin \mathcal{K}_i \cup (-\mathcal{K}_i)$, then $\partial_B \Pi_{\mathcal{K}_i}(z) = \left\{ M_i \left(\frac{z_0}{||\overline{z}||_2}, \frac{\overline{z}}{||\overline{z}||_2} \right) \right\}$;
(d) If $z \in \operatorname{bd}^+(\mathcal{K}_i)$, then $\partial_B \Pi_{\mathcal{K}_i}(z) = \left\{ \mathbb{I}_{m_i}, M_i \left(1, \frac{\overline{z}}{||\overline{z}||_2} \right) \right\}$;
(e) If $z \in \operatorname{bd}^+(-\mathcal{K}_i)$, then $\partial_B \Pi_{\mathcal{K}_i}(z) = \left\{ 0, M_i \left(-1, \frac{\overline{z}}{||\overline{z}||_2} \right) \right\}$;
(f) If $z = 0$, then $\partial_B \Pi_{\mathcal{K}_i}(z) = \{0, \mathbb{I}_{m_i}\} \cup \{M_i(\xi, w) \colon |\xi| \leq 1, ||w||_2 = 1\}$.

To the best of our knowledge, the first specialized study on second-order necessary conditions for (NSOCP) is credited to Bonnans and Ramírez [28, Thm. 30], where they assume nondegeneracy and the so-called *second-order* *growth condition* (or *uniform growth condition*). Fukuda and Fukushima [37, Thm. 4.5] also developed second-order conditions via squared slack variables, under nondegeneracy and strict complementarity. Our contribution to this discussion is to draw attention to the fact that the nondegeneracy assumption can be strictly weakened and that strict complementarity is not necessary when considering WSOC, which is also the main result of this section.

Theorem 7 *Let $x^\star$ be a local minimizer of (NSOCP) satisfying Robinson's CQ and the WCR property. Then, there are some Lagrange multipliers $\omega^\star \in \mathcal{K}$ and $\mu^\star \in \mathbb{R}^p$ such that the KKT conditions and WSOC hold.*

Proof Let $x^\star$ be a local minimizer of (NSOCP). Then, by Theorem 3.1, for any given $\{\rho_k\}_{k\in\mathbb{N}} \to +\infty$, there exists a sequence $\{x^k\}_{k\in\mathbb{N}} \to x^\star$ such that x^k is a local minimizer of $F_k(x)$ for each k, where

$$F_k(x) = f(x) + \frac{1}{4}||x - x^\star||_2^4 + \frac{\rho_k}{2}\left(\sum_{i=1}^{r} ||\Pi_{\mathcal{K}_i}(-g_i(x))||_2^2 + ||h(x)||_2^2\right).$$

From the local optimality of x^k, we obtain

$$\nabla F_k(x^k) = \nabla f(x^k) + ||x^k - x^\star||_2^2(x^k - x^\star) - \\ - \sum_{i=1}^{r} Dg_i(x^k)^\top \rho_k \Pi_{\mathcal{K}_i}(-g_i(x^k)) + \rho_k Dh(x^k)^\top h(x^k) = 0$$

and by Theorem 2.2, for every $d \in \mathbb{R}^n$ and every $i \in \{1, \dots, r\}$, there exists some $\chi_i^k \in \partial(\Pi_{\mathcal{K}_i} \circ -g_i)(x^k)$ such that $d^\top \nabla^2 F_k(x^k) d \geqslant 0$, where we denote by $\nabla^2 F_k(x^k)$ the element of the generalized Hessian of F_k at x^k that is defined in terms of χ_i^k, by an abuse of notation. That is,

$$\begin{aligned}\nabla^2 F_k(x^k) \doteq\ & \nabla^2 f(x^k) + ||x^k - x^\star||_2^2\, \mathbb{I}_n + 2(x^k - x^\star)(x^k - x^\star)^\top \\ & - \sum_{i=1}^{r}\left(\sum_{j=1}^{m_i} (\rho_k \Pi_{\mathcal{K}_i}(-g_i(x^k)))_j \nabla^2 g_{ij}(x^k) - \rho_k Dg_i(x^k)^\top \chi_i^k\right) \\ & + \sum_{j=1}^{p}\left(\rho_k h_j(x^k)\nabla^2 h_j(x^k) + \rho_k \nabla h_j(x^k)\nabla h_j(x^k)^\top\right).\end{aligned}$$

Following Theorem 3.1, we define $\omega_i^k \doteq \rho_k \Pi_{\mathcal{K}_i}(-g_i(x^k))$ for all $i \in \{1, \dots, r\}$, and $\mu^k \doteq \rho_k h(x^k)$ for every $k \in \mathbb{N}$, which satisfy (13). Also, it follows from Theorem 2.3 that there exists some $V_i^k \in \partial \Pi_{\mathcal{K}_i}(-g_i(x^k))$ such that

$$\chi_i^k = V_i^k \circ -Dg_i(x^k) = -V_i^k Dg_i(x^k),$$

where $\circ$ denotes a composition of linear operators. Hence, the expression $d^\top \nabla^2 F_k(x^k) d \geqslant 0$ can be rewritten as

$$d^\top\left(\nabla_x^2 L(x^k, \omega^k, \mu^k) + \rho_k \sum_{i=1}^{r} Dg_i(x^k)^\top V_i^k Dg_i(x^k) + \\ + \rho_k \sum_{j=1}^{p} \nabla h_j(x^k)\nabla h_j(x^k)^\top\right) d \geqslant -d^\top \Delta^k d, \tag{17}$$

where $\Delta^k \doteq ||x^k - x^\star||_2^2 \mathbb{I}_n + 2(x^k - x^\star)(x^k - x^\star)^\top \to 0$.

Under Robinson's CQ, the sequence $\{(\omega^k, \mu^k)\}_{k\in\mathbb{N}}$ is bounded (see the proof of [4, Thm. 3.3]). Then, for every limit point $(\omega^\star, \mu^\star)$ of $\{(\omega^k, \mu^k)\}_{k\in\mathbb{N}}$, note that $x^\star$ satisfies the KKT conditions. Without loss of generality, we assume $\{(\omega^k, \mu^k)\}_{k\in\mathbb{N}} \to (\omega^\star, \mu^\star)$. Now, from WCR and Lemma 3.1, we know that the mapping $x \mapsto S(x, x^\star)$ as in (16) is inner semicontinuous at $x^\star$, then for each $d \in S(x^\star)$ there exists a sequence $\{d^k\}_{k\in\mathbb{N}} \to d$ such that $d^k \in S(x^k, x^\star)$ for all $k \in \mathbb{N}$.

For each $i \in \{1, \dots, r\}$, define

$$u_i^k = \left([u_i^k]_0, \overline{u_i^k}\right) \doteq Dg_i(x^k)d^k.$$

Our next step is to compute $\rho_k (u_i^k)^\top V_i^k u_i^k$ and its limit points in three independent cases:

1. If $i \in I_I(x^\star)$, we have $g_i(x^k) \in \text{int}(\mathcal{K}_i)$ for all k sufficiently large. Then, from Lemma 3.2 item (a), $V_i^k = 0$ and $\rho_k (u_i^k)^\top V_i^k u_i^k = 0$ for such k;
2. If $i \in I_0(x^\star)$, recalling that $d^k \in S(x^k, x^\star)$, we have $u_i^k = 0$ for all $k \in \mathbb{N}$, which means $\rho_k (u_i^k)^\top V_i^k u_i^k = 0$ in this case as well;
3. If $i \in I_B(x^\star)$, the sequence $\{g_i(x^k)\}_{k\in\mathbb{N}}$ can be essentially split into three subsequences, which have distinct influences over $\rho_k (u_i^k)^\top V_i^k u_i^k$. Hence, they are separately analysed below, where N_1, N_2, and N_3 constitute a partition of $\mathbb{N}$:
   (i) $\{g_i(x^k)\}_{k\in N_1} \subset \text{int}(\mathcal{K}_i)$. Here, $\omega_i^k = 0$ for every $k \in N_1$. Also, by item (a) of Lemma 3.2, $V_i^k = 0$ and $\rho_k (u_i^k)^\top V_i^k u_i^k = 0$ for every $k \in N_1$;
   (ii) $\{g_i(x^k)\}_{k\in N_2} \subset \mathbb{R}^m \setminus (\mathcal{K}_i \cup -\mathcal{K}_i)$. From Lemma 3.2 item (c) we obtain

$$V_i^k = M_i \left(-\frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2}, -\frac{\overline{g_i(x^k)}}{||\overline{g_i(x^k)}||_2} \right)$$

   which can be explicitly written as

$$V_i^k = \frac{1}{2} \begin{bmatrix} 1 & -\dfrac{\overline{g_i(x^k)}^\top}{||\overline{g_i(x^k)}||_2} \\ -\dfrac{\overline{g_i(x^k)}}{||\overline{g_i(x^k)}||_2} & Z^k \end{bmatrix},$$

   where

$$Z^k \doteq \left(1 - \frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2}\right)\mathbb{I}_{m_i-1} + \left(\frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2}\right) \frac{\overline{g_i(x^k)}\;\overline{g_i(x^k)}^\top}{||\overline{g_i(x^k)}||_2^2},$$

and it is elementary to see that

$$(u_i^k)^\top V_i^k u_i^k = \frac{1}{2}\left([u_i^k]_0^2 - \frac{2[u_i^k]_0 \overline{g_i(x^k)}^\top \overline{u_i^k}}{||\overline{g_i(x^k)}||_2} + \left(1 - \frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2}\right)||\overline{u_i^k}||_2^2 + \right.$$
$$\left. + \frac{[g_i(x^k)]_0 (\overline{g_i(x^k)}^\top \overline{u_i^k})^2}{||\overline{g_i(x^k)}||_2^3}\right).$$

Also, since $d^k \in S(x^k, x^\star)$ and $i \in I_B(x^\star)$ we have $\tilde{g}_i(x^k)^\top \Gamma_i u_i^k = 0$, or equivalently, $\overline{g_i(x^k)}^\top \overline{u_i^k} = ||\overline{g_i(x^k)}||_2 [u_i^k]_0$. Replacing this in the above expression, we obtain:

$$\begin{aligned}
&(u_i^k)^\top V_i^k u_i^k \\
&= \frac{1}{2}\left([u_i^k]_0^2 - 2[u_i^k]_0^2 + \left(1 - \frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2}\right)||\overline{u_i^k}||_2^2 + \frac{[u_i^k]_0^2 [g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2}\right) \\
&= -\frac{1}{2}\left(1 - \frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2}\right)\left([u_i^k]_0^2 - ||\overline{u_i^k}||_2^2\right). \qquad (18)
\end{aligned}$$

It follows from our specific choice of approximate multiplier that

$$\omega_i^k = \rho_k \Pi_{\mathcal{K}_i}(-g_i(x^k)) = \rho_k \frac{||\overline{g_i(x^k)}||_2 - [g_i(x^k)]_0}{2}\left(1, -\frac{\overline{g_i(x^k)}}{||\overline{g_i(x^k)}||_2}\right).$$

Hence, we have

$$\frac{[\omega_i^k]_0}{||\overline{g_i(x^k)}||_2} = \frac{\rho_k}{2}\left(1 - \frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2}\right)$$

and from (18), we obtain

$$\rho_k (u_i^k)^\top V_i^k u_i^k = -\frac{[\omega_i^k]_0}{||\overline{g_i(x^k)}||_2}\left([u_i^k]_0^2 - ||\overline{u_i^k}||_2^2\right) = -\frac{[\omega_i^k]_0}{||\overline{g_i(x^k)}||_2}(u_i^k)^\top \Gamma_i u_i^k. \qquad (19)$$

(iii) $\{g_i(x^k)\}_{k \in N_3} \subset \mathrm{bd}^+(\mathcal{K}_i)$. For every $k \in N_3$, we have $\omega_i^k = 0$. Also, for all such k, Lemma 3.2 item (e) implies

$$V_i^k = \tau M_i \left(-1, -\frac{\overline{g_i(x^k)}}{||\overline{g_i(x^k)}||_2}\right),$$

for some $\tau \in [0,1]$. Then, note that

$$M_i\left(-1, -\frac{\overline{g_i(x^k)}}{||\overline{g_i(x^k)}||_2}\right) = M_i\left(-\frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2}, -\frac{\overline{g_i(x^k)}}{||\overline{g_i(x^k)}||_2}\right),$$

which means that simply multiplying (19) by τ is enough to obtain $\rho_k(u_i^k)^\top V_i^k u_i^k = 0$ as well, since $[\omega_i^k]_0 = 0$.

Considering exclusively any infinite subsequence indexed by N_1, N_2, or N_3, based on our previous analyses we observe that, for k sufficiently large, (17) implies

$$\liminf_{k\to\infty} (d^k)^\top\left(\nabla_x^2 L(x^k,\omega^k,\mu^k) - \sum_{i\in I_B(x^\star)} \frac{[\omega_i^k]_0}{||\overline{g_i(x^k)}||_2} Dg_i(x^k)^\top \Gamma_i Dg_i(x^k)\right)d^k \geq 0.$$

Since $x^k \to x^\star$, $\omega^k \to \omega^\star$, $\mu^k \to \mu^\star$, $d^k \to d \in S(x^\star)$ and $||\overline{g_i(x^k)}||_2 \to ||\overline{g_i(x^\star)}||_2 = [g_i(x^\star)]_0$ when $i \in I_B(x^\star)$, we conclude that

$$d^\top\left(\nabla_x^2 L(x^\star,\omega^\star,\mu^\star) + \sum_{i\in I_{BB}(x^\star)} \sigma_i(x^\star,\omega^\star)\right)d \geq 0 \quad \text{for all } d \in S(x^\star),$$

where $\sigma_i(x^\star,\omega^\star)$ is defined as in (10). Therefore, $x^\star$ satisfies WSOC. □

Note that Theorem 3.2 contains a proof for the fact that every feasible limit point of any sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by an *external penalty method* must satisfy WSOC if it satisfies Robinson's CQ and WCR. Moreover, with minor adaptations, it is possible to prove that the same holds for every feasible limit point of a modified extension of the *augmented Lagrangian method* for NLP considered in [25]. And finally, we remark that if $m_1 = \ldots = m_r = 1$, that is, if (NSOCP) reduces to a NLP problem, then Theorem 3.2 recovers a result by Andreani et al. [13, Crlr. 4.2 and Crlr. 4.3] with an alternative proof.

4 Semidefinite programming

In this section, $\mathbb{S}^m$ is the linear space of all $m \times m$ symmetric matrices with real entries, equipped with the (*Frobenius*) inner product given by $\langle M, N\rangle \doteq \operatorname{trace}(MN)$ and the norm $||M||_F \doteq \sqrt{\langle M, M\rangle}$, for every $M, N \in \mathbb{S}^m$. We define $M \odot N$ as the (*Hadamard*) entry-wise product between M and N. Also, the cone of all symmetric positive semidefinite matrices is denoted by $\mathbb{S}^m_+$ and $\succeq$ is the partial order induced by it, that is, $M \succeq N$ if, and only if, $M - N \in \mathbb{S}^m_+$. Similarly, $M \succ N$ when $M - N \in \operatorname{int}(\mathbb{S}^m_+)$.

Recall that every $M \in \mathbb{S}^m$ has a *spectral decomposition* in the form $M = U\Lambda U^\top$, where U is an orthogonal matrix whose columns are eigenvectors of M and $\Lambda = \operatorname{Diag}(\lambda_1^U(M), \ldots, \lambda_m^U(M))$ is a diagonal matrix whose entries are the eigenvalues of M respective to the columns of U. It is well-known that the orthogonal projection of M onto $\mathbb{S}^m_+$ under $||\cdot||_F$ is given by

$$\Pi_{\mathbb{S}^m_+}(M) \doteq U\operatorname{Diag}(\max\{0, \lambda_1^U(M)\}, \ldots, \max\{0, \lambda_m^U(M)\})U^\top.$$

The specialization of (NCP) to an NSDP is obtained by setting $\mathbb{E} = \mathbb{S}^m$ and $\mathcal{K} = \mathbb{S}^m_+$, and it is often stated in the form

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & g(x) \succeq 0, \\ & h(x) = 0. \end{aligned} \tag{NSDP}$$

Here, for simplicity, we consider a single conic constraint since it is enough to cover all major aspects of the problem and the notation would be unnecessarily heavy otherwise. Similarly to the NSOCP case, several concepts of general conic programming can be specialized and explicitly characterized here, for example, the tangent cone to $\mathbb{S}^m_+$ at some $M \in \mathbb{S}^m_+$ can be written as

$$T_{\mathbb{S}^m_+}(M) = \{E \in \mathbb{S}^m \colon V^\top E V \in \mathbb{S}^{|\beta|}_+\},$$

where $V \in \mathbb{R}^{m\times|\beta|}$ is any matrix with orthonormal columns that form a basis for $\mathrm{Ker}(M)$ and $|\beta|$ is its dimension (see [65] for details).

Let x be a feasible point of (NSDP). In this section we always consider spectral decompositions of $g(x)$ that keep zero and nonzero eigenvalues separated, for example,

$$g(x) = U \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} U^\top,$$

where $\mathbb{S}^\alpha \ni \Lambda \succ 0$ and $\alpha \doteq \alpha(x)$ is the set of indices of the positive eigenvalues of $g(x)$. Let $\beta \doteq \beta(x)$ be the set of indices of the null eigenvalues of $g(x)$ and partition the columns U with respect to α and β as follows: $U \doteq [U_\alpha, U_\beta]$. For every $d \in \mathbb{R}^n$, define $D\tilde{g}(x)[d] \doteq U^\top Dg(x)[d]U$ as a reverse conjugation of $Dg(x)[d]$ around $g(x)$ and set

$$D\tilde{g}(x)[d] = \begin{bmatrix} (D\tilde{g}(x)[d])_{\alpha\alpha} & (D\tilde{g}(x)[d])_{\alpha\beta} \\ (D\tilde{g}(x)[d])^\top_{\alpha\beta} & (D\tilde{g}(x)[d])_{\beta\beta} \end{bmatrix}$$

as a partition of $D\tilde{g}(x)[d]$ with respect to α and β. Note that since U is an orthogonal matrix, the inner product is invariant to reverse conjugation in terms of U, that is,

$$\langle A, B\rangle = \mathrm{trace}(AB) = \mathrm{trace}(UU^\top AUU^\top B) = \mathrm{trace}(U^\top AUU^\top BU) = \langle \tilde{A}, \tilde{B}\rangle$$

for all $A, B \in \mathbb{S}^m$.

The critical cone of (NSDP) at a feasible point $x^\star$ is given by

$$C(x^\star) = \{d \in \mathbb{R}^n \colon \nabla f(x^\star)^\top d = 0, Dh(x^\star)d = 0, (D\tilde{g}(x^\star)[d])_{\beta\beta} \succeq 0\}.$$

Under Robinson's CQ, if $x^\star$ is a KKT point associated with some Lagrange multipliers $\omega^\star \in \mathbb{S}^m_+$ and $\mu^\star \in \mathbb{R}^p$ that satisfy strict complementarity, then the critical cone becomes equal to the critical subspace

$$S(x^\star) = \{d \in \mathbb{R}^n \colon Dh(x^\star)d = 0, (D\tilde{g}(x^\star)[d])_{\beta\beta} = 0\}$$

since $\tilde{\omega}^\star_{\alpha\alpha} = 0$ and $\nabla f(x^\star)^\top d = \langle Dg(x^\star)[d], \omega^\star \rangle - \langle Dh(x^\star)d, \mu^\star \rangle = 0$ for every $d \in \mathbb{R}^n$.

In [15], Andreani, Haeser, and Viana proposed an extension of the AKKT condition from NLP to (NSDP) as well. We state it as follows:

Definition 4 (AKKT for NSDP) A feasible point $x^\star$ of (NSDP) satisfies the AKKT condition when there are sequences $\{x^k\}_{k\in\mathbb{N}} \to x^\star$, $\{\omega^k\}_{k\in\mathbb{N}} \subset \mathbb{S}^m_+$, and $\{\mu^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p$ such that

$$\nabla_x L(x^k, \omega^k, \mu^k) \to 0 \tag{20}$$

and

$$\lambda_i^U(g(x^\star)) > 0 \Rightarrow \lambda_i^{S^k}(\omega^k) = 0, \tag{21}$$

for sufficiently large k, where U diagonalizes $g(x^\star)$, S^k diagonalizes ω^k for each k, and $S^k \to U$.

If $x^\star$ is a KKT point of (NSDP) associated with multipliers $\omega^\star \in \mathbb{S}^m_+$ and $\mu^\star \in \mathbb{R}^p$, note that the complementarity condition $\langle g(x^\star), \omega^\star \rangle = 0$ holds for $g(x^\star), \omega^\star \in \mathbb{S}^m_+$ if, and only if, $g(x^\star)\omega^\star = 0$, then it is elementary to check that $g(x^\star)$ and $\omega^\star$ must be simultaneously diagonalizable (i.e. they commute) in this case. In light of this, note that Definition 4.1 relaxes the commutativity between $g(x^\star)$ and $\omega^\star$ by requiring $S^k \to U$.

Also in [15], the authors prove that AKKT as in Definition 4.1 is a necessary optimality condition, independently of the fulfilment of constraint qualifications. We state it below in the same form as Theorem 3.1, with some emphasis on how the sequences that compose it are generated.

Theorem 8 *Let $x^\star$ be a local minimizer of (NSDP). Then, for any sequence $\{\rho_k\}_{k\in\mathbb{N}} \to +\infty$, there exists some $\{x^k\}_{k\in\mathbb{N}} \to x^\star$, such that for every k, x^k is a local minimizer of the regularized penalty function*

$$F_k(x) \doteq f(x) + \frac{1}{4}||x - x^\star||_2^4 + \frac{\rho_k}{2}\left(||\Pi_{\mathbb{S}^m_+}(-g(x))||_F^2 + ||h(x)||_2^2\right).$$

Also, the multiplier sequences given by $\omega^k \doteq \rho_k \Pi_{\mathbb{S}^m_+}(-g(x^k))$ and $\mu^k \doteq \rho_k h(x^k)$ satisfy (20) and (21) with $\{x^k\}_{k\in\mathbb{N}}$. Consequently, since ω^k and $g(x^k)$ are simultaneously diagonalizable in this case for every $k \in \mathbb{N}$, $x^\star$ satisfies AKKT.

Under Robinson's CQ, the sequences $\{\omega^k\}_{k\in\mathbb{N}}$ and $\{\mu^k\}_{k\in\mathbb{N}}$ are bounded, and also all limit points of those sequences are Lagrange multipliers associated with $x^\star$ [15, Thm. 6.1]. That is, AKKT implies KKT in this case. An augmented Lagrangian algorithm is also presented in [15] for NSDP, whose global convergence theory is built around AKKT. Such results were sharpened in [5] and further extended in [6], for the general (NCP).

4.1 Second-order analysis

As mentioned before, there are many different works that deal with a specialized second-order analysis for (NSDP), which mainly differ in the assumptions required for it and the techniques employed to characterize the sigma-term. As far as we know, the first work on this topic is due to Shapiro [65, Sec. 4], who obtained the very useful and practical expression

$$\sigma(x^\star, \omega^\star) = [2\langle \omega^\star, \partial_i g(x^\star) g(x^\star)^\dagger \partial_j g(x^\star)\rangle]_{i,j\in\{1,\ldots,n\}}, \tag{22}$$

where $g(x^\star)^\dagger$ is the *Moore-Penrose pseudoinverse* of $g(x^\star)$. Shapiro's idea was to write the semidefinite cone using the second-order directional derivative of the least eigenvalue function $\lambda_{\min}$, as follows:

$$\mathbb{S}^m_+ = \{M \in \mathbb{S}^m : \lambda_{\min}(M) \geqslant 0\},$$

and the expression of the sigma-term comes from the expression of the second-order directional derivative of $\lambda_{\min}$. Moreover, his second-order analysis was based on the uniqueness of the Lagrange multiplier, via nondegeneracy, and strict complementarity (Theorem 2.4). Then, Jarre [48, Thm. 2] presented another way of achieving Shapiro's characterization of the sigma-term, and consequently an alternative proof for Theorem 2.4, using a locally equivalent formulation of (NSDP) based on the Schur complement of $g(x^\star)_{\alpha\alpha}$, which turned out to be a more elementary proof. Later, Lourenço, Fukuda, and Fukushima [56, Props. 5.1 and 5.2], studied a characterization of the semidefinite cone with squared slack variables

$$\mathbb{S}^m_+ = \{M \in \mathbb{S}^m : \exists Z \in \mathbb{S}^m,\ M - ZZ = 0\},$$

which induces a reformulation of (NSDP) as a NLP problem. Then, the authors related the classical second-order conditions for NLP with the second-order conditions for (NSDP) (with the curvature term), under the same hypotheses as Shapiro and Jarre. Forsgren [35, Thm. 2], on the other hand, proved that strict complementarity was not needed for WSOC when assuming a different notion of regularity that treats structural sparsity and also uses Schur complements. In this section, we use the characterization

$$\mathbb{S}^m_+ = \{M \in \mathbb{S}^m : ||\Pi_{\mathbb{S}^m_+}(-M)||^2 = 0\},$$

and the generalized derivative of the orthogonal projection, to obtain second-order results that do not require uniqueness of multipliers, nor strict complementarity.

Our approach is based on extracting second-order information from AKKT and Theorem 4.1 and we do this in a similar manner of the previous section, which means we begin by exhibiting a characterization of the derivative of

$\Pi_{\mathbb{S}^m_+}$, then we extend the WCR condition from NLP to NSDP and, at last, we compute the sigma-term using the second-order (generalized) derivative of F_k.

Based on the works of Bonnans at al. [26] and Pang et al. [60], Sun [67] characterized the B-subdifferential of the projection onto the semidefinite cone. To make a proper reference, we define the following matrix:

$$\mathcal{B}(\lambda^U(M)) \doteq \left[\frac{\max\{\lambda_i^U(M), 0\} + \max\{\lambda_j^U(M), 0\}}{|\lambda_i^U(M)| + |\lambda_j^U(M)|}\right]_{i,j\in\{1,\ldots,m\}},$$

where $0/0$ is set as 1 and U is an orthogonal matrix that diagonalizes M. Next, we make a slightly adapted transcription of a proposition by Qi [62, Prop. 2.5] summarizing Sun's result:

Proposition 9 *Suppose that $M = U\Lambda U^\top$ is the spectral decomposition of $M \in \mathbb{S}^m$ and let α, β and γ be the sets of indices of the positive, zero and negative eigenvalues of M, respectively. Without loss of generality, assume those three blocks are separated and that $U \doteq [U_\alpha, U_\beta, U_\gamma]$. Then, for any $V \in \partial_B \Pi_{\mathbb{S}^m_+}(-M)$ there exists a $V_{|\beta|} \in \partial_B \Pi_{\mathbb{S}^{|\beta|}_+}(0)$ such that*

$$V[H] = U \begin{bmatrix} 0 & 0 & \tilde{H}_{\alpha\gamma} \odot \mathcal{B}(\lambda^U(M))_{\alpha\gamma} \\ 0 & V_{|\beta|}[\tilde{H}_{\beta\beta}] & \tilde{H}_{\beta\gamma} \\ \tilde{H}_{\alpha\gamma}^\top \odot \mathcal{B}(\lambda^U(M))_{\alpha\gamma}^\top & \tilde{H}_{\beta\gamma}^\top & \tilde{H}_{\gamma\gamma} \end{bmatrix} U^\top \qquad (23)$$

for every $H \in \mathbb{S}^m$, where $\tilde{H} \doteq U^\top H U$. Conversely, for every $V_{|\beta|} \in \partial_B \Pi_{\mathbb{S}^{|\beta|}_+}(0)$, there exists some $V \in \partial_B \Pi_{\mathbb{S}^m_+}(-M)$ such that (23) holds.

Even though we assume the eigenvalues are separated by sign, the ordering inside each partition is not relevant. Note that Proposition 4.1 is still true if we replace the B-subdifferential for the Clarke subdifferential.

Corollary 1 *Under the hypotheses of Proposition 4.1, for any $V \in \partial \Pi_{\mathbb{S}^m_+}(-M)$ there exists a $V_{|\beta|} \in \partial \Pi_{\mathbb{S}^{|\beta|}_+}(0)$ such that (23) holds. Conversely, for every $V_{|\beta|} \in \partial \Pi_{\mathbb{S}^{|\beta|}_+}(0)$, there exists some $V \in \partial \Pi_{\mathbb{S}^m_+}(-M)$ such that (23) holds.*

Proof Let $V \in \partial \Pi_{\mathbb{S}^m_+}(-M)$. Then, $V = \sum_{i=1}^s a_i V^i$, for some $s \in \mathbb{N}$, some $a_i \geqslant 0$, and some $V^i \in \partial_B \Pi_{\mathbb{S}^m_+}(-M)$, $i \in \{1, \ldots, s\}$, with $\sum_{i=1}^s a_i = 1$. This means there are $V^i_{|\beta|} \in \partial_B \Pi_{\mathbb{S}^{|\beta|}_+}(0)$, $i \in \{1, \ldots, s\}$, such that (23) holds. Hence, for every $H \in \mathbb{S}^m$, we have $V[H] = \sum_{i=1}^s a_i V^i[H]$ and the proof is over, because $\sum_{i=1}^s a_i V^i_{|\beta|}[\tilde{H}_{\beta\beta}] \in \partial \Pi_{\mathbb{S}^{|\beta|}_+}(0)$. The converse is analogous. □

In order to study perturbations of the critical subspace around a given point $x^\star$ via WCR, let $\overline{\alpha}$ and $\overline{\beta}$ represent the indices of positive and zero eigenvalues of $g(x^\star)$, respectively, regarding the decomposition

$$g(x^\star) = U\text{Diag}(\lambda^U(g(x^\star)))U^\top, \tag{24}$$

where $U \doteq [U_{\overline{\alpha}}, U_{\overline{\beta}}]$ is a matrix whose columns are eigenvectors of $g(x^\star)$ and, in particular, the columns of $U_{\overline{\beta}}$ form a basis for $\text{Ker}(g(x^\star))$. Moreover, we will use a construction from Bonnans and Shapiro's book [29, Ex. 3.98 and Ex. 3.140], which will be stated as a lemma below:

Lemma 3 *Let $M^\star \in \mathbb{S}^m_+$, set $\overline{\beta}$ as the indices of zero eigenvalues of M, and let $U_{\overline{\beta}}$ be a matrix with orthonormal columns that span $\text{Ker}(M^\star)$. There exists a neighborhood $\mathcal{N}$ of $M^\star$ and an analytic matrix function $\mathcal{U}_{\overline{\beta}}\colon \mathcal{N} \to \mathbb{R}^{m\times|\overline{\beta}|}$ such that $\mathcal{U}_{\overline{\beta}}(M^\star) = U_{\overline{\beta}}$ and, for every $M \in \mathcal{N}$, the columns of $\mathcal{U}_{\overline{\beta}}(M)$ form an orthonormal basis for the space spanned by the eigenvectors associated with the $|\overline{\beta}|$ smallest eigenvalues of M.*

This construction allows us to approximate the critical subspace around $x^\star$. Indeed, let $\mathcal{N}$ be the neighborhood of $g(x^\star)$ and $\mathcal{U}_{\overline{\beta}}\colon \mathcal{N} \to \mathbb{R}^{m\times|\overline{\beta}|}$ be the function given by Lemma 4.1 such that $\mathcal{U}_{\overline{\beta}}(g(x^\star)) = U_{\overline{\beta}}$. Then, for every x close enough to $x^\star$ so that $g(x) \in \mathcal{N}$, consider the following set:

$$S(x, x^\star) = \{d \in \mathbb{R}^n \colon \mathcal{U}_{\overline{\beta}}(g(x))^\top Dg(x)[d]\mathcal{U}_{\overline{\beta}}(g(x)) = 0, Dh(x)d = 0\},$$

which will be called *perturbed critical subspace* at x, centered at $x^\star$.

Extending WCR from NLP to (NSDP) is not a trivial task because the notion of "rank" of the three-dimensional tensor $Dg(x)$ may have multiple meanings. Fortunately, there is a useful characterization of nondegeneracy by Shapiro and Fan [66], which provides some insight on how to talk about rank in NSDP. Next, we make a transcription of this result as stated in [65, Prop. 6], for completeness.

Proposition 10 *Suppose that the dimension of $\text{Ker}(g(x^\star))$ is $|\overline{\beta}|$ and let $U_{\overline{\beta}} \doteq [u_1, \ldots, u_{|\overline{\beta}|}] \in \mathbb{R}^{m\times|\overline{\beta}|}$ be a matrix whose columns form a basis for $\text{Ker}(g(x^\star))$. Then, nondegeneracy holds at a feasible point $x^\star$ of (NSDP) if, and only if, the set of n-dimensional vectors $\{v_{ij} \colon 1 \leqslant i \leqslant j \leqslant |\overline{\beta}|\} \cup \{\nabla h_i(x^\star) \colon i \in \{1, \ldots, p\}\}$ is linearly independent, where $v_{ij} \doteq [u_i^\top \partial_\ell g(x^\star) u_j]_{\ell\in\{1,\ldots,n\}}$.*

Inspired by this characterization, we define WCR as follows:

Definition 5 (WCR for NSDP) Let $x^\star$ be a feasible point of (NSDP) and let (24) be a spectral decomposition of $g(x^\star)$. We say that $x^\star$ satisfies the weak constant rank (WCR) property when there exists a neighborhood $\mathcal{N}$ of $x^\star$ such that the set

$$\{\overline{v}_{ij}(x)\colon 1 \leqslant i \leqslant j \leqslant |\overline{\beta}|\} \cup \{\nabla h_i(x)\colon i \in \{1,\dots,p\}\}$$

has the same rank for every $x \in \mathcal{N}$, where

$$\overline{v}_{ij}(x) \doteq [\overline{u}_i(x)^\top \partial_\ell g(x) \overline{u}_j(x)]_{\ell\in\{1,\dots,n\}}$$

and $\overline{u}_1(x),\dots,\overline{u}_{|\overline{\beta}|}(x) \in \mathbb{R}^m$ denote the columns of $\mathcal{U}_{\overline{\beta}}(g(x))$.

Also, in the following lemma we prove that WCR as in Definition 4.2 is equivalent to the inner semicontinuity of the mapping $x \mapsto S(x, x^\star)$ at $x^\star$.

Lemma 4 *A feasible point $x^\star$ satisfies WCR if, and only if, the set-valued mapping $x \mapsto S(x, x^\star)$ is inner semicontinuous at $x^\star$.*

Proof First, we shall prove that, for every $x \in \mathbb{R}^n$,

$$S(x, x^\star) = \left\{ d \in \mathbb{R}^n \colon \overline{v}_{ij}(x)^\top d = 0,\ 1 \leqslant i \leqslant j \leqslant |\overline{\beta}|;\ \ Dh(x)d = 0 \right\}. \qquad (25)$$

Note that for each $\ell \in \{1,\dots,n\}$ we have

$$\mathcal{U}_{\overline{\beta}}(g(x^\star))^\top \partial_\ell g(x) \mathcal{U}_{\overline{\beta}}(g(x^\star)) = \begin{bmatrix} \overline{u}_1(x)^\top \partial_\ell g(x) \overline{u}_1(x) & \cdots & \overline{u}_1(x)^\top \partial_\ell g(x) \overline{u}_{|\overline{\beta}|}(x) \\ \vdots & \ddots & \vdots \\ \overline{u}_{|\overline{\beta}|}(x)^\top \partial_\ell g(x) \overline{u}_1(x) & \cdots & \overline{u}_{|\overline{\beta}|}(x)^\top \partial_\ell g(x) \overline{u}_{|\overline{\beta}|}(x) \end{bmatrix},$$

hence, considering that $\overline{u}_j(x)^\top \partial_\ell g(x) \overline{u}_i(x) = \overline{u}_i(x)^\top \partial_\ell g(x) \overline{u}_j(x) = (\overline{v}_{ij}(x))_\ell$ for every i, j, ℓ, we obtain

$$\mathcal{U}_{\overline{\beta}}(g(x^\star))^\top Dg(x)[d] \mathcal{U}_{\overline{\beta}}(g(x^\star)) = \begin{bmatrix} \overline{v}_{11}(x)^\top d & \cdots & \overline{v}_{1|\overline{\beta}|}(x)^\top d \\ \vdots & \ddots & \vdots \\ \overline{v}_{1|\overline{\beta}|}(x)^\top d & \cdots & \overline{v}_{|\overline{\beta}||\overline{\beta}|}(x)^\top d \end{bmatrix},$$

whence follows (25).

Now, similarly to the NSOCP case, since $\partial_\ell g$ and $\mathcal{U}_{\overline{\beta}}$ are continuous, $\overline{v}_{ij}$ is also continuous, then a result from Facchinei and Pang [31, Prop. 3.2.9] tells us that WCR is equivalent to the outer semicontinuity of the mapping $x \mapsto S(x, x^\star)^\circ$ at $x^\star$, where

$$S(x, x^\star)^\circ = \left\{ \sum_{1 \leqslant i \leqslant j \leqslant |\overline{\beta}|} a_{ij} \overline{v}_{ij}(x) + \sum_{i=1}^{p} b_i \nabla h_i(x) \ :\ a_{ij} \in \mathbb{R}, b_i \in \mathbb{R} \right\}$$

using the characterization in (25). Then, the desired result follows from [20, Thm. 1.1.8], which states that the inner semicontinuity of a set-valued mapping at a given point is equivalent to the outer semicontinuity of its polar at that point. □

Clearly, WCR as in Definition 4.2 is implied by nondegeneracy, in view of Proposition 4.2. Also, let us assume for a moment that $g(x)$ is a structurally diagonal matrix constraint whose diagonal elements are denoted by $g_1(x),\dots,g_m(x)$, and let $x^\star$ be such that $g(x) \in \mathbb{S}^m_+$. Without loss of generality, let us assume that $g_1(x^\star) > \ldots > g_{|\overline{\alpha}|}(x^\star) > g_{|\overline{\alpha}|+1}(x^\star) = \ldots = g_{|\overline{\alpha}|+|\overline{\beta}|}(x^\star) = 0$. Then, we can take

$$\mathcal{U}_{\overline{\beta}}(g(x)) \doteq \begin{bmatrix} 0 \\ \mathbb{I}_{|\overline{\beta}|} \end{bmatrix}$$

as a constant function to obtain that $\overline{v}_{ii}(x) = \nabla g_i(x)$ for every $i \in \{1, \ldots, |\overline{\beta}|\}$ and $\overline{v}_{ij}(x) = 0$ when $i \neq j$. That is, the WCR condition as in Definition 4.2 recovers the NLP definition of WCR when such NLP constraints are modelled as a single structurally diagonal matrix constraint, with this choice of $\mathcal{U}_{\overline{\beta}}$. It is important to keep in mind, however, that even if the constraints $g_1(x) \geqslant 0, \ldots, g_m(x) \geqslant 0$ satisfy LICQ at $x^\star$, nondegeneracy may not hold at $x^\star$, as observed by Shapiro in [65, p. 309]. The converse, on the other hand, is true. Now, recall that [16, Ex. 5.2] exhibits an NLP problem with a feasible point that satisfies "MFCQ+WCR", but not LICQ, and the above discussion tells us that it can be used again to prove that nondegeneracy is strictly stronger than "Robinson's CQ+WCR". Moreover, nondegeneracy and LICQ are equivalent when considering multiple unidimensional semidefinite constraints, and so are Definition 4.2 and the NLP version of WCR. Furthermore, Forsgren [35, Sec. 2.3] and Andreani et al. [9, Def. 3.2] considered regularity notions different from nondegeneracy, that also recover the standard LICQ in NLP. Thus, in all cases, regardless of modelling, the example of [16, Ex. 5.2] can be used to conclude that "Robinson's CQ+WCR" is strictly weaker than all existing notions of nondegeneracy.

With this in mind, we proceed to the main result of this section:

Theorem 11 *If $x^\star$ is a local minimizer such that Robinson's CQ and the WCR property hold, then there are some Lagrange multipliers $\omega^\star \in \mathbb{S}^m_+$ and $\mu^\star \in \mathbb{R}^p$ such that the KKT conditions and WSOC hold for this pair of multipliers.*

Proof If $x^\star$ is a local minimizer of (NSDP), Theorem 4.1 tells us that for any given $\{\rho_k\}_{k\in\mathbb{N}} \to +\infty$, there is some sequence $\{x^k\}_{k\in\mathbb{N}} \to x^\star$ such that, for every $k \in \mathbb{N}$, x^k is a local minimizer of the penalty function

$$F_k(x) = f(x) + \frac{1}{4}||x - x^\star||_2^4 + \frac{\rho_k}{2}\left(||\Pi_{\mathbb{S}^m_+}(-g(x))||_F^2 + ||h(x)||_2^2\right).$$

Hence, it satisfies the first-order stationarity condition

$$\nabla F(x^k) = \nabla f(x^k) + ||x^k - x^\star||_2^2(x^k - x^\star) + Dh(x^k)^\top(\rho_k h(x^k)) - \\ - Dg(x^k)^*[\rho_k \Pi_{\mathbb{S}^m_+}(-g(x^k))] = 0.$$

Setting the approximate multipliers $\omega^k \doteq \rho_k \Pi_{\mathbb{S}^m_+}(-g(x^k))$ and $\mu^k \doteq \rho_k h(x^k)$, we obtain (20) and (21) due to Theorem 4.1. Also, x^k is second-order stationary in the nonsmooth sense (see Theorem 2.2), which means that, for each unitary vector

$d \in \mathbb{R}^n$, there exists some $\chi^k \in \partial(\Pi_{\mathbb{S}^m_+} \circ -g)(x^k)$ such that

$$
\begin{aligned}
& d^\top \nabla^2 F(x^k) d \\
& = d^\top \left(\nabla^2 f(x^k) - D^2 g(x^k)^*[\rho_k \Pi_{\mathbb{S}^m_+}(-g(x^k))] + \sum_{i=1}^{p} \mu_i^k \nabla^2 h(x^k) \right) d + \\
& \quad + \rho_k (Dh(x^k)d)^\top Dh(x^k) d - d^\top \left(Dg(x^k)^* \left[\rho_k \chi^k[d] \right] \right) + d^\top \Delta^k d \\
& \geqslant 0
\end{aligned}
\tag{26}
$$

where $\Delta^k \doteq ||x^k - x^\star||_2^2 \mathbb{I}_n + 2(x^k - x^\star)(x^k - x^\star)^\top$ and $\nabla^2 F(x^k)$ denotes the element of $\partial^2 F_k(x^k)$ that is defined in terms of χ^k, as an abuse of notation. By Theorem 2.3, there exists some $V^k \in \partial \Pi_{\mathbb{S}^m_+}(-g(x^k))$, such that

$$
\chi^k = V^k \circ -Dg(x^k),
$$

for every $k \in \mathbb{N}$.

Under Robinson's CQ, the sequences $\{\omega^k\}_{k\in\mathbb{N}}$ and $\{\mu^k\}_{k\in\mathbb{N}}$ are bounded, so they have convergent subsequences which we will consider to be themselves from now on, without loss of generality. Denote their limits by $\omega^\star$ and $\mu^\star$, respectively. In [15, Thm. 6.1], the authors also prove that $\omega^\star$ and $\mu^\star$ are Lagrange multipliers associated with $x^\star$.

Now, let $d \in S(x^\star)$. By WCR there is a sequence $\{d^k\}_{k\in\mathbb{N}} \to d$ such that $d^k \in S(x^k, x^\star)$ for every k. Rewriting (26) in terms of d^k, V^k, ω^k, and μ^k, we obtain

$$
\begin{aligned}
& (d^k)^\top \nabla_x^2 L(x^k, \omega^k, \mu^k) d^k + \rho_k (Dh(x^k) d^k)^\top Dh(x^k) d^k + \\
& \qquad + \rho_k \left\langle Dg(x^k)[d^k], V^k \left[Dg(x^k)[d^k] \right] \right\rangle \geqslant -\delta^k,
\end{aligned}
\tag{27}
$$

where $\delta^k \doteq (d^k)^\top \Delta^k d^k \to 0$. The following paragraphs prove that (27) implies

$$
d^\top \nabla_x^2 L(x^\star, \omega^\star, \mu^\star) d + 2 \left\langle Dg(x^\star)[d], \omega^\star Dg(x^\star)[d] g(x^\star)^\dagger \right\rangle \geqslant 0, \tag{28}
$$

which is enough to complete the proof since

$$
2 \left\langle \omega^\star, Dg(x^\star)[d] g(x^\star)^\dagger Dg(x^\star)[d] \right\rangle = d^\top \sigma(x^\star, \omega^\star) d,
$$

for every $d \in \mathbb{R}^n$ due to (22).

To complete that task, we proceed to analyse the behaviour of the sequence $\{\rho_k \langle Dg(x^k)[d^k], V^k[Dg(x^k)[d^k]] \rangle\}_{k\in\mathbb{N}}$ in distinct cases. In the following paragraphs, we let $\alpha \doteq \alpha(x^k)$, $\beta \doteq \beta(x^k)$, and $\gamma \doteq \gamma(x^k)$ be the sets of indices of the positive, zero and negative eigenvalues of $g(x^k)$, respectively, regarding the spectral decomposition

$$
g(x^k) = S^k \mathrm{Diag}(\lambda^{S^k}(g(x^k)))(S^k)^\top
$$

with $S^k \to U$. Recall that, by construction, the columns of $\mathcal{U}_{\overline{\beta}}(g(x^k))$ span the eigenspace associated with the $\overline{\beta}$ smallest eigenvalues of $g(x^k)$, for all k sufficiently large. Denote the submatrix of S^k that has the eigenvectors associated with the $\overline{\beta}$ smallest eigenvectors of $g(x^k)$ in its columns by $S^k_{\overline{\beta}}$, and since $d^k \in S(x^k, x^\star)$, we have $(S^k_{\overline{\beta}})^\top Dg(x^k)[d^k] S^k_{\overline{\beta}} = 0$ for every k large enough. We proceed by analysing a few cases:

1. If $g(x^\star) \succ 0$, then $-g(x^k) \prec 0$ for k sufficiently large. For such k, since $\gamma(x^k) = \beta(x^k) = \emptyset$ and $\alpha(x^k) = \{1,\ldots,m\}$, we obtain and $V^k[Dg(x^k)[d^k]] = 0$ from Proposition 4.1, which implies
$$\rho_k \langle Dg(x^k)[d^k],\ V^k[Dg(x^k)[d^k]]\rangle = 0.$$
Also, note that $\sigma(x^\star, \omega^\star) = 0$ in this case, because $\langle g(x^\star), \omega^\star\rangle = 0$ implies $\omega^\star = 0$.
2. If $g(x^\star) = 0$, then $g(x^\star)^\dagger = 0$ and $\sigma(x^\star, \omega^\star) = 0$ as well. On the other hand, note that
$$\begin{aligned}\langle Dg(x^k)[d^k],\ V^k[Dg(x^k)[d^k]]\rangle &= \langle D\tilde{g}(x^k)[d^k],\ (U^k)^\top V^k[Dg(x^k)[d^k]]U^k\rangle \\ &= 0,\end{aligned}$$
because $\overline{\beta} = \{1,\ldots,m\}$ and $d^k \in S(x^k, x^\star)$ implies
$$D\tilde{g}(x^k)[d^k] = (S^k)^\top Dg(x^k)[d^k] S^k = 0$$
in this case.
3. If $g(x^\star) \succeq 0$, but $g(x^\star) \neq 0$, assume the diagonalization is taken such that nonzero eigenvalues are separated from the others and the common zero eigenvalues between $g(x^\star)$ and $\omega^\star$ are discriminated, that is,
$$g(x^\star) = U \begin{bmatrix} \Lambda & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} U^\top \quad \text{and} \quad \omega^\star = U \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Gamma \end{bmatrix} U^\top,$$
where $\mathbb{S}^{|\overline{\alpha}|} \ni \Lambda \succ 0$ and $\mathbb{S}^{|\overline{\gamma}|} \ni \Gamma \succ 0$ are diagonal matrices, $\overline{\kappa} \cup \overline{\gamma}$ is a partition of $\overline{\beta}$, and U is orthogonal. Denoting $H \doteq Dg(x^\star)[d]$, since $d \in S(x^\star)$ we get $\tilde{H}_{\overline{\beta}\overline{\beta}} = 0$ and
$$\begin{aligned} g(x^\star)^\dagger H \omega^\star &= U \begin{bmatrix} \Lambda^{-1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} U^\top U \begin{bmatrix} \tilde{H}_{\overline{\alpha}\overline{\alpha}} & \tilde{H}_{\overline{\alpha}\overline{\kappa}} & \tilde{H}_{\overline{\alpha}\overline{\gamma}} \\ \tilde{H}_{\overline{\kappa}\overline{\alpha}} & 0 & 0 \\ \tilde{H}_{\overline{\gamma}\overline{\alpha}} & 0 & 0 \end{bmatrix} U^\top U \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Gamma \end{bmatrix} U^\top \\ &= U \begin{bmatrix} 0 & 0 & \Lambda^{-1}\tilde{H}_{\overline{\alpha}\overline{\gamma}}\Gamma \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} U^\top. \end{aligned}$$
Conveniently,
$$\Lambda^{-1}\tilde{H}_{\overline{\alpha}\overline{\gamma}}\Gamma = \left[\frac{\lambda_j^U(\omega^\star)}{\lambda_i^U(g(x^\star))} \tilde{H}_{ij} \right]_{i\in\overline{\alpha}, j\in\overline{\gamma}} \doteq A \odot \tilde{H}_{\overline{\alpha}\overline{\gamma}}, \tag{29}$$
where $\mathbb{R}^{|\overline{\alpha}|\times|\overline{\gamma}|} \ni A \doteq \left[\lambda_j^U(\omega^\star)\lambda_i^U(g(x^\star))^{-1}\right]_{i\in\overline{\alpha}, j\in\overline{\gamma}}$. Also, note that

$$\langle H, g(x^\star)^\dagger H\omega^\star\rangle = \left\langle \tilde{H}, \frac{1}{2}\begin{bmatrix} 0 & 0 & A\odot \tilde{H}_{\overline{\alpha\gamma}} \\ 0 & 0 & 0 \\ (A\odot \tilde{H}_{\overline{\alpha\gamma}})^\top & 0 & 0 \end{bmatrix}\right\rangle.$$

In view of this characterization of the sigma-term over d, its relation with (27) can be made explicit. Consider the following spectral decomposition of $g(x^k)$:

$$g(x^k) = S^k \begin{bmatrix} \Lambda_+^k & 0 & 0 & 0 \\ 0 & \Lambda_-^k & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Psi^k \end{bmatrix} (S^k)^\top, \tag{30}$$

where we separate the eigenvalues of $g(x^k)$ primarily by their sign and, secondarily, by their limit points. For instance, $\Lambda_+^k \in \mathbb{S}^{|\alpha_+|}$ are the positive ones that converge to Λ, while $\Lambda_-^k \in \mathbb{S}^{|\alpha_-|}$ are the positive ones that converge to zero. The squared block of zeros in the diagonal of (30) is of dimension β and $\Psi^k \in \mathbb{S}^{|\gamma|}$ contains the negative eigenvalues of $g(x^k)$. Also, $|\alpha_+|+|\alpha_-|+|\beta|+|\gamma| = m$. Recall that S^k simultaneously diagonalizes $g(x^k)$ and ω^k, by definition of ω^k. In order to simplify the notation, define

$$H^k \doteq Dg(x^k)[d^k]$$

and

$$B_{\alpha\gamma}^k \doteq \tilde{H}_{\alpha\gamma}^k \odot \mathcal{B}(\lambda^{S^k}(-g(x^k)))_{\alpha\gamma}.$$

Using the characterization of V^k provided in (23) from Proposition 4.1 (and Corollary 4.1), we obtain

$$V^k[H^k] = S^k \begin{bmatrix} 0 & 0 & 0 & B_{\alpha_+\gamma}^k \\ 0 & 0 & 0 & B_{\alpha_-\gamma}^k \\ 0 & 0 & V_{|\beta|}[\tilde{H}_{\beta\beta}] & \tilde{H}_{\beta\gamma}^k \\ (B_{\alpha_+\gamma}^k)^\top & (B_{\alpha_-\gamma}^k)^\top & (\tilde{H}_{\beta\gamma}^k)^\top & \tilde{H}_{\gamma\gamma}^k \end{bmatrix} (S^k)^\top.$$

Since $\rho_k\langle H^k, V^k[H^k]\rangle = \langle \tilde{H^k}, \rho_k \tilde{V^k[H^k]}\rangle$, it is fundamental to note that for every $i \in \alpha_+$ and $j \in \gamma$,

$$\Big(\rho_k \mathcal{B}(\lambda^{S^k}(-g(x^k)))\Big)_{ij} = \frac{\rho_k \lambda_j^{S^k}(-g(x^k))}{\lambda_j^{S^k}(-g(x^k)) - \lambda_i^{S^k}(-g(x^k))} \to \frac{\lambda_j^U(\omega^\star)}{\lambda_i^U(g(x^\star))}, \tag{31}$$

because $\rho_k\lambda_j^{S^k}(-g(x^k)) = \lambda_j^{S^k}(\omega^k)$ and $\lambda_j^{S^k}(-g(x^k)) \to 0$. Also, keep in mind that

$$\lambda_i^{S^k}(-g(x^k)) = -\lambda_i^{S^k}(g(x^k)).$$

The blocks indexed by α_-, β, and γ, are all blocks of zeros because if k is large enough, we must have $|\alpha_-| + |\beta| + |\gamma| = \overline{\beta}$ and, on the other hand, since $d^k \in S(x^k, x^\star)$ we also have that $\tilde{H}_{\overline{\beta\beta}} \doteq (S^k_{\overline{\beta}})^\top Dg(x^k)[d^k] S^k_{\overline{\beta}} = 0$. Similarly, $Dh(x^k)d^k = 0$. Thus

$$\lim_{k\to\infty} \rho_k \langle H^k, V^k[H^k]\rangle = 2\langle H, g(x^\star)^\dagger H\omega^\star\rangle$$

and, consequently, (27) implies (28), which means $x^\star$ satisfies the WSOC with the multiplier $\omega^\star$.

□

In the presence of nondegeneracy, the set of Lagrange multipliers is a singleton and Theorem 4.2 recovers the classical result of [65], but even without assuming uniqueness of the Lagrange multiplier it ensures there will be at least one multiplier satisfying WSOC. Moreover, in contrast with [48, 56, 65], our proof does not require strict complementarity; but nevertheless, if it does hold, then the proof of Theorem 4.2 can be significantly simplified, since in this case the sequence $\{g(x^k)\}_{k\in\mathbb{N}}$ is nonsingular and we can avoid the use of subdifferentials.

5 Conclusion

In this paper, we proved that every local minimizer of a nonlinear semidefinite program or a nonlinear second-order cone program satisfies the weak second-order necessary optimality condition under Robinson's constraint qualification and the so-called weak constant rank property (WCR), which was extended from NLP [16]. This joint condition is strictly weaker than nondegeneracy in NLP, NSOCP, and NSDP. We also stress that we do not assume strict complementarity, which is common in second-order analyses for conic programming. In contrast, our second-order necessary condition is based on the lineality space of the critical cone, and not the critical cone itself. This is consistent with the algorithmic practice of second-order algorithms as no algorithm is known to achieve a stronger second-order necessary optimality condition (see the extended version of [22] for details).

In the context of conic programming, several different approaches are known for obtaining second-order necessary optimality conditions [28, 30, 35, 37, 48, 56, 65]. We present a novel approach by extending the existing theory of first-order sequential optimality conditions to the second-order context. In particular, it is remarkable to see the appearance of the *sigma-term* in such a variety of approaches, which contributes to the understanding of this concept.

Our approach has a heavy algorithmic taste, as our proof is based on the construction of a sequence of approximate solutions of penalized subproblems, very similarly to a sequence generated by practical algorithms. In particular, a similar first-order approach has recently led to several improvements of global convergence theory of augmented Lagrangian methods in conic contexts [4–6, 15].

Thus, this paper opens the path to the development of second-order algorithms in conic optimization, which, as far as we know, has not been considered yet in the literature. In particular, augmented Lagrangian and interior point methods [25, 41] are expected to be well suited to the techniques we develop here. In this context, the joint condition "Robinson's CQ+WCR" is the natural candidate for a condition to guarantee global convergence to a second-order stationary point.

Statements and declarations

The authors have no relevant financial or non-financial interests to disclose. Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

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

[2] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Second-order negative-curvature methods for box-constrained and general constrained optimization. *Computational Optimization and Applications*, 45:209–236, 2010.

[3] R. Andreani, C. E. Echagüe, and M. L. Schuverdt. Constant-rank condition and second-order constraint qualification. *Journal of Optimization theory and Applications*, 146:255–266, 2010.

[4] R. Andreani, E. H. Fukuda, G. Haeser, D. O. Santos, and L. D. Secchin. Optimality conditions for nonlinear second-order cone programming and symmetric cone programming. Technical report, 2019. Available at Optimization Online: http://www.optimization-online.org/DB_HTML/2019/10/7436.html.

[5] R. Andreani, E. H. Fukuda, G. Haeser, D. O. Santos, and L. D. Secchin. On the use of Jordan Algebras for improving global convergence of an augmented Lagrangian method in nonlinear semidefinite programming. *Computational Optimization and Applications*, 79, 2021.

[6] R. Andreani, W. Gómez, G. Haeser, L. M. Mito, and A. Ramos. On optimality conditions for nonlinear conic programming. *To appear in Mathematics of Operations Research*, 2022. DOI: 10.1287/moor.2021.1203.

[7] R. Andreani, G. Haeser, and J. M. Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60:627–641, 2011.

[8] R. Andreani, G. Haeser, L. M. Mito, H. Ramırez, D. O. Santos, and T. P. Silveira. Naive constant rank-type constraint qualifications for multifold

second-order cone programming and semidefinite programming. *To appear in Optimization Letters*, 2021. DOI: 10.1007/s11590-021-01737-w.

[9] R. Andreani, G. Haeser, L. M. Mito, and H. Ramírez C. Weak notions of nondegeneracy in nonlinear semidefinite programming. Technical report, 2020. Available at ArXiv:2012.14810.

[10] R. Andreani, G. Haeser, L. M. Mito, and H. Ramírez C. Sequential constant rank constraint qualifications for nonlinear semidefinite programming with applications. Technical report, 2021. Available at ArXiv:2106.00775v2.

[11] R. Andreani, G. Haeser, L M. Mito, H. Ramírez C., and T. P. Silveira. First- and second-order optimality conditions for second-order cone and semidefinite programming under a constant rank condition. Technical report, 2021. Available at ArXiv:2107.04693v1.

[12] R. Andreani, G. Haeser, L M. Mito, H. Ramírez C., and T. P. Silveira. Sequential constant rank for nonlinear second-order cone programming problems. Technical report, 2021.

[13] R. Andreani, G. Haeser, A. Ramos, and P. J. S. Silva. A second-order sequential optimality condition associated to the convergence of optimization algorithms. *IMA Journal of Numerical Analysis*, 37(4):1902–1929, 2017.

[14] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. Two new weak constraint qualifications and applications. *SIAM Journal on Optimization*, 22:1109–1135, 2012.

[15] R. Andreani, G. Haeser, and D. S. Viana. Optimality conditions and global convergence for nonlinear semidefinite programming. *Mathematical Programming*, 180(1-2):203–235, 2020.

[16] R. Andreani, J. M. Martínez, and M. L. Schuverdt. On second-order optimality conditions for nonlinear programming. *Optimization*, 56:529–542, 2007.

[17] R. Andreani, J. M. Martínez, and B.F. Svaiter. A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization*, 20:3533–3554, 2010.

[18] Roberto Andreani and Leonardo 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.

[19] M. F. Anjos and J. B. Lassere (eds.). *Handbook on Semidefinite, Conic and Polynomial Optimization.* International Series in Operations Research and Management Science. Springer US. New York, 2012.

[20] J. P. Aubin and H. Frankowska. *Set-Valued Analysis.* Birkhauser, 1990.

[21] A. Baccari. On the Classical Necessary Second-Order Optimality Conditions. *Journal of Optimization Theory and Applications*, 123(1):213–221, 2004.

[22] R. Behling, G. Haeser, A. Ramos, and D. S. Viana. On a conjecture in second-order optimality conditions. *Journal of Optimization Theory and Applications*, 176(3):625–633, 2018. Extended version at arXiv:1706.07833.

[23] H. Y. Benson and R. J. Vanderbei. Solving problems with semidefinite and related constraints using interior-point methods for nonlinear programming. *Mathematical Programming*, 95:279–302, 2003.

[24] E. G. Birgin, W. Gómez, G. Haeser, L. M. Mito, and D. O. Santos. An Augmented Lagrangian algorithm for nonlinear semidefinite programming applied to the covering problem. *Computational and Applied Mathematics*, 39(10), 2020.

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

[26] J. F. Bonnans, R. Cominetti, and A. Shapiro. Sensitivity analysis of optimization problems under second order regular constraints. *Mathematics of Operations Research*, 23(4):806–831, 1998.

[27] J. F. Bonnans, R. Cominetti, and A. Shapiro. Second order optimality conditions based on parabolic second order tangent sets. *SIAM Journal on Optimization*, 9(2):466–492, 1999.

[28] J. F. Bonnans and H. Ramírez C. Perturbation analysis of second-order cone programming problems. *Mathematical Programming*, 104:205–227, 2005.

[29] J. F. Bonnans and A. Shapiro. *Pertubation Analysis of Optimization Problems.* Springer-Verlag. Berlin, 2000.

[30] R. Cominetti. Metric regularity, tangent sets, and second-order optimality conditions. *Applied Mathematics and Optimization*, 287:265–287, 1990.

[31] F. Facchinei and J.S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume I.* Springer Series in Operations Research. Springer-Verlag. New York, 2003.

[32] B. Fares, P. Apkarian, and D. Noll. An augmented Lagrangian method for a class of LMI-constrained problems in robust control theory. *International Journal of Control*, 74(4):348–360, 2001.

[33] B. Fares, D. Noll, and P. Apkarian. Robust control via sequential semidefinite programming. *SIAM Journal on Control and Optimization*, 40:1791–1820, 2002.

[34] S. Fitzpatrick and R. R. Phelps. Differentiability of the metric projection in Hilbert space. *Transactions of the American Mathematical Society*, 270(2):483–501, 1982.

[35] A. Forsgren. Optimality conditions for nonconvex semidefinite programming. *Mathematical Programming*, 88:105–128, 2000.

[36] R. W. Freund, F. Jarre, and C. H. Vogelbusch. Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling. *Mathematical Programming*, 109:581–611, 2007.

[37] E. H. Fukuda and M. Fukushima. The use of squared slack variables in nonlinear second-order cone programming. *Journal of Optimization Theory and Applications*, 170(2):394–418, 2016.

[38] E. H. Fukuda, P. J. S. Silva, and M. Fukushima. Differentiable exact penalty functions for nonlinear second-order cone programs. *SIAM Journal on Optimization*, 22(4):1607–1633, 2012.

[39] P. E. Gill, V. Kungurtsev, and D. P. Robinson. A stabilized SQP method: global convergence. *IMA Journal of Numerical Analysis*, 2016. DOI: 10.1093/imanum/drw004.

[40] Nicholas IM Gould and Philippe L Toint. A note on the convergence of barrier algorithms to second-order necessary points. *Mathematical programming*, 85(2):433–438, 1999.

[41] G. Haeser. A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms. *Computational Optimization and Applications*, 70(2):615–639, 2018.

[42] G. Haeser. Some theoretical limitations of second-order algorithms for smooth constrained optimization. *Operations Research Letters*, 46(3):295–299, 2018.

[43] S. Hayashi, N. Yamashita, and M. Fukushima. A combined smoothing and regularization method for monotone second-order cone complementarity problems. *SIAM Journal on Optimization*, 15:593–615, 2005.

[44] M. R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4:303–320, 1969.

[45] J.-B. Hiriart-Urruty, J.-J. Strodiot, and V. H. Nguyen. Generalized Hessian matrix and second-order optimality conditions for problems with C1,1 data. *Applied Mathematics and Optimization*, 11:43–56, 1984.

[46] R. Janin. Direction derivative of the marginal function in nonlinear programming. *Mathematical Programming Studies*, 21:127–138, 1984.

[47] F. Jarre. An interior method for nonconvex semidefinite programs. *Optimization and Engineering*, 1:347–372, 2000.

[48] F. Jarre. Elementary optimality conditions for nonlinear SDPs. In M. F. Anjos and J. B. Lassere, editors, *Handbook on Semidefinite, Conic and Polynomial Optimization*. International Series in Operations Research and Management Science, 2012.

[49] Y. Kanno and I. Takewaki. Sequential semidefinite program for maximum robustness design of structures under load uncertainty. *Journal of Optimization Theory and Applications*, 130:265–287, 2006.

[50] C. Kanzow, I. Ferenczi, and M. Fukushima. On the local convergence of semismooth Newton methods for linear and nonlinear second-order cone programs without strict complementarity. *SIAM Journal on Optimization*, 20:297–320, 2009.

[51] C. Kanzow, C. Nagel, H. Kato, and M. Fukushima. Successive linearization methods for nonlinear semidefinite programs. *Computational Optimization and Applications*, 31:251–273, 2005.

[52] H. Kato and M. Fukushima. An SQP-type algorithm for nonlinear second-order cone programs. *Optimization Letters*, 1:129–144, 2007.

[53] H. Kawasaki. An envelope-like effect of infinitely many inequality constraints on second-order necessary conditions for minimization problems. *Mathematical Programming*, 41:73–96, 1988.

[54] M. Kočvara and M. Stingl. Solving nonconvex SDP problems of structural optimization with stability control. *Optimization Methods and Software*, 19(5):595–609, 2004.

[55] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and Applications*, 284:193–228, 1998.

[56] B. F. Lourenço, E. H. Fukuda, and M. Fukushima. Optimality conditions for nonlinear semidefinite programming via squared slack variables. *Mathematical Programming*, 168(1-2):177–200, 2018.

[57] J. M. Morguerza and F. J. Prieto. An augmented Lagrangian interior-point method using directions of negative curvature. *Mathematical Programming*, 95:3:573–616, 2003.

[58] K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39(2):117–129, 1987.

[59] Z. Páles and V. Zeidan. Generalized Jacobian for functions with infinite dimensional range and domain. *Set-Valued Analysis*, 15:331–375, 2007.

[60] J. S. Pang, D. Sun, and J. Sun. Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems. *Mathematics of Operations Research*, 28(1):39–63, 2003.

[61] M. J. D. Powell. *A method for nonlinear constraints in minimization problems.* Academic Press. New York, 1982.

[62] H. D. Qi. Local duality of nonlinear semidefinite programming. *Mathematics of Operations Research*, 34(1):124–141, 2009.

[63] S. M. Robinson. First-order conditions for general nonlinear optimization. *SIAM Journal on Applied Mathematics*, 30(4):597–610, 1976.

[64] R. T. Rockafellar. Lagrange multipliers and optimality. *SIAM Review*, 35:183–238, 1993.

[65] A. Shapiro. First and second order analysis of nonlinear semidefinite programs. *Mathematical Programming, Series B*, 77(2):301–320, 1997.

[66] A. Shapiro and M. K. H. Fan. On Eigenvalue optimization. *SIAM Journal on Optimization*, 5(3):552–569, 1995.

[67] D. Sun. The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Mathematics of Operations Research*, 31(4):761–776, 2006.

[68] H. Wolkowicz, R. Saigal, and L. (eds.) Vandenberghe. *Handbook of Semidefinite Programming: Theory, Algorithms, an Applications*. International Series in Operations Research and Management Science. Springer

US. New York, 2000.

[69] H. Yamashita and H. Yabe. A primal-dual interior point method for nonlinear optimization over second-order cones. *Optimization Methods and Software*, 24:407–426, 2009.

[70] H. Yamashita and H. Yabe. A survey of numerical methods for nonlinear semidefinite programming. *Journal of the Operations Research Society of Japan*, 58(1):24–60, 2015.

[71] X. Y. Zhao, D. Sun, and K. C. Toh. A Newton-CG augmented Lagrangian method for semidefinite programming. *SIAM Journal on Optimization*, 20(4):1737–1765, 2010.

Appendix F

External reference VI

Title: Weak notions of nondegeneracy in nonlinear semidefinite programming.

Status: Under review.

What is attached: A preprint that has *not* been peer reviewed yet.

Weak notions of nondegeneracy in nonlinear semidefinite programming

Roberto Andreani * Gabriel Haeser † Leonardo M. Mito † Héctor Ramírez ‡

January 1, 2021

Abstract

The constraint nondegeneracy condition is one of the most relevant and useful constraint qualifications in nonlinear semidefinite programming. It can be characterized in terms of any fixed orthonormal basis of the, let us say, ℓ-dimensional kernel of the constraint matrix, by the linear independence of a set of $\ell(\ell+1)/2$ derivative vectors. We show that this linear independence requirement can be equivalently formulated in a smaller set, of ℓ derivative vectors, by considering all orthonormal bases of the kernel instead. This allows us to identify that not all bases are relevant for a constraint qualification to be defined, giving rise to a strictly weaker variant of nondegeneracy related to the global convergence of an external penalty method. Also, by exploiting the sparsity structure of the constraints, we were able to define another weak variant of nondegeneracy by removing the null entries from consideration. In particular, both our new constraint qualifications reduce to the linear independence constraint qualification for nonlinear programming when considering a diagonal semidefinite constraint. More generally, when the problem has a diagonal block structure, the conditions formulated as a single block diagonal matrix constraint are equivalent to their analogues formulated with several semidefinite matrices.

Keywords: Semidefinite programming, Constraint qualifications, Constraint nondegeneracy.

1 Introduction

The study of *linear* and *nonlinear semidefinite programming* (for short, SDP and NSDP, respectively) problems has been consistently growing over the last decades. There are several models for real world problems that can be reformulated as SDPs or NSDPs (we refer to the handbooks [9, Part 4] and [29, Part 3] for a vast collection of applications), which motivate and are motivated by the development of theoretical results regarding optimality conditions and *constraint qualifications* (CQs) for (N)SDPs. Loosely speaking, CQs are assumptions over the feasible set of an optimization problem that ensure that it can be locally described in terms of its first-order approximation. This leads to the possibility of characterizing all solutions of an (N)SDP problem in terms of the derivatives of the functions that describe it, which gives CQs a pivotal role in building convergence theories for practical algorithms. The standard way to do this is to prove that every feasible limit point of the output sequence of the algorithm satisfies the Karush-Kuhn-Tucker (KKT) conditions under a given CQ. Thus, employing a weaker CQ leads to a more robust convergence theory.

One of the most relevant CQs in the literature of (N)SDP is the so-called *nondegeneracy* (or *transversality) condition*, introduced by Shapiro and Fan in [26, Sec. 2] in the context of eigenvalue optimization, and later reformulated by Shapiro [24, Def. 4] for general NSDPs. This condition has been widely used for characterizing sensitivity results (see, for instance, [13, 16, 18, 19, 20, 27]), and also for proving global convergence and the rate of convergence of numerical algorithms (we refer to Yamashita and Yabe [30, Secs. 3, 4, and 5] for a survey on this topic). However, it is known that even in the linear case, the solutions of large scale SDP problems tend to be degenerate, even though nondegeneracy is expected to hold in a generic sense. Besides, when the constraint of an NSDP problem has some sparsity structure near one of its solutions – for instance, a diagonal structure – then nondegeneracy is not satisfied at that solution [24]. This means that the convergence theory of an algorithm supported by nondegeneracy does not cover such points.

The authors received financial support from CEPID - FAPESP (grant 2013/07375-0), FAPESP (grants 2018/24293-0, 2017/18308-2, and 2017/17840-2), 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 ANID PIA AFB 170001, both from ANID (Chile).

*Department of Applied Mathematics, State University of Campinas, Campinas, SP, Brazil. Email: andreani@unicamp.br

†Department of Applied Mathematics, University of São Paulo, São Paulo, SP, Brazil. Emails: ghaeser@ime.usp.br, leokoto@ime.usp.br

‡Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático (AFB170001 - CNRS UMI 2807), Universidad de Chile, Santiago, Chile. Email: hramirez@dim.uchile.cl

The explanation for such kind of issue, in our opinion, is the degree of generality of the nondegeneracy condition. That is, although it was born in NSDP, nondegeneracy does not capture any particularity of the constraints, being straigthforwardly extended for any general conic optimization problem, as long as the cone is closed and convex. However, embedding specific traits of matrix-valued functions into nondegeneracy may be more or less direct, depending on how it is characterized. In this paper, instead of defining nondegeneracy as the transversality of two particular subspaces – which is the most usual definition – we exploit an equivalent characterization by Shapiro [24, Prop. 6], which is phrased in terms of the gradients of the entries of a special function.

The contributions of this paper revolve around the following results:

- We provide a new characterization of nondegeneracy that induces a weaker variant of it, here called *weak-nondegeneracy*, which uses information of the eigenvectors of the constraints evaluated at nearby points;
- We incorporate a sparsity treatment in [24, Prop. 6], which leads to another weak variant of nondegeneracy, called *sparse-nondegeneracy*.

These conditions are designed with the sole goal of assisting in proving global convergence of algorithms by means of sequential optimality conditions [3, 8]; however, we envision that they may be further employed in sensitivity analysis, second-order analysis, among other applications. Both variants are proved to be constraint qualifications strictly weaker than nondegeneracy. We also show that when weak- and sparse-nondegeneracy are applied to diagonal matrices, they are reduced to the *linear independence constraint qualification* (LICQ) from nonlinear programming (NLP). More generally, both conditions are invariant to block representations of (N)SDP problems as a single semidefinite block diagonal matrix or as multiple semidefinite constraints. Then, we compare our definitions with other CQs from the literature.

This paper is structured as follows: In Section 2, we introduce our notation; in Section 3 we recall the nondegeneracy condition and we prove a new characterization of it, which is where the definition of weak-nondegeneracy comes from. In Section 4, we present our definition of sparse-nondegeneracy. Section 5 briefly describes extensions of the previous definitions and results to Robinson's CQ. Finally, in Section 6, we discuss some possibilities for prospective work.

2 Preliminaries

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ and $G\colon \mathbb{R}^n \to \mathbb{S}^m$ be continuously differentiable functions, where $\mathbb{S}^m$ is the linear space of all $m\times m$ symmetric matrices, and let $\mathbb{S}^m_+$ be the closed convex pointed cone of all $m\times m$ positive semidefinite matrices. The problem of interest in this paper is the following:

$$\begin{aligned} &\underset{x\in\mathbb{R}^n}{\text{Minimize}} && f(x), \\ &\text{subject to} && G(x)\succeq 0, \end{aligned} \tag{NSDP}$$

where $\succeq$ is the partial order induced by $\mathbb{S}^m_+$, characterized by the relation: $M \succeq N \Leftrightarrow M - N \in \mathbb{S}^m_+$, for all $M, N \in \mathbb{S}^m$. It is worth pointing out that all results in this paper can be straightforwardly extended to NSDP problems with separate equality constraints, but we omit them for simplicity. The feasible set of (NSDP) will be denoted by $\mathcal{F} \doteq G^{-1}(\mathbb{S}^m_+)$. It is well-known that $\mathbb{S}^m$ is an Euclidean space when equipped with the (*Frobenius*) inner product $\langle M, N\rangle \doteq \text{trace}(MN) \doteq \sum_{i,j=1}^m M_{ij}N_{ij}$.

The derivative of G at a point $x \in \mathbb{R}^n$ is the linear mapping $DG(x)\colon \mathbb{R}^n \to \mathbb{S}^m$ that can be described (in the canonical basis of $\mathbb{R}^n$) by the action

$$d \mapsto DG(x)[d] \doteq \sum_{i=1}^n D_{x_i}G(x)d_i$$

for all $d = (d_1, \dots, d_n) \in \mathbb{R}^n$, where $D_{x_i}G(x) \in \mathbb{S}^m$ is the partial derivative of G with respect to the variable x_i at $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Also, for each fixed x, the *adjoint* of $DG(x)$ is the unique linear mapping $DG(x)^*\colon \mathbb{S}^m \to \mathbb{R}^n$ that satisfies $\langle DG(x)[d], M\rangle = \langle d, DG(x)^*[M]\rangle$, for all $(d, M) \in \mathbb{R}^n \times \mathbb{S}^m$. Hence,

$$DG(x)^*[M] = \begin{bmatrix} \langle D_{x_1}G(x), M\rangle \\ \vdots \\ \langle D_{x_n}G(x), M\rangle \end{bmatrix} = \sum_{i,j=1}^m M_{ij}\nabla G_{ij}(x)$$

for all $M \in \mathbb{S}^m$, where $\nabla G_{ij}(x)$ denotes the *gradient* of the (i, j)-th entry of G as a function of x. Similarly, we shall denote the gradient of any real-valued function $F\colon \mathbb{R}^n \to \mathbb{R}$ at a point $x \in \mathbb{R}^n$ by $\nabla F(x)$.

For any given $M \in \mathbb{S}^m$, we consider its spectral decomposition in the form

$$M = \sum_{i=1}^m \lambda_i(M)u_i(M)u_i(M)^\top,$$

where $\lambda_i(M) \in \mathbb{R}$ denotes the i-th eigenvalue of M arranged in non-increasing order (that is, $\lambda_1(M) \geqslant \lambda_2(M) \geqslant \ldots \geqslant \lambda_m(M)$), and $u_i(M) \in \mathbb{R}^m$ corresponds to any associated eigenvector such that the set

$\{u_i(M)\colon i \in \{1,\ldots,m\}\}$ is an orthonormal basis of $\mathbb{R}^m$ (that is, $u_i(M)^T u_i(M) = 1$ and $u_i(M)^T u_j(M) = 0$ when $i \neq j$, for all $i, j \in \{1,\ldots,m\}$).

A useful fact for our analyses is that the orthogonal projection of M onto $\mathbb{S}^m_+$ with respect to the induced (Frobenius) norm, denoted by $\Pi_{\mathbb{S}^m_+}(M)$, can be characterized in terms of its spectral decomposition as follows:

$$\Pi_{\mathbb{S}^m_+}(M) = \sum_{i=1}^{m} [\lambda_i(M)]_+ u_i(M) u_i(M)^\top,$$

where $[\lambda]_+ \doteq \max\{0, \lambda\}$ for all $\lambda \in \mathbb{R}$.

Given any $\overline{x} \in \mathcal{F}$ and any orthogonal matrix $\overline{U} \in \mathbb{R}^{m\times m}$ whose columns are eigenvectors of $G(\overline{x})$, we partition $\overline{U} = [\overline{P}, \overline{E}]$ such that the columns of $\overline{P} \in \mathbb{R}^{m\times r}$ correspond to the eigenvectors associated with the positive eigenvalues of $G(\overline{x})$ and the columns of $\overline{E} \in \mathbb{R}^{m\times m-r}$ correspond to the eigenvectors associated with the null eigenvalues of $G(\overline{x})$, where $r = \operatorname{rank}(G(\overline{x}))$. To abbreviate, as an abuse of notation and language, we will say that $\overline{E}$ *spans* $\operatorname{Ker} G(\overline{x})$ in this context. That is, $\overline{E}$ spans $\operatorname{Ker} G(\overline{x})$ if, and only if, $\overline{E}^\top \overline{E} = \mathbb{I}_{m-r}$ and $G(\overline{x})\overline{E} = 0$, where $\mathbb{I}_{m-r}$ denotes an $(m-r)$-dimensional identity matrix.

There are multiple ways of describing optimality in NSDP problems, but in this paper we direct our attention to necessary optimality conditions that are based on the classical *Karush-Kuhn-Tucker* (KKT) conditions:

Definition 2.1 (KKT)**.** *We say that a point $\overline{x} \in \mathcal{F}$ satisfies the* KKT conditions *when there exists some $\overline{Y} \succeq 0$ such that*

$$\begin{aligned} \nabla_x L(\overline{x}, \overline{Y}) \doteq \nabla f(\overline{x}) - DG(\overline{x})^*[\overline{Y}] = 0,\\ \langle G(\overline{x}), \overline{Y}\rangle = 0, \end{aligned} \tag{KKT}$$

where $L : \mathbb{R}^n \times \mathbb{S}^m \to \mathbb{R}$ is the Lagrangian function *of* (NSDP)*, given by*

$$L(x, Y) \doteq f(x) - \langle G(x), Y\rangle.$$

As usual, the matrix $\overline{Y}$ is called a *Lagrange multiplier* associated with $\overline{x}$ and we denote the set of all Lagrange multipliers associated with $\overline{x}$ by $\Lambda(\overline{x})$. When $\Lambda(\overline{x}) \neq \emptyset$, $\overline{x}$ is called a *KKT point* of (NSDP). Let r be the rank of $G(\overline{x})$ and let $\overline{E} \in \mathbb{R}^{m\times m-r}$ be a matrix that spans $\operatorname{Ker} G(\overline{x})$; then, for any $\overline{Y} \in \Lambda(\overline{x})$, since both $\overline{Y}$ and $G(\overline{x})$ are positive semidefinite, the complementarity relation $\langle G(\overline{x}), \overline{Y}\rangle = 0$ is equivalent to $G(\overline{x})\overline{Y} = 0$, which is in turn equivalent to saying that $\operatorname{Im} \overline{Y} \subseteq (\operatorname{Im} G(\overline{x}))^\perp = \operatorname{Ker} G(\overline{x})$, where $(\operatorname{Im} G(\overline{x}))^\perp$ denotes the orthogonal complement of $\operatorname{Im} G(\overline{x})$. Therefore, $\overline{Y}$ is complementary to $G(\overline{x})$ if, and only if, it has the form

$$\overline{Y} = \overline{E}\tilde{Y}\overline{E}^\top, \tag{1}$$

where $\tilde{Y} \in \mathbb{S}^{m-r}_+$ is not necessarily a diagonal matrix. Moreover, note that $\tilde{Y}$ is not necessarily positive definite; that is, $\dim(\operatorname{Ker} \overline{Y})$ does not necessarily coincide with r. When they do coincide, $\overline{x}$ and $\overline{Y}$ are said to be *strictly complementary* [24].

It is known that the KKT conditions are not necessary for local optimality unless they are paired with a constraint qualification. For instance, one of the most studied constraint qualifications for (NSDP) is *Robinson's CQ* [23, Def. 3], which holds at a point $\overline{x} \in \mathcal{F}$ if there exists $d \in \mathbb{R}^n$ such that

$$G(\overline{x}) + DG(\overline{x})[d] \in \operatorname{int} \mathbb{S}^m_+,$$

where $\operatorname{int} \mathbb{S}^m_+$ denotes the topological interior of $\mathbb{S}^m_+$, which in turn coincides with the set of $m \times m$ symmetric positive definite matrices. Alternatively, following Bonnans and Shapiro [10, Prop. 2.97], it is possible to say that (the dual form of) Robinson's CQ holds at $\overline{x} \in \mathcal{F}$ if, and only if,

$$\left.\begin{aligned} DG(\overline{x})^*[Y] = 0\\ \langle G(\overline{x}), Y\rangle = 0\\ Y \succeq 0 \end{aligned}\right\} \Rightarrow Y = 0. \tag{2}$$

Another well-known fact is that, for every local minimizer $\overline{x} \in \mathcal{F}$, the set $\Lambda(\overline{x})$ is nonempty and compact if, and only if, Robinson's CQ holds at $\overline{x}$ (see [10, Props. 3.9 and 3.17] for details). This makes Robinson's CQ the natural analogue of the *Mangasarian-Fromovitz CQ* (MFCQ), from NLP, in NSDP.

3 The nondegeneracy condition for NSDP

In this section, we discuss the well-known nondegeneracy condition introduced by Shapiro and Fan [26, Sec. 2]. We derive a different characterization for it that suggests a way of obtaining a weaker constraint qualification with potentially interesting properties. But firstly, we briefly recall some elements of convex analysis.

The (*Bouligand*) tangent cone to a set C at a point $y \in C$ is defined as

$$T_C(y) \doteq \left\{ d\colon \begin{array}{l} \exists\{d^k\}_{k\in\mathbb{N}} \to d,\ \exists\{t^k\}_{k\in\mathbb{N}} \to 0,\ t^k > 0, \\ \forall k \in \mathbb{N},\ y + t^k d^k \in C \end{array} \right\}.$$

In particular, when $C = \mathbb{S}^m_+$, at a given $M \in \mathbb{S}^m_+$, it can be characterized as follows

$$T_{\mathbb{S}^m_+}(M) = \left\{ N \in \mathbb{S}^m : d^\top N d \geqslant 0,\ \forall d \in \operatorname{Ker} M \right\}.$$

Therefore, for every feasible $\overline{x}$ we have

$$T_{\mathbb{S}^m_+}(G(\overline{x})) = \left\{ N \in \mathbb{S}^m : \overline{E}^\top N \overline{E} \succeq 0 \right\}, \tag{3}$$

whenever $\overline{E}$ spans $\operatorname{Ker} G(\overline{x})$.

It is clear from (3) that the largest subspace contained in $T_{\mathbb{S}^m_+}(G(\overline{x}))$, that is, its *lineality space*, can be characterized as follows:

$$\operatorname{lin}(T_{\mathbb{S}^m_+}(G(\overline{x}))) = \left\{ N \in \mathbb{S}^m : \overline{E}^\top N \overline{E} = 0 \right\}. \tag{4}$$

The nondegeneracy condition of Shapiro and Fan is verified at $\overline{x}$ when the linear subspaces $\operatorname{Im} DG(\overline{x})$ and $\operatorname{lin}(T_{\mathbb{S}^m_+}(G(\overline{x})))$ of $\mathbb{S}^m$ meet transversally, which is why it was originally called *transversality* in [26]. In mathematical language:

Definition 3.1 (Def. 4 from [24])**.** *A point $\overline{x} \in \mathcal{F}$ is said to satisfy the* nondegeneracy *condition when the following relation is satisfied:*

$$\operatorname{Im} DG(\overline{x}) + \operatorname{lin}(T_{\mathbb{S}^m_+}(G(\overline{x}))) = \mathbb{S}^m. \tag{5}$$

If $\overline{x}$ is a local solution of (NSDP), then nondegeneracy implies that $\Lambda(\overline{x})$ is a singleton; and the converse is also true in the presence of strict complementarity (see [25, Thm. 2.2 and Sect. 3]). Hence, Definition 3.1 is generally seen as an analogue of LICQ, from NLP, in NSDP. However, this analogy is tied to how the link between NLP and NSDP is made [24]. For example, when an NLP problem with constraints $g_1(x) \geqslant 0, \ldots, g_m(x) \geqslant 0$ is modelled as an NSDP with a single *structurally diagonal* conic constraint; that is, with G in the form

$$G(x) \doteq \begin{bmatrix} g_1(x) & & \\ & \ddots & \\ & & g_m(x) \end{bmatrix} \succeq 0; \tag{6}$$

then Definition 3.1 fails whenever there is some $\overline{Y} \in \Lambda(\overline{x})$ and some nonzero $H \in \mathbb{S}^m$ with only zeros in its diagonal, such that $H \succeq -\overline{Y}$, regardless of the linear independence of the set $\{\nabla g_1(\overline{x}), \ldots, \nabla g_m(\overline{x})\}$. In fact, structurally diagonal NSDP problems are in general expected to lack uniqueness of the Lagrange multiplier.

On the other hand, it is well-known (cf. [10, Sect. 4.6.1]) that a feasible point $\overline{x}$ satisfies the nondegeneracy condition if, and only if, either $\operatorname{Ker} G(\overline{x}) = \{0\}$ or the linear mapping $\psi_{\overline{x}} : \mathbb{R}^n \to \mathbb{S}^{m-r}$, defined by

$$\psi_{\overline{x}}(d) \doteq \overline{E}^\top DG(\overline{x})[d]\overline{E}, \tag{7}$$

is surjective for any $\overline{E}$ that spans $\operatorname{Ker} G(\overline{x})$. As a direct consequence of the equivalence above, it is possible to characterize Definition 3.1 as follows:

Proposition 3.1 (Prop. 6 from [24])**.** *Let $\overline{x} \in \mathcal{F}$ and let r denote the rank of $G(\overline{x})$. Then, $\overline{x}$ satisfies the nondegeneracy condition if, and only if, either* $\operatorname{Ker} G(\overline{x}) = \{0\}$ *or the vectors*

$$\begin{aligned} v_{ij}(\overline{x}, \overline{E}) &\doteq \left[\overline{e}_i^\top D_{x_1} G(\overline{x}) \overline{e}_j, \ldots, \overline{e}_i^\top D_{x_n} G(\overline{x}) \overline{e}_j \right]^\top \\ &= DG(\overline{x})^* \left[\frac{\overline{e}_i \overline{e}_j^\top + \overline{e}_j \overline{e}_i^\top}{2} \right], \end{aligned} \qquad 1 \leq i \leq j \leq m-r \tag{8}$$

are linearly independent, where $\overline{E} \in \mathbb{R}^{m \times m-r}$ is an arbitrary fixed matrix that spans $\operatorname{Ker} G(\overline{x})$*, and $\overline{e}_i$ denotes the i-th column of $\overline{E}$, for all $i \in \{1, \ldots, m-r\}$.*

Now, inspired by Proposition 3.1, we present a similar characterization of nondegeneracy that evaluates the linear independence of a narrower set of vectors at the cost of looking at all possible choices of $\overline{E}$ instead of a fixed one.

Proposition 3.2. *A point $\overline{x} \in \mathcal{F}$ satisfies nondegeneracy if, and only if, either* $\operatorname{Ker} G(\overline{x}) = \{0\}$ *or* $\left\{ v_{ii}(\overline{x}, \overline{E}) : i \in \{1, \ldots, m-r\} \right\}$ *is linearly independent for every matrix $\overline{E} \in \mathbb{R}^{m \times m-r}$ that spans* $\operatorname{Ker} G(\overline{x})$*, where $r = \operatorname{rank}(G(\overline{x}))$.*

Proof. Let us assume that $r < m$ since the result follows trivially otherwise. If there exists some matrix $\overline{E}$ that spans $\operatorname{Ker} G(\overline{x})$, such that $\left\{ v_{ii}(\overline{x}, \overline{E}) : i \in \{1, \ldots, m-r\} \right\}$ is linearly dependent, then the larger set $\left\{ v_{ij}(\overline{x}, \overline{E}) : 1 \leqslant i \leqslant j \leqslant m-r \right\}$ is also linearly dependent and it follows that $\overline{x}$ violates nondegeneracy.

Conversely, let B be an arbitrary matrix with orthonormal columns that span the kernel of $G(\overline{x})$ and consider the linear operator (7) defined in terms of B:

$$\psi_{\overline{x}}(d) \doteq B^\top DG(\overline{x})[d]B.$$

It suffices to prove that $\psi_{\overline{x}}$ is surjective or, equivalently, that its adjoint $\psi_{\overline{x}}^*$ given by

$$\psi_{\overline{x}}^*[Y] = \left[\left\langle B^\top D_{x_\ell} G(\overline{x}) B, Y \right\rangle \right]_{\ell \in \{1, \ldots, n\}}$$

is injective. Indeed, let $Y \in \mathbb{S}^{m-r}$ be such that $\psi_{\overline{x}}^*[Y] = 0$, and let $C \in \mathbb{R}^{m-r \times m-r}$ be an orthogonal matrix such that $C^\top Y C = \mathrm{Diag}(y_1, \ldots, y_{m-r})$, where $\mathrm{Diag}(y_1, \ldots, y_{m-r}) \in \mathbb{S}^{m-r}$ is a diagonal matrix whose i-th diagonal entry is y_i, with $i \in \{1, \ldots, m-r\}$. Then, it follows that

$$\begin{aligned} 0 &= \psi_{\overline{x}}^* \left[C\mathrm{Diag}(y_1, \ldots, y_{m-r}) C^\top \right] \\ &= \left[\left\langle B^\top D_{x_\ell} G(\overline{x}) B, C\mathrm{Diag}(y_1, \ldots, y_{m-r}) C^\top \right\rangle \right]_{\ell \in \{1,\ldots,n\}} \\ &= \left[\left\langle (BC)^\top D_{x_\ell} G(\overline{x}) BC, \mathrm{Diag}(y_1, \ldots, y_{m-r}) \right\rangle \right]_{\ell \in \{1,\ldots,n\}} . \end{aligned} \tag{9}$$

Set $\overline{E} \doteq BC$, which, since C is nonsingular, is also a matrix with orthonormal columns that spans $\mathrm{Ker}\, G(\overline{x})$, and denote its i-th column by $\overline{e}_i$. From (9), we obtain that

$$\sum_{i=1}^{m-r} y_i v_{ii}(\overline{x}, \overline{E}) = \left[\sum_{i=1}^{m} \left(\overline{e}_i^T D_{x_\ell} G(\overline{x}) \overline{e}_i \right) y_i \right]_{\ell \in \{1,\ldots,n\}} = 0. \tag{10}$$

By hypothesis, it follows that $y_i = 0$ for all $i \in \{1, \ldots, m-r\}$ and, hence, $Y = 0$. ■

The characterization of nondegeneracy from Proposition 3.2 may seem less practical than the one from Proposition 3.1, but it reveals a clear path for defining a weaker CQ by ruling out some particular choices of $\overline{E}$, which is the main result of the next subsection.

We mention that Wachsmuth [28] proved for NLPs that LICQ is equivalent to the uniqueness of the Lagrange multiplier for any objective function f (the unique multiplier may vary with f). Thanks to Proposition 3.2 this characterization can be straightforwardly extended to NSDP replacing LICQ by nondegeneracy, which we omit.

3.1 Sequences of eigenvectors and weak-nondegeneracy

In [8], Andreani et al. introduce a constructive technique for proving the existence of Lagrange multipliers for (NSDP), which is based on the so-called *sequential optimality conditions* from NLP [3]. The core idea of their proof is to apply an external penalty algorithm to (NSDP) after regularizing it around a given local minimizer, to obtain a sequence of approximate KKT points converging to it, as follows:

Theorem 3.1 (Thm. 3.2 from [8]). *Let $\overline{x}$ be a local minimizer of* (NSDP). *Then, for any sequence $\{\rho_k\}_{k\in\mathbb{N}} \to +\infty$, there exists some $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, such that for every $k \in \mathbb{N}$, x^k is a local minimizer of the regularized penalty function*

$$f(x) + \frac{1}{2}\|x - \overline{x}\|_2^2 + \frac{\rho_k}{2}\|\Pi_{\mathbb{S}_+^m}(-G(x))\|^2.$$

In particular, computing derivatives we obtain $\nabla_x L(x^k, Y^k) \to 0$, where $Y^k \doteq \rho_k \Pi_{\mathbb{S}_+^m}(-G(x^k))$.

With this result at hand, the authors prove that the sequence $\{Y^k\}_{k\in\mathbb{N}}$ must be bounded in the presence of Robinson's CQ, and that all of its limit points are Lagrange multipliers associated with $\overline{x}$ [8, Thm. 6.1]. Furthermore, the proof of this fact under nondegeneracy follows easily by contradiction: suppose that $\{Y^k\}_{k\in\mathbb{N}}$ is unbounded, and take any limit point $\overline{Y}$ of the sequence $\{Y^k/\|Y^k\|\}_{k\in\mathbb{N}}$; then:

1. It follows from $\nabla_x L(x^k, Y^k) \to 0$ that $DG(\overline{x})^*[\overline{Y}] = 0$, which means $\overline{Y} \in \mathrm{Ker}\, DG(\overline{x})^* = \mathrm{Im}\, DG(\overline{x})^\perp$;
2. By the definition of Y^k, we have $0 \neq \overline{Y} \succeq 0$ and $\langle G(\overline{x}), \overline{Y} \rangle = 0$, so $\overline{Y} \in \mathrm{lin}(T_{\mathbb{S}_+^m}(G(\overline{x})))^\perp$;

Hence, $\overline{Y} \in \mathrm{Im}\, DG(\overline{x})^\perp \cap \mathrm{lin}(T_{\mathbb{S}_+^m}(G(\overline{x})))^\perp$, which contradicts nondegeneracy.

With a single extra step, which is to take a spectral decomposition of Y^k for each k, the reasoning of the previous paragraph can be put in the same terms as Proposition 3.2. Indeed, observe that $\lambda_i(Y^k) = [\rho_k \lambda_i(-G(x^k))]_+ = 0$ for all $i \in \{m-r+1, \ldots, m\}$ and all k large enough, because

$$\lambda_i(-G(x^k)) = -\lambda_{m-i+1}(G(x^k)).$$

So

$$\nabla_x L(x^k, Y^k) = \nabla f(x^k) - \sum_{i=1}^{m-r} [\rho_k \lambda_i(-G(x^k))]_+ v_{ii}(x^k, E^k) \to 0,$$

where $E^k \in \mathbb{R}^{m\times m-r}$ is a matrix whose i-th column is $u_{m-i+1}(G(x^k))$. Then, note that if E^k can be chosen such that at least one of its limit points $\overline{E}$ ensures linear independence of $\{v_{ii}(\overline{x}, \overline{E}) : i \in \{1, \ldots, m-r\}\}$, then $\{Y^k\}_{k\in\mathbb{N}}$ must be bounded. Although the first clause of the previous sentence resembles nondegeneracy (as in Proposition 3.2), note that asking for the linear independence of the set $\{v_{ii}(\overline{x}, \overline{E}) : i \in \{1, \ldots, m-r\}\}$ when $\overline{E}$ is not a limit point of some sequence $\{E^k\}_{k\in\mathbb{N}}$ of eigenvectors of $G(x^k)$ seems unnecessary for defining a constraint qualification. This motivates us to propose a weaker variant of nondegeneracy as follows:

Definition 3.2 (Weak-nondegeneracy). *Let $\overline{x} \in \mathcal{F}$ and let r be the rank of $G(\overline{x})$. We say that* weak-nondegeneracy *holds at $\overline{x}$ if either $\mathrm{Ker}\, G(\overline{x}) = \{0\}$ or: for every sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there exists some sequence of matrices with orthonormal columns $\{E^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^{m\times m-r}$ such that:*

1. *The columns of E^k are eigenvectors associated with the $m-r$ smallest eigenvalues of $G(x^k)$, for each $k \in \mathbb{N}$;*
2. *There exists a limit point $\overline{E}$ of $\{E^k\}_{k\in\mathbb{N}}$ such that the set $\left\{v_{ii}(\overline{x},\overline{E})\colon i \in \{1,\ldots,m-r\}\right\}$, as defined in* (8)*, is linearly independent.*

There are a couple of nuances about Definition 3.2 that should be properly addressed (see also the discussion after Proposition 3.3). First, we recall that the eigenvector functions $u_i(G(x))$, $i \in \{m-r+1,\ldots,m\}$ are not necessarily continuous at a given point $\overline{x}$; so weak-nondegeneracy relies on the "sequential continuity of eigenvectors" along a given path. Second, for any fixed $\overline{x} \in \mathcal{F}$ and any $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, the sequence $\{E^k\}_{k\in\mathbb{N}}$ described in Definition 3.2 is well-defined for k sufficiently large, since the r largest eigenvalues of $G(x^k)$ are necessarily bounded away from zero.

The discussion that motivated Definition 3.2 already suggests that weak-nondegeneracy is indeed a genuine constraint qualification, and it also provides an outline of how to prove it. Nevertheless, we state and prove this fact with appropriate mathematical rigor below.

Theorem 3.2. *Every local minimizer $\overline{x} \in \mathcal{F}$ of* (NSDP) *that satisfies weak-nondegeneracy also satisfies the KKT conditions.*

Proof. Let $\overline{x}$ be a local minimizer of (NSDP) that satisfies weak-nondegeneracy and let $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and $\{Y^k\}_{k\in\mathbb{N}}$ be the sequences described in Theorem 3.1, for an arbitrary sequence $\{\rho_k\}_{k\in\mathbb{N}} \to \infty$. If $r=m$, set $\overline{Y}=0$ as a Lagrange multiplier associated with $\overline{x}$ and we are done; so let us assume that $r<m$ from now on. From the local optimality of x^k, for each $k \in \mathbb{N}$, we obtain

$$\nabla f(x^k) + (x^k - \overline{x}) + DG(x^k)^*[Y^k] = 0. \tag{11}$$

Recall that we assume, without loss of generality, that $\lambda_1(-G(x^k)) \geqslant \ldots \geqslant \lambda_m(-G(x^k))$, for every k; and note that when k is large enough, say greater than some $k_0 \in \mathbb{N}$, we necessarily have $\lambda_i(-G(x^k)) < 0$ for all $i \in \{m-r+1,\ldots,m\}$ since $G(x^k) \to G(\overline{x})$ and eigenvalues $\lambda_i(\cdot)$ are continuous mappings. Then, for each $k > k_0$, we have

$$Y^k = \sum_{i=1}^{m-r} \alpha_i^k e_i^k (e_i^k)^\top,$$

where $\alpha_i^k \doteq [\rho_k \lambda_i(-G(x^k))]_+$ and $e_i^k \doteq u_{m-i+1}(G(x^k))$ is an arbitrary unitary eigenvector associated with $\lambda_{m-i+1}(G(x^k))$, for each $i \in \{1,\ldots,m-r\}$. Set $E^k \doteq [e_1^k,\ldots,e_{m-r}^k]$. Since $\{E^k\}_{k\in\mathbb{N}}$ is bounded, we may pick any of its limit points $\overline{E} = [\overline{e}_1,\ldots,\overline{e}_{m-r}]$ and assume, taking a subsequence if necessary, that it converges to $\overline{E}$, which spans $\operatorname{Ker} G(\overline{x})$. Then, observe that (11) implies

$$\nabla f(x^k) - \sum_{i=1}^{m-r} \alpha_i^k DG(x^k)^*[e_i^k (e_i^k)^\top] \to 0,$$

but since $DG(x^k)^*[e_i^k(e_i^k)^\top] = v_{ii}(x^k, E^k)$ (see (8)), we can rewrite it as

$$\nabla f(x^k) - \sum_{i=1}^{m-r} \alpha_i^k v_{ii}(x^k, E^k) \to 0. \tag{12}$$

If $\left\{(\alpha_i^k,\ldots,\alpha_{m-r}^k)\right\}_{k\in\mathbb{N}}$ has any convergent subsequence, denote its limit point by $\overline{\alpha} \doteq (\overline{\alpha}_i,\ldots,\overline{\alpha}_{m-r})$, and note that $\overline{\alpha}$ generates a Lagrange multiplier for $\overline{x}$, which is $\overline{Y} \doteq \sum_{i=1}^{m-r} \overline{\alpha}_i \overline{e}_i \overline{e}_i^\top$. Hence, it suffices to prove that $\{\alpha_i^k\}_{k\in\mathbb{N}}$, $i \in \{1,\ldots,m-r\}$, must be bounded under weak-nondegeneracy. Let us assume for a moment that the sequences $\{\alpha_i^k\}_{k\in\mathbb{N}}$ are unbounded, which means

$$m^k \doteq \max\left\{\alpha_i^k \colon i \in \{1,\ldots,m-r\}\right\} \to \infty.$$

Note that $\left\{(\alpha_1^k,\ldots,\alpha_{m-r}^k)/m^k\right\}_{k\in\mathbb{N}}$ must be bounded and it must also have a nonzero limit point, which we will denote by $(\tilde{\alpha}_1,\ldots,\tilde{\alpha}_{m-r})$. We assume without loss of generality, that $\left\{(\alpha_1^k,\ldots,\alpha_{m-r}^k)/m^k\right\}_{k\in\mathbb{N}} \to (\tilde{\alpha}_1,\ldots,\tilde{\alpha}_{m-r})$. After dividing (12) by m^k for each k and taking limit $k \to +\infty$, we obtain

$$\sum_{i=1}^{m-r} \tilde{\alpha}_i v_{ii}(\overline{x},\overline{E}) = 0,$$

which means $\left\{v_{ii}(\overline{x},\overline{E})\colon i \in \{1,\ldots,m-r\}\right\}$ is linearly dependent. However, since our analyses hold for any arbitrary choice of $\{E^k\}_{k\in\mathbb{N}}$ and any $\overline{E}$, this contradicts weak-nondegeneracy. ■

Let us briefly analyse a direct application of weak-nondegeneracy: As an intermediary step of the proof of Theorem 3.2, we proved that every feasible limit point of a sequence described in Theorem 3.1 must satisfy the KKT conditions under weak-nondegeneracy. However, the sequences $\{x^k\}_{k\in\mathbb{N}}$ and $\{Y^k\}_{k\in\mathbb{N}}$ described in Theorem 3.2 are precisely the ones that are generated by a standard external penalty method (that is, [8, Algorithm 1] with the parameter Ω^{max} fixed at zero). Thus, every feasible limit point of the external penalty method that satisfies weak-nondegeneracy must also satisfy the KKT conditions.

Another interesting property of weak-nondegeneracy is that it is equivalent to LICQ when G is a structurally diagonal matrix function (as in (6)) that models an NLP problem, which in some sense resolves the inconsistency between nondegeneracy and LICQ noted by Shapiro [24, Page 309].

Proposition 3.3. *When G is structurally diagonal, as in* (6)*, then $\overline{x} \in \mathcal{F}$ satisfies weak-nondegeneracy if, and only if, the set $\{\nabla g_i(\overline{x})\colon g_i(\overline{x}) = 0\}$ is linearly independent.*

Proof. If $r = m$, the result follows trivially, so let us assume that $r < m$. Also, suppose that $\{i \in \{1, \dots, m\}\colon g_i(\overline{x}) = 0\} = \{r + 1, \dots, m\}$, where r is the rank of $G(\overline{x})$. Clearly, if $\{\nabla g_{r+1}(\overline{x}), \dots, \nabla g_m(\overline{x})\}$ is linearly independent, then we may take

$$E^k \doteq \begin{bmatrix} 0 \\ \mathbb{I}_{m-r} \end{bmatrix} \in \mathbb{R}^{m \times m-r}$$

for all sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ to conclude that $\overline{x}$ satisfies weak-nondegeneracy. Conversely, suppose that weak-nondegeneracy holds at $\overline{x}$, take any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and any $\{E^k\}_{k\in\mathbb{N}} \to \overline{E} \doteq [\overline{e}_i, \dots, \overline{e}_{m-r}]$ such that $\left\{v_{ii}(\overline{x}, \overline{E})\colon i \in \{1, \dots, m - r\}\right\}$ is linearly independent. Note that $\overline{E}$ must have the form

$$\overline{E} = \begin{bmatrix} 0 \\ Q \end{bmatrix}, \text{ where } Q \in \mathbb{R}^{m-r \times m-r} \text{ is orthonormal,}$$

due to the diagonal structure of G and the fact that $g_i(\overline{x}) \neq 0$ for all $i \in \{1, \dots, r\}$. Hence,

$$v_{ii}(\overline{x}, \overline{E}) = \sum_{j=r+1}^{m} \nabla g_j(\overline{x}) Q_{i,j-r}^2 = Dg(\overline{x})^\top (Q_i \odot Q_i), \tag{13}$$

where $Dg(\overline{x})$ is the Jacobian matrix of $g(x) \doteq (g_{r+1}(x), \dots, g_m(x))$ at $\overline{x}$, the operator $\odot$ is the (Hadamard) entry-wise vector product, and Q_i is the i-th column of Q, with $i \in \{1, \dots, m - r\}$. Then,

$$\operatorname{span}\left\{v_{ii}(\overline{x}, \overline{E})\colon i \in \{1, \dots, m - r\}\right\} \subseteq \operatorname{Im} Dg(\overline{x})^\top$$

and, consequently,

$$\begin{aligned} m - r &= \dim(\operatorname{span}\left\{v_{ii}(\overline{x}, \overline{E})\colon i \in \{1, \dots, m - r\}\right\}) \\ &\leqslant \dim(\operatorname{Im} Dg(\overline{x})^\top) \\ &= \operatorname{rank}(Dg(\overline{x})^\top) \leqslant m - r \end{aligned}$$

Hence, $\operatorname{rank}(Dg(\overline{x})^\top) = m - r$, which means that $\{\nabla g_{r+1}(\overline{x}), \dots, \nabla g_m(\overline{x})\}$ is linearly independent. ■

In light of Proposition 3.3, let us go back to Definition 3.2 for a moment, to clarify that weak-nondegeneracy is not equivalent to requiring linear independence of $\left\{v_{ii}(\overline{x}, \overline{E})\colon i \in \{1, \dots, m - r\}\right\}$ for all $\overline{E}$ in the set

$$\mathcal{B}(\overline{x}) \doteq \left\{\overline{E} \in \mathbb{R}^{m\times m-r}\colon \exists\{x^k\}_{k\in\mathbb{N}} \to \overline{x}, \exists\{E^k\}_{k\in\mathbb{N}} \to \overline{E},\ x^k \neq \overline{x},\ \text{item 1 of Def. 3.2 holds}\right\}.$$

That is, we do not require linear independence of $\{v_{ii}(\overline{x}, \overline{E})\colon i \in \{1, \dots, m - r\}\}$ for all matrices $\overline{E}$ that are limits of matrices E^k of eigenvectors of $G(x^k)$. This is relevant when the eigenvalues of $G(x^k)$ are non-simple, and thus E^k is not uniquely defined. The following example illustrates that.

Example 3.1. *Consider the following constraint:*

$$G(x) \doteq \begin{bmatrix} x_1 & 0 & 0 \\ 0 & x_2 & 0 \\ 0 & 0 & x_3 \end{bmatrix} \succeq 0$$

and the point $\overline{x} \doteq 0$. Take any sequence $\{x^k\}_{k\in\mathbb{N}} \doteq \{(x_1^k, 0, 0)\}_{k\in\mathbb{N}} \to \overline{x}$ with $x_1^k \neq 0$ for every k, and note that

$$\overline{E} \doteq E^k \doteq \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix},$$

satisfies item 1 of Definition 3.2, so $\overline{E} \in \mathcal{B}(\overline{x})$, but $v_{22}(\overline{x}, \overline{E}) = v_{33}(\overline{x}, \overline{E}) = (0, 1/2, 1/2)$, and then they are, of course, linearly dependent. However, $\overline{x}$ satisfies weak-nondegeneracy due to Proposition 3.3 and, in fact, given any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, one should pick $E^k \doteq \mathbb{I}_3$ for all $k \in \mathbb{N}$ to verify that weak-nondegeneracy holds.

It is also worth mentioning that weak-nondegeneracy does not guarantee uniqueness of multipliers, even when they are constrained to be in the form $\overline{E}\tilde{Y}\overline{E}^\top$ for some $\overline{E} \in \mathcal{B}(\overline{x})$. For instance, consider the example above with a constant objective function $f(x) \doteq 0$; then, 0 is a Lagrange multiplier associated with $\overline{x}$, but so is $\overline{E}\mathrm{Diag}(0, -1, 1)\overline{E}^\top \neq 0$.

It is clear from Proposition 3.2 that weak-nondegeneracy is implied by nondegeneracy; and we see in Example 3.1 that the converse is not true since there is no uniqueness of Lagrange multipliers. Moreover, note that weak-nondegeneracy imposes a less demanding dimensional constraint over (NSDP); in fact, in order to verify nondegeneracy, one must have $n \geqslant (m-r)(m-r+1)/2$, while weak-nondegeneracy may hold as long as $n \geqslant m-r$ (Proposition 3.3).

Remark 3.1. *If we replace the sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ by matrix sequences $\{M^k\}_{k\in\mathbb{N}} \to G(\overline{x})$ in Definition 3.2, then we recover the nondegeneracy condition. Indeed, for any $\overline{E} \in \mathbb{R}^{m\times m-r}$ that spans $\operatorname{Ker} G(\overline{x})$, consider*

$$M^k \doteq \overline{U}\Lambda^k\overline{U}^\top, \quad \text{with} \quad \overline{U} \doteq [\overline{E}, u_{m-r+1}(G(\overline{x})), \ldots, u_m(G(\overline{x}))],$$

and $\Lambda^k \doteq \operatorname{Diag}(y^k)$ such that $y_i^k \doteq i/k$ for $i \in \{1,\ldots,m-r\}$, and $y_i^k \doteq \lambda_i(G(\overline{x}))$ otherwise. So, clearly $M^k \to G(\overline{x})$ and the only convergent sequence E^k to $\overline{E}$ is $\overline{E}$ itself. Consequently, when we assume Definition 3.2 it necessarily follows that $\{v_{ii}(\overline{x},\overline{E})\colon i \in \{1,\ldots,m-r\}\}$ is linearly independent. Then, since $\overline{E}$ was chosen arbitrary, Proposition 3.2 implies that nondegeneracy holds true.

Remark 3.2. *Proposition 3.3 can be straightforwardly extended to structurally block diagonal matrix constraints, such as*

$$G(x) \doteq \begin{bmatrix} G_1(x) & & \\ & \ddots & \\ & & G_q(x) \end{bmatrix} \succeq 0, \qquad \text{(Block-NSDP)}$$

where each "block" is defined by a continuously differentiable function $G_\ell\colon \mathbb{R}^n \to \mathbb{S}^{m_\ell}$, with $\ell \in \{1,\ldots,q\}$, and $m_1 + \ldots + m_q = m$. In fact, let $\overline{x} \in \mathcal{F}$ and $r_\ell \doteq \operatorname{rank}(G_\ell(\overline{x}))$ for each ℓ; and, for simplicity, let us assume that $r_\ell < m_\ell$ for all ℓ. Since $\operatorname{Ker} G(\overline{x}) = \operatorname{Ker} G_1(\overline{x}) \times \ldots \times \operatorname{Ker} G_q(\overline{x})$, then $r = r_1 + \ldots + r_q$. Then, weak-nondegeneracy holds at $\overline{x}$ if, and only if, for all sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there are sequences of matrices $\{E_\ell^k\}_{k\in\mathbb{N}}$ such that:

- *The columns of E_ℓ^k are unitary eigenvectors associated with the $m_\ell - r_\ell$ smallest eigenvalues of $G_\ell(x^k)$, for each $k \in \mathbb{N}$ and each $\ell \in \{1,\ldots,q\}$;*
- *There are limit points $\overline{E}_\ell$ of $\{E_\ell^k\}_{k\in\mathbb{N}}$, $\ell \in \{1,\ldots,q\}$, such that the set*

$$\bigcup_{\ell=1}^{q} \left\{ v_{ii}^\ell(\overline{x},\overline{E}_\ell)\colon i \in \{1,\ldots,m_\ell - r_\ell\} \right\}$$

  is linearly independent, where

$$v_{ij}^\ell(\overline{x},\overline{E}_\ell) \doteq \left[\overline{e}_{\ell,i}^\top D_{x_1} G_\ell(\overline{x})\overline{e}_{\ell,j} \ , \ \ldots \ , \ \overline{e}_{\ell,i}^\top D_{x_n} G_\ell(\overline{x})\overline{e}_{\ell,j}\right]^\top, \qquad (14)$$

  and $\overline{e}_{\ell,1},\ldots,\overline{e}_{\ell,m_\ell-r_\ell}$ denote the columns of $\overline{E}_\ell$, for each $\ell \in \{1,\ldots,q\}$.

The proof of this fact is elementary with [14, Lem. 1.3.10] and (8) at hand. Moreover, note that this is precisely the way weak-nondegeneracy would be defined for an equivalent multifold NSDP with constraints $G_1(x) \succeq 0, \ldots, G_q(x) \succeq 0$. Thus, weak-nondegeneracy is invariant to block diagonal and multifold representations of (Block-NSDP).

4 Dealing with structural sparsity

In this section, we take inspiration from a regularity condition introduced by Forsgren [12, Sect. 2.3], whose primary goal was to prove second-order optimality conditions for (NSDP). However, what makes Forsgren's condition specially interesting for us is the fact it can benefit from some sparsity structure of a certain Schur complement related to the constraint function. The main objective of this section is to present a more straightforward way of enjoying sparsity, based on Forsgren's results and Section 3. But before that, we present some of the notation used by Forsgren.

Given a point $\overline{x}$ and a matrix-valued function $F\colon \mathbb{R}^n \to \mathbb{S}^\beta$, consider the set $\mathcal{S}(F,\overline{x})$ defined as follows:

$$\begin{aligned} \mathcal{S}(F,\overline{x}) &\doteq \left\{M \in \mathbb{S}^\beta\colon M_{ij} = 0 \text{ if } F_{ij}(x) \text{ is structurally zero near } \overline{x}\right\} \\ &= \left\{M \in \mathbb{S}^\beta\colon M_{ij} = 0 \text{ if } \exists \varepsilon > 0 \text{ such that } F_{ij}(x) = 0, \forall x \in B(\overline{x},\varepsilon)\right\}. \end{aligned}$$

For example, if $\beta = 3$ and for all x close to $\overline{x}$, we are able to identify non trivial mappings F_{ij} such that

$$F(x) = \begin{bmatrix} F_{11}(x) & 0 & F_{13}(x) \\ 0 & F_{22}(x) & 0 \\ F_{13}(x) & 0 & F_{33}(x) \end{bmatrix}, \text{ then } M \in \mathcal{S}(F,\overline{x}) \Leftrightarrow M = \begin{bmatrix} M_{11} & 0 & M_{13} \\ 0 & M_{22} & 0 \\ M_{13} & 0 & M_{33} \end{bmatrix},$$

where M_{11}, M_{13}, M_{22}, and M_{33} may or may not be zero. Also, we define

$$\mathcal{I}(F,\overline{x}) \doteq \{(i,j)\colon \forall\varepsilon > 0, \exists x \in B(\overline{x},\varepsilon) \text{ such that } F_{ij}(x) \neq 0, \ 1 \leqslant i \leqslant j \leqslant \beta\}$$

as the set of indices that define the elements of $\mathcal{S}(F,\overline{x})$.

Forsgren's results are obtained in terms of the function

$$\tilde{G}(x) \doteq G(x) - G(x)\overline{P}(\overline{P}^\top G(x)\overline{P})^{-1}\overline{P}^\top G(x),$$

where $\overline{U} = [\overline{P}, \overline{E}]$ has columns that form an orthonormal eigenvector basis for $G(\overline{x})$, such that $\overline{E}$ spans the kernel of $G(\overline{x})$ and $\overline{P}^\top G(\overline{x})\overline{P} \succ 0$. Note that $\overline{E}^\top \tilde{G}(x)\overline{E}$ is the Schur complement of $\overline{P}^\top G(x)\overline{P}$ inside

$$\overline{U}^\top G(x)\overline{U} = \begin{bmatrix} \overline{P}^\top G(x)\overline{P} & \overline{P}^\top G(x)\overline{E} \\ \overline{E}^\top G(x)\overline{P} & \overline{E}^\top G(x)\overline{E} \end{bmatrix}.$$

Moreover, following Forsgren [12, Lem. 1], we see that $\tilde{G}(x) \succeq 0$ if, and only if $G(x) \succeq 0$, for all x sufficiently close to $\overline{x}$, so the original NSDP problem can be locally reformulated as a minimization problem over $\tilde{G}(x) \succeq 0$, around $\overline{x}$. In fact, since

$$\begin{aligned} \overline{P}(\overline{P}^\top G(\overline{x})\overline{P})^{-1}\overline{P}^\top &= \overline{P}\lambda_+(G(\overline{x}))^{-1}\overline{P}^\top \\ &= \overline{U} \begin{bmatrix} \lambda_+(G(\overline{x}))^{-1} & 0 \\ 0 & 0 \end{bmatrix} \overline{U}^\top \\ &= G(\overline{x})^\dagger, \end{aligned}$$

where $G(\overline{x})^\dagger$ is the Moore-Penrose pseudoinverse of $G(\overline{x})$, it follows that $\tilde{G}(\overline{x}) = 0$ [12, Lem. 2], so $\tilde{G}$ can be considered a reduction to the kernel of $G(\overline{x})$ near $\overline{x}$.

The regularity condition introduced by Forsgren is as follows:

Definition 4.1 (Forsgren's CQ)**.** *Let $\overline{x} \in \mathcal{F}$ and let $\overline{U} \doteq [\overline{P}, \overline{E}]$ be an orthogonal matrix that diagonalizes $G(\overline{x})$, such that the columns of $\overline{E}$ span* $\operatorname{Ker} G(\overline{x})$*. Then,* Forsgren's CQ *holds at $\overline{x}$ with respect to $\overline{U}$ when*

$$\operatorname{span}\left\{\overline{E}^\top D_{x_i}G(\overline{x})\overline{E} \colon i \in \{1, \ldots, n\}\right\} = \overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x})\overline{E} \tag{F1}$$

and

$$\exists M \in \overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x})\overline{E}, \quad \text{such that} \quad M \succ 0. \tag{F2}$$

Forsgren's CQ is indeed a constraint qualification, for when (F1) holds, then (F2) is equivalent to Robinson's CQ [12, Lem. 5]. However, although Forsgren states that any choice of $\overline{U}$ leads to a valid CQ, there is no discussion on the effects of this choice over the condition proposed. Under a specific condition, Forsgren's CQ provides uniqueness of the Lagrange multiplier [12, Thm. 1], but this condition varies with $\overline{U}$. Thus, different choices of $\overline{U}$ are likely to define different variants of Forsgren's CQ. This is not necessarily a negative point, but a comparison among those variants would be appropriate. For instance, from the practical point of view, one may be interested in knowing which choice of $\overline{U}$ defines the weakest CQ, or which one is easier to compute.

A result from Dorsch, Gómez, and Shikhman [11] shows that, ignoring the sparsity treatment, (F1) becomes equivalent to nondegeneracy.

Lemma 4.1 (Lem. 5 from [11])**.** *Let $\overline{x} \in \mathcal{F}$ and assume that $\mathcal{S}(\tilde{G}, \overline{x}) = \mathbb{S}^m$. Then, condition* (F1) *of Forsgren's CQ holds if, and only if, nondegeneracy holds at $\overline{x}$.*

However, similarly to weak-nondegeneracy, Forsgren's CQ also reduces to LICQ from NLP when G is structurally diagonal (as in (6)), contrasting with nondegeneracy. To put Forsgren's CQ in the same terms as the previous sections, we present an elementary characterization of it using the vectors $v_{ij}(\overline{x}, \overline{E})$ defined in Proposition 3.1:

Proposition 4.1. *Let $\overline{x} \in \mathcal{F}$ and let $\overline{E} \in \mathbb{R}^{m \times m-r}$ span* $\operatorname{Ker} G(\overline{x})$*. Then, condition* (F1) *of Forsgren's CQ holds at $\overline{x}$ if, and only if,*

$$\sum_{i=1}^{m-r} \sum_{j=i}^{m-r} M_{ij} v_{ij}(\overline{x}, \overline{E}) = 0, \quad M \in \overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x})\overline{E} \quad \Rightarrow \quad M = 0,$$

where $r = \operatorname{rank}(G(\overline{x}))$*.*

Proof. Let us assume that $r < m$, since otherwise the proof is trivial. We employ [12, Lem. 2], which states that $\overline{E}^\top D_{x_i}G(\overline{x})\overline{E} = \overline{E}^\top D_{x_i}\tilde{G}(\overline{x})\overline{E}$ for all $i \in \{1, \ldots, n\}$, to ensure that the linear operator $\psi : \mathbb{R}^n \to \overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x})\overline{E}$, defined by the action $\psi(d) \doteq \overline{E}^\top DG(\overline{x})[d]\overline{E}$ is well-defined.

With this in mind, note that

$$\operatorname{Im}(\psi) = \operatorname{span}\left\{\overline{E}^\top D_{x_i}G(\overline{x})\overline{E} \colon i \in \{1, \ldots, n\}\right\} = \overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x})\overline{E}$$

if, and only if,

$$\operatorname{Ker}(\psi^*) = \left\{M \in \overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x})\overline{E} \colon \langle \overline{E}^\top D_{x_\ell}G(\overline{x})\overline{E}, M\rangle = 0, \ \forall \ell \in \{1, \ldots, n\}\right\} = \{0\}, \tag{15}$$

whence the result follows since

$$\overline{E}^\top D_{x_\ell}G(\overline{x})\overline{E} = [(v_{ij}(\overline{x}, \overline{E}))_\ell]_{i,j \in \{1, \ldots, m-r\}},$$

where $(v_{ij}(\overline{x}, \overline{E}))_\ell$ is the ℓ-th entry of the vector $v_{ij}(\overline{x}, \overline{E})$. ∎

As an abuse of language, (F1) consists of the "linear independence" of $\{v_{ij}(\overline{x},\overline{E})\colon 1 \leqslant i \leqslant j \leqslant m-r\}$ with respect to the set $\overline{E}^\top \mathcal{S}(\tilde{G},\overline{x})\overline{E}$. In particular, when $G(\overline{x})=0$ and (F2) holds, take $\overline{U}=\overline{E}=\mathbb{I}_m$ and note that Forsgren's CQ holds for this particular choice of $\overline{U}$ if, and only if, the set $\{\nabla G_{ij}(\overline{x})\colon (i,j)\in\mathcal{I}(G,\overline{x})\}$ is linearly independent, with $(i,i)\in\mathcal{I}(G,\overline{x})$ for all $i\in\{1,\ldots,m\}$.

Remark 4.1. *As far as we understand, the relation between Forsgren's CQ and nondegeneracy was not formally established in [12]. To clarify this important detail, note that it is clear from Propositions 4.1 and 3.1 that nondegeneracy implies Forsgren's CQ. Moreover, Example 3.1 shows that this implication is strict.*

The above discussion leads us to deal with sparsity in a more straightforward way, namely without taking Schur complements, which induces another weak variant of nondegeneracy.

4.1 A sparse variant of nondegeneracy

For any matrix $\overline{E}$ that spans $\operatorname{Ker} G(\overline{x})$, consider the function

$$\widehat{G}^{\overline{E}}(x) \doteq \overline{E}^\top G(x)\overline{E}$$

and note that $\nabla \widehat{G}^{\overline{E}}_{ij}(\overline{x}) = v_{ij}(\overline{x},\overline{E})$ for all $i,j\in\{1,\ldots,m-r\}$ with $i\leqslant j$. We incorporate structural sparsity into nondegeneracy directly, but in a similar style of Forsgren's CQ (as characterized in Proposition 4.1), to introduce a new constraint qualification.

Definition 4.2 (Sparse-nondegeneracy). *We say that* sparse-nondegeneracy *holds at $\overline{x}\in\mathcal{F}$ when either* $\operatorname{Ker} G(\overline{x})=\{0\}$ *or there exists a matrix $\overline{E}\in\mathbb{R}^{m\times m-r}$ that spans* $\operatorname{Ker} G(\overline{x})$ *and such that:*

1. *The set $\left\{v_{ij}(\overline{x},\overline{E})\colon (i,j)\in\mathcal{I}(\widehat{G}^{\overline{E}},\overline{x}), 1\leqslant i\leqslant j\leqslant m-r\right\}$ is linearly independent;*
2. *$(i,i)\in\mathcal{I}(\widehat{G}^{\overline{E}},\overline{x})$ for all $i\in\{1,\ldots,m-r\}$.*

There are two natural questions about sparse-nondegeneracy that we shall answer in the following paragraphs. The first one consists of knowing whether the sparse-nondegeneracy condition is a genuine constraint qualification; and the second one concerns about the relation between Definition 4.2 and other constraint qualifications, such as nondegeneracy, Forsgren's CQ, and Robinson's CQ. To address these questions, we first prove an elementary characterization of sparse-nondegeneracy:

Lemma 4.2. *Let $\overline{x}\in\mathcal{F}$ be such that $\operatorname{Ker} G(\overline{x})\neq\{0\}$, and let $\overline{E}$ be a matrix that spans $\operatorname{Ker} G(\overline{x})$. Then, item 1 of Definition 4.2 holds at $\overline{x}$ if, and only if, there is no nonzero $\tilde{Y}\in\mathcal{S}(\widehat{G}^{\overline{E}},\overline{x})$ such that $DG(\overline{x})^*[\overline{E}\tilde{Y}\overline{E}^\top]=0$.*

Proof. The result follows directly by noticing that

$$\sum_{(i,j)\in\mathcal{I}(\widehat{G}^{\overline{E}},\overline{x})} v_{ij}(\overline{x},\overline{E})\tilde{Y}_{ij} = DG(\overline{x})^*[\overline{E}\tilde{Y}\overline{E}^\top] \tag{16}$$

for every $\tilde{Y}\in\mathcal{S}(\widehat{G}^{\overline{E}},\overline{x})$. ■

Next, we prove that sparse-nondegeneracy implies Robinson's CQ, which also shows that it is indeed a constraint qualification.

Proposition 4.2. *If $\overline{x}\in\mathcal{F}$ satisfies sparse-nondegeneracy, then it also satisfies Robinson's CQ.*

Proof. The result follows trivially when $\operatorname{Ker} G(\overline{x})=\{0\}$, so let us assume that $r=\operatorname{rank}(G(\overline{x}))<m$. Suppose that sparse-nondegeneracy holds at $\overline{x}\in\mathcal{F}$, and take any $Z\succeq 0$ such that $\langle Z,G(\overline{x})\rangle=0$ and $DG(\overline{x})^*[Z]=0$, then there exists some $Y\in\mathbb{S}^{m-r}_+$ such that $Z=\overline{E}Y\overline{E}^\top$. Define the matrix $\tilde{Y}\in\mathcal{S}(\widehat{G}^{\overline{E}},\overline{x})$ whose (i,j)-th entry is given by

$$\tilde{Y}_{ij} \doteq \begin{cases} Y_{ij}, & \text{if } (i,j)\in\mathcal{I}(\widehat{G}^{\overline{E}},\overline{x}) \\ 0, & \text{otherwise,}\end{cases}$$

and note that

$$DG(\overline{x})^*[Z] = DG(\overline{x})^*[\overline{E}Y\overline{E}^\top] = DG(\overline{x})^*[\overline{E}\tilde{Y}\overline{E}^\top]=0, \tag{17}$$

so $\tilde{Y}=0$ due to Lemma 4.2. Moreover, from item 2 of Definition 4.2, $(i,i)\in\mathcal{I}(\widehat{G}^{\overline{E}},\overline{x})$ for all $i\in\{1,\ldots,m-r\}$, so the diagonal of Y must consist only of zeros, which implies that $Y=0$ and, consequently, $Z=0$. Since Z is arbitrary, Robinson's CQ holds. ■

We highlight that item 2 of Definition 4.2 is not superfluous, for removing it may cause us to lose the property of being a constraint qualification. Indeed, the following example illustrates that:

Example 4.1. *Consider the problem:*

$$\begin{aligned} &\underset{x\in\mathbb{R}^2}{\text{Minimize}} && x_2, \\ &\text{subject to} && G(x) \doteq \begin{bmatrix} x_1 & x_2 \\ x_2 & 0 \end{bmatrix} \succeq 0, \end{aligned}$$

which has $\overline{x} \doteq (0,0)$ as one of its solutions. The point $\overline{x}$ satisfies Definition 4.2 after removing item 2, with $\overline{E} \doteq \mathbb{I}_2$, because $v_{11}(\overline{x},\overline{E}) = (1,0)$ and $v_{12}(\overline{x},\overline{E}) = (0,1)$ are linearly independent; but $\overline{x}$ does not satisfy the KKT conditions since there is no $\overline{Y} \succeq 0$ such that $\overline{Y}_{11} = 0$ and $\overline{Y}_{12} = \overline{Y}_{21} = 1/2$. Thus, Definition 4.2 is not a constraint qualification without item 2.

Remark 4.2. *Let us show that when item 2 fails, the problem can be reformulated such that it holds. Let $\overline{x} \in \mathcal{F}$ and $\overline{E}$ be a matrix that spans $\operatorname{Ker} G(\overline{x})$. If item 2 of Definition 4.2 is not satisfied, then let $J \doteq \{i \in \{1,\dots,m-r\} \colon (i,i) \notin \mathcal{I}(\widehat{G}^{\overline{E}},\overline{x})\}$ and note that there exists some $\varepsilon > 0$ such that*

$$G(x) \in \mathbb{S}^m_+ \text{ if, and only if, } G(x) \in \mathbb{S}^m_+ \bigcap_{i\in J} \{\overline{e}_i\overline{e}_i^\top\}^\perp,$$

for every $x \in B(\overline{x},\varepsilon)$, where $\overline{e}_i$ denotes the i-th column of $\overline{E}$. That is, the feasible set $\mathcal{F}$ coincides locally with the preimage of the face $F \doteq \mathbb{S}^m_+ \bigcap_{i\in J}\{\overline{e}_i\overline{e}_i^\top\}^\perp$ of $\mathbb{S}^m_+$. Moreover, since F is a face of $\mathbb{S}^m_+$, then there is an orthogonal matrix $V \doteq [V_1, V_2] \in \mathbb{R}^{m\times m}$ such that

$$V^\top F V = \left\{ \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix} : M \in \mathbb{S}^{m-\omega}_+ \right\},$$

where ω is the cardinality of J [21, Eq. 2.3]. This means that it is possible to locally replace the original constraint of (NSDP) *by the equality constraint $V_2^\top G(x) = 0$ and a smaller semidefinite constraint $\mathcal{G}(x) \doteq V_1^\top G(x) V_1 \in \mathbb{S}^{m-\omega}_+$. If F is minimal, then the new constraint $\mathcal{G}(x) \in \mathbb{S}^{m-\omega}_+$ satisfies item 2 of Definition 4.2 at $\overline{x}$. Otherwise, this process can be repeated until the minimal face is reached. Thus, every problem can be equivalently reformulated (reducing dimension if necessary), such that item 2 always holds. In particular, when G is an affine function, then this procedure can be computed via a popular preprocessing technique called* facial reduction *(we refer to Pataki [21] and references therein for more details about it). When $G(\overline{x}) = 0$ and $\overline{E} = \mathbb{I}_m$, this procedure can be done by simply removing the i-th row and the i-th column of G, for every i such that $(i,i) \notin \mathcal{I}(G,\overline{x})$, and including the correspondent equality constraints into the problem. We recall that all of our results can be easily extended to NSDP problems with separate equality constraints.*

Let us illustrate this procedure using Example 4.1. In this case we have $\overline{e}_2^\top G(x)\overline{e}_2 = 0$ for every x; then $x \in \mathcal{F}$ if, and only if, $G(x) \in F$, where

$$F \doteq \mathbb{S}^2_+ \bigcap \left\{ \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}^\perp = \left\{ \begin{bmatrix} \alpha & 0 \\ 0 & 0 \end{bmatrix} : \alpha \geqslant 0, \right\}$$

which means that the constraint of the problem can be equivalently written as $x_2 = 0$ and $x_1 \geqslant 0$; for which $\overline{x}$ satisfies Definition 4.2 and the KKT conditions.

The next proposition tells us that when sparse-nondegeneracy is applied to a structurally diagonal problem (6), it becomes equivalent to LICQ for its NLP reformulation. Consequently, it is easy to build a diagonal counterexample for the converse of Proposition 4.2. For instance, take $m = 2$ and set $\overline{x} = 0$; then, define the constraint

$$G(x) \doteq \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix}, \tag{18}$$

and note that $v_{11}(\overline{x},\overline{E}) = v_{22}(\overline{x},\overline{E}) = 1$ for every matrix $\overline{E}$ that spans $\operatorname{Ker} G(\overline{x})$. Hence, sparse-nondegeneracy does not hold, although Robinson's CQ does.

Proposition 4.3. *If G is structurally diagonal as in* (6)*, then $\overline{x}$ satisfies sparse-nondegeneracy if, and only if, the set $\{\nabla g_i(\overline{x}) \colon g_i(\overline{x}) = 0\}$ is linearly independent.*

Proof. Let us assume that no diagonal entry of G is a structural zero; otherwise, neither sparse-nondegeneracy nor LICQ hold. The result is also trivial when $\operatorname{Ker} G(\overline{x}) = \{0\}$, so we assume otherwise. Without loss of generality, let us assume also that $g_{r+1}(\overline{x}) = \ldots = g_m(\overline{x}) = 0$, where $r = \operatorname{rank}(G(\overline{x}))$. Then, suppose that sparse-nondegeneracy holds at $\overline{x}$ and let $\overline{E}$ be the associated matrix whose columns form an orthonormal basis of $\operatorname{Ker} G(\overline{x})$. Note that

$$\nabla \widehat{G}^{\overline{E}}_{ij}(\overline{x}) = \sum_{\ell=r+1}^{m} \overline{e}_{\ell i}\overline{e}_{\ell j} \nabla G_{\ell\ell}(\overline{x}),$$

so $\operatorname{span}\left(\left\{\nabla \widehat{G}^{\overline{E}}_{ij}(\overline{x}) \colon (i,j) \in \mathcal{I}(\widehat{G}^{\overline{E}}\overline{x})\right\}\right) \subseteq \operatorname{span}\left(\{\nabla g_i(\overline{x}) \colon i \in \{r+1,\dots m\}\}\right)$, whence

$$\begin{aligned} m - r &\leqslant \operatorname{rank}\left(\left\{\nabla \widehat{G}^{\overline{E}}_{ij}(\overline{x}) \colon (i,j) \in \mathcal{I}(\widehat{G}^{\overline{E}}\overline{x})\right\}\right) \\ &\leqslant \operatorname{rank}(\{\nabla g_i(\overline{x}) \colon i \in \{r+1,\dots m\}\}), \end{aligned}$$

where the first inequality follows from item 2 of Definition 4.2. Hence,

$$\operatorname{rank}(\{\nabla g_i(\overline{x})\colon i \in \{r+1,\dots m\}\}) = m - r.$$

Conversely, take $\overline{E} = [0, \mathbb{I}_{m-r}]^\top$ and the result follows easily. ■

Remark 4.3. *Similarly to weak-nondegeneracy, the sparse-nondegeneracy condition is also invariant to block diagonal and multifold representations of a problem in the form* (Block-NSDP)*. That is, assuming the same notation as Remark 3.2, sparse nondegeneracy holds at a feasible point $\overline{x}$ of* (Block-NSDP) *if, and only if, for each $\ell \in \{1,\dots,q\}$ there is some matrix $\overline{E}_\ell$ that spans $\operatorname{Ker} G_\ell(\overline{x})$, such that:*

- *For all $i \in \{1,\dots,m_\ell - r_\ell\}$, we have $(i,i) \in \mathcal{I}(G_\ell^{\overline{E}_\ell}, \overline{x})$;*
- *The set*

$$\bigcup_{\ell=1}^{q} \left\{ v_{ij}^\ell(\overline{x}, \overline{E}_\ell)\colon (i,j) \in \mathcal{I}(G_\ell^{\overline{E}_\ell}, \overline{x}) \right\}$$

  is linearly independent, where $v_{ij}^\ell(\overline{x}, \overline{E}_\ell)$ is defined as in (14)*.*

Note that this is how sparse-nondegeneracy would be defined for a multifold equivalent representation of (Block-NSDP)*, with constraints $G_1(x) \succeq 0, \dots, G_q(x) \succeq 0$.*

Proposition 4.3 reveals a similarity among sparse-nondegeneracy, Forsgren's CQ, and weak-nondegeneracy, which is the fact they all reduce to LICQ when considering a diagonal matrix constraint. Moreover, it follows directly from Propositions 3.1 and 3.2 that nondegeneracy also strictly implies sparse-nondegeneracy. However, to make a rough comparison between Forsgren's CQ and sparse-nondegeneracy, note that both evaluate linear independence of the set $\{v_{ij}(\overline{x}, \overline{E})\colon 1 \leqslant i \leqslant j \leqslant m - r\}$, but while item 1 of Definition 4.2 takes coefficients structured as in $\mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})$, condition (F1) takes coefficients structured as in $\overline{E}^\top \mathcal{S}(\tilde{G}, \overline{x}) \overline{E}$. This suggests that they are different conditions. In fact, the next example shows that neither weak- nor sparse-nondegeneracy imply Forsgren's CQ.

Example 4.2. *Consider the constraint:*

$$G(x) \doteq \begin{bmatrix} x_1 & x_2 \\ x_2 & x_1 \end{bmatrix}$$

and the point $\overline{x} \doteq (0,0)$, and we have the following:

- **Weak-nondegeneracy holds at $\overline{x}$:** *take*

$$E^k = \overline{E} \doteq \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$$

  for all sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and note that $v_{11}(\overline{x}, \overline{E}) = [1,-1]$ and $v_{22}(\overline{x}, \overline{E}) = [1,1]$ are linearly independent;

- **Sparse-nondegeneracy holds at $\overline{x}$:** *take the same $\overline{E}$ as above and we have*

$$\widehat{G}^{\overline{E}}(x) \doteq \begin{bmatrix} x_1 - x_2 & 0 \\ 0 & x_1 + x_2 \end{bmatrix}$$

  and $\mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x}) = \{(1,1),(2,2)\}$;

- **Forsgren's CQ does not hold at $\overline{x}$:** *in this case Forsgren's CQ is equivalent to nondegeneracy, which does not hold because if $\overline{E} \doteq \mathbb{I}_m$, then $v_{11}(\overline{x}, \overline{E}) = v_{22}(\overline{x}, \overline{E}) = [1,0]$.*

Thus, neither weak- nor sparse-nondegeneracy imply Forsgren's CQ.

An elementary consequence of Lemma 4.2 is that sparse-nondegeneracy guarantees uniqueness of the Lagrange multiplier with respect to a fixed sparsity pattern, which is similar to a result proven for Forsgren's CQ [12].

Proposition 4.4. *Let $\overline{x}$ be a KKT point of* (NSDP) *that satisfies item 1 of Definition 4.2 and let $\overline{E}$ be the matrix that certifies it, which spans $\operatorname{Ker} G(\overline{x})$. Then, $\Lambda(\overline{x}) \bigcap \left(\overline{E}\mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})\overline{E}^\top\right)$ is a singleton.*

Proof. Let $Y_1, Y_2 \in \Lambda(\overline{x}) \bigcap \left(\overline{E}\mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})\overline{E}^\top\right)$ be Lagrange multipliers associated with $\overline{x}$, define $Y \doteq Y_1 - Y_2$, and by definition there exists some $Z \in \mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})$ such that $Y = \overline{E}Z\overline{E}^\top$ and $DG(\overline{x})^*[\overline{E}Z\overline{E}^\top] = 0$. By Lemma 4.2 we must have $Z = 0$ and, consequently, $Y_1 = Y_2$. ■

Another important property of sparse-nondegeneracy is that the number of structural zeros of $\widehat{G}^{\overline{E}}$, at points that satisfy it, remains the same regardless of $\overline{E}$.

Proposition 4.5. *Let $\overline{x} \in \mathcal{F}$ be such that $\operatorname{Ker} G(\overline{x}) \neq \{0\}$, and let $\overline{E}$ and $\overline{W}$ be matrices that span $\operatorname{Ker} G(\overline{x})$, such that item 1 of Definition 4.2 holds. Then, $\#\mathcal{I}(\widehat{G}^{\overline{E}}, \overline{x}) = \#\mathcal{I}(\widehat{G}^{\overline{W}}, \overline{x})$.*

Proof. Let $Z \doteq [z_{\ell s}]_{\ell,s\in\{1,\ldots,m-r\}}$ be an invertible matrix such that $\overline{E}Z = \overline{W}$ and note that $\widehat{G}^{\overline{W}}(\overline{x}) = Z^\top \widehat{G}^{\overline{E}}(\overline{x})Z$, so $\widehat{G}^{\overline{W}}_{ij}(\overline{x}) = \langle \widehat{G}^{\overline{E}}(\overline{x}), z_i z_j^\top\rangle = \sum_{\ell,s=1}^{r} z_{\ell i} z_{sj} \widehat{G}^{\overline{E}}_{\ell s}(\overline{x})$, where z_i denotes the i-th column of Z, and

$$\nabla \widehat{G}^{\overline{W}}_{ij}(\overline{x}) = \sum_{\ell,s=1}^{r} z_{\ell i} z_{sj} \nabla \widehat{G}^{\overline{E}}_{\ell s}(\overline{x}).$$

Rephrasing,

$$\nabla \widehat{G}^{\overline{W}}_{ij}(\overline{x}) = \underbrace{\begin{bmatrix} | & & | & | & & | \\ \nabla \widehat{G}^{\overline{E}}_{11}(\overline{x}) & \ldots & \nabla \widehat{G}^{\overline{E}}_{m-r,1}(\overline{x}) & \nabla \widehat{G}^{\overline{E}}_{12}(\overline{x}) & \ldots & \nabla \widehat{G}^{\overline{E}}_{m-r,m-r}(\overline{x}) \\ | & & | & | & & | \end{bmatrix}}_{\doteq\ \mathrm{unfold}(D\widehat{G}^{\overline{E}}(\overline{x}))\colon\ n\times(m-r)^2} \cdot \underbrace{\begin{bmatrix} z_{1i}z_{1j} \\ \vdots \\ z_{m-r,i}z_{1j} \\ z_{1i}z_{2j} \\ \vdots \\ z_{m-r,i}z_{m-r,j} \end{bmatrix}}_{\doteq\ \mathrm{vec}(z_i z_j^\top)\colon\ (m-r)^2\times 1},$$

where unfold: $\mathbb{R}^{m-r\times m-r\times n} \to \mathbb{R}^{n\times(m-r)^2}$ is an *unfolding* operator for the tensor $D\widehat{G}^{\overline{E}}(\overline{x})$ when it is seen as an $m-r \times m-r$ matrix with n-dimensional entries. Also, vec: $\mathbb{R}^{m-r\times m-r} \to \mathbb{R}^{(m-r)^2}$ is the usual *vectorization* operator, which transforms a matrix into a vector by stacking up its columns, from left to right. Consequently,

$$\underbrace{\mathrm{unfold}(D\widehat{G}^{\overline{W}}(\overline{x}))}_{n\times(m-r)^2} = \mathrm{unfold}(D\widehat{G}^{\overline{E}}(\overline{x}))\cdot \underbrace{\begin{bmatrix} | & & | & | & & | \\ \mathrm{vec}(z_1 z_1^\top) & \ldots & \mathrm{vec}(z_1 z_r^\top) & \mathrm{vec}(z_2 z_1^\top) & \ldots & \mathrm{vec}(z_r z_r^\top) \\ | & & | & | & & | \end{bmatrix}}_{(m-r)^2\times(m-r)^2},$$

which can be rephrased in terms of the *Kronecker product* as $\mathrm{unfold}(D\widehat{G}^{\overline{W}}(\overline{x})) = \mathrm{unfold}(D\widehat{G}^{\overline{E}}(\overline{x}))Z\otimes Z$. But since Z is invertible, $Z\otimes Z$ is also invertible, which means that

$$\mathrm{span}\left(\left\{\nabla \widehat{G}^{\overline{W}}_{ij}(\overline{x})\colon 1\leqslant i\leqslant j\leqslant m-r\right\}\right) = \mathrm{span}\left(\left\{\nabla \widehat{G}^{\overline{E}}_{ij}(\overline{x})\colon 1\leqslant i\leqslant j\leqslant m-r\right\}\right).$$

Then, since $\nabla \widehat{G}^{\overline{E}}_{ij}(\overline{x}) = 0$ for all $(i,j)\in\mathcal{I}(\widehat{G}^{\overline{E}},\overline{x})$ (and the same holds for $\overline{W}$), it follows that

$$\mathrm{span}\left(\left\{\nabla \widehat{G}^{\overline{W}}_{ij}(\overline{x})\colon (i,j)\in\mathcal{I}(\widehat{G}^{\overline{W}},\overline{x})\right\}\right) = \mathrm{span}\left(\left\{\nabla \widehat{G}^{\overline{E}}_{ij}(\overline{x})\colon (i,j)\in\mathcal{I}(\widehat{G}^{\overline{E}},\overline{x})\right\}\right),.$$

Finally, since item 1 of Definition 4.2 holds for both $\overline{E}$ and $\overline{W}$, we conclude that $\#\mathcal{I}(\widehat{G}^{\overline{E}},\overline{x}) = \#\mathcal{I}(\widehat{G}^{\overline{W}},\overline{x})$. ■

Proposition 4.5 tells us that the strength of sparse-nondegeneracy is invariant with respect to $\overline{E}$. That is, if there are multiple matrices $\overline{E}$ certifying sparse-nondegeneracy at a point $\overline{x}$, then they all induce similar conditions. In our opinion, this is an advantage with respect to Forsgreen's CQ. As for weak-nondegeneracy, we were not able to find any counterexample nor prove any relation between them. In fact, finding this relation seems a challenging task since there is no clear relation between the eigenvectors of $G(x)$ and its sparsity structure, in general.

5 A comment on Robinson's condition

At this point, one question that naturally arises is whether the conclusions from the previous sections could be extended to Robinson's CQ. We provide a partial answer to this question. But first, let us recall that in contrast with nondegeneracy, the analogy between Robinson's CQ and MFCQ does not depend on modelling. For instance, let us consider again an NLP problem with constraints $g_1(x)\geqslant 0,\ldots,g_m(x)\geqslant 0$ and its reformulation as an NSDP of the form (6). Also, for simplicity, let $\overline{x}$ be such that $G(\overline{x}) = 0$. Then, it is elementary to verify that Robinson's CQ holds at $\overline{x}$ if, and only if, the set $\{\nabla g_1(\overline{x}),\ldots,\nabla g_m(\overline{x})\}$ is *positive linearly independent*; that is, if there are no real numbers $\alpha_1\geqslant 0,\ldots,\alpha_m\geqslant 0$ such that $\sum_{i=1}^m \alpha_i\nabla g_i(\overline{x}) = 0$ and at least one α_i is nonzero, which is one possible characterization of MFCQ.

Inspired by Proposition 3.2 and MFCQ, we characterize Robinson's CQ in terms of the positive linear independence of the same set as Proposition 3.2, that also makes distinction between different choices of $\overline{E}$ but deals with only $m-r$ vectors per choice.

Proposition 5.1. *Let $\overline{x}\in\mathcal{F}$. Robinson's CQ holds at $\overline{x}$ if, and only if, either* $\mathrm{Ker}\, G(\overline{x}) = \{0\}$ *or the set* $\{v_{ii}(\overline{x},\overline{E})\colon i\in\{1,\ldots,m-r\}\}$ *is positive linearly independent for all matrices $\overline{E}$ that span* $\mathrm{Ker}\, G(\overline{x})$.

Proof. If $\operatorname{Ker} G(\overline{x}) = \{0\}$, there is nothing to prove, so let us assume otherwise. For any fixed $\overline{E}$, note that $\{v_{ii}(\overline{x}, \overline{E})\colon i \in \{1, \dots, m-r\}\}$ is positive linearly independent if, and only if, the following holds: if the scalars $\alpha_1 \geqslant 0, \dots, \alpha_{m-r} \geqslant 0$ satisfy

$$\sum_{i=1}^{m-r} \alpha_i DG(\overline{x})^* [\overline{e}_i \overline{e}_i^\top] = \sum_{i=1}^{m-r} \alpha_i \left[\overline{e}_i^\top D_{x_j} G(\overline{x}) \overline{e}_i\right]_{j \in \{1,\dots,n\}} = 0, \tag{19}$$

then one must have $\alpha_1 = \dots = \alpha_{m-r} = 0$. That is, the studied set is positive linearly independent if, and only if, for every matrix Y of the form

$$Y \doteq \sum_{i=1}^{m-r} \alpha_i \overline{e}_i \overline{e}_i^\top = \overline{E} \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_{m-r} \end{bmatrix} \overline{E}^\top \tag{20}$$

where $\alpha_1 \geqslant 0, \dots, \alpha_{m-r} \geqslant 0$, we have that

$$DG(\overline{x})^*[Y] = 0 \;\Rightarrow\; Y = 0. \tag{21}$$

With this in mind, let us assume that $\{v_{ii}(\overline{x}, \overline{E})\colon i \in \{1, \dots, m-r\}\}$ is positive linearly independent for all $\overline{E}$. Now let Y be such that $DG(\overline{x})^*[Y] = 0$, $\langle G(\overline{x}), Y\rangle = 0$ and $Y \succeq 0$, so there exists some matrix $\overline{E}$ that spans $\operatorname{Ker} G(\overline{x})$, such that Y has the form (20). It follows from our hypothesis that $Y = 0$ and since Y is arbitrary, Robinson's CQ holds. Conversely, assume that Robinson's CQ holds and let $\overline{E}$ and $\alpha_1 \geqslant 0, \dots, \alpha_{m-r} \geqslant 0$ be such that (19) holds. Then, define Y as in (20) and it follows from Robinson's CQ that $Y = 0$, which means $\alpha_1 = \dots = \alpha_{m-r} = 0$. Thus, $\{v_{ii}(\overline{x}, \overline{E})\colon i \in \{1, \dots, m-r\}\}$ is positive linearly independent, for every $\overline{E}$. ■

Similarly to Proposition 3.2, the main advantage of Proposition 5.1 is the suggestive way of potentially relaxing Robinson's CQ by considering only the limit points of eigenvector sequences associated with the sequences that converge to $\overline{x}$.

5.1 Weak-Robinson's CQ

Next, we present a definition with the same spirit as Definition 3.2.

Definition 5.1 (Weak-Robinson's CQ). *Let $\overline{x} \in \mathcal{F}$ and $r = \operatorname{rank}(\operatorname{Ker} G(\overline{x}))$. We say that* weak-Robinson's CQ *holds at $\overline{x}$ when either* $\operatorname{Ker} G(\overline{x}) = \{0\}$ *or: for every sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there exists some sequence of matrices $\{E^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^{m\times m-r}$ such that:*

1. *The columns of E^k are orthonormal eigenvectors associated with the $m-r$ smallest eigenvalues of $G(x^k)$, for each $k \in \mathbb{N}$;*
2. *There exists a limit point $\overline{E}$ of $\{E^k\}_{k\in\mathbb{N}}$ such that the set $\{v_{ii}(\overline{x}, \overline{E})\colon i \in \{1, \dots, m-r\}\}$ is positive linearly independent.*

It is clear from their definitions that weak-nondegeneracy implies weak-Robinson's CQ; and it is possible to show that the converse is not necessarily true. For instance, consider again problem (3.1), which we recall below:

$$G(x) \doteq \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix},$$

and note that all orthogonal matrices $\overline{E} \in \mathbb{R}^{2\times 2}$ have in their columns eigenvectors of $G(x)$, for every x. Since $v_{11}(\overline{x}, \overline{E}) = v_{22}(\overline{x}, \overline{E}) = 1$ for every $\overline{E}$, it follows that weak-nondegeneracy and weak-Robinson's CQ are equivalent to their strong counterparts in this case. Thus, from Propositions 3.2 and 5.1 we see that (weak-)nondegeneracy does not hold, while (weak-)Robinson's CQ does.

It is also clear from Proposition 5.1 that Robinson's CQ implies weak-Robinson's CQ; however, we were not capable of finding a counterexample for the converse. We conjecture that they are equivalent. Nevertheless, even if weak-Robinson's CQ happens to be strictly weaker than Robinson's CQ, it would still be a constraint qualification. This fact is proved below:

Theorem 5.1. *If $\overline{x}$ is a local minimizer of* (NSDP) *that satisfies weak-Robinson's CQ, then the KKT conditions hold at $\overline{x}$.*

Proof. This result can be proved by following the exact same steps as the proof of Theorem 3.2. To see this, observe that the scalar sequences $\{\alpha_i^k\}_{k\in\mathbb{N}}$ in the proof of Theorem 3.2 are nonnegative for every $k \in \mathbb{N}$ and every $i \in \{1, \dots, m-r\}$. ■

Furthermore, if sparse- and weak-nondegeneracy turn out to be independent, both of them imply weak-Robinson's CQ, which helps us in understanding where weak-nondegeneracy is positioned among the other constraint qualifications. This relationship is illustrated in the diagram in Figure 1.

It is noteworthy that it is also possible to define another variant of Robinson's CQ that enjoys sparsity, by replacing $\tilde{Y} \in \mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x})$ by $\tilde{Y} \in \mathcal{S}(\widehat{G}^{\overline{E}}, \overline{x}) \cap \mathbb{S}_+^{m-r}$ in Lemma 4.2. This definition is strictly implied by sparse-nondegeneracy (see the example given in (18)). Moreover, it is clear that this variant of Robinson's CQ is implied by Robinson's CQ, but the converse is also an open question. The proof that this is a CQ follows similarly to the proof of Theorem 3.2.

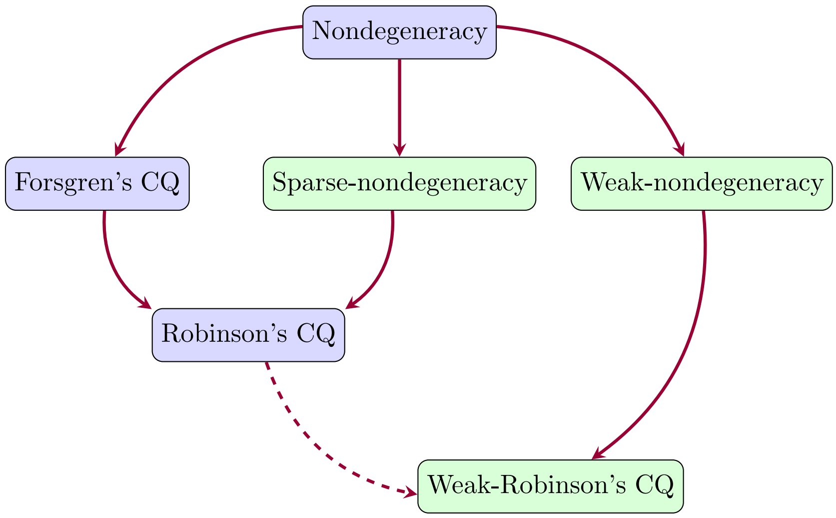


Figure 1: Relationship among some CQs for NSDP. Classical CQs are in blue boxes, while new CQs are in green boxes. Arrows indicate strict implications, except for the dashed arrow where the reverse implication is unknown.

6 Conclusions

In this paper, we studied the nondegeneracy condition of Shapiro and Fan [26] with the purpose of incorporating some matrix structure into it, such as spectral decompositions and structural sparsity. Our work was motivated by a well-known limitation of nondegeneracy, which is the fact it generally fails in the presence of structural sparsity in the constraint function. For example, we recall that a NSDP problem with multiple constraints may be equivalently reformulated as a single block diagonal constraint, but nondegeneracy is not expected to be preserved in the process. This limitation may have important consequences in practice, since many algorithms are theoretically supported by nondegeneracy and, on the other hand, structural sparsity is a very common trait of optimization models of real world problems.

To address this issue, we proposed two variants of nondegeneracy, here called weak-nondegeneracy and sparse-nondegeneracy. They were proven to be strictly weaker than the classical nondegeneracy. In particular, both new constraint qualifications only require the dimension constraint $n \geqslant m-r$, which is considerably less demanding than the constraint $n \geqslant (m-r)(m-r+1)/2$ imposed by nondegeneracy. Also, they are both invariant to multifold or block diagonal formulations of (NSDP) and, consequently, they recover the LICQ condition from NLP when the constraint function is structurally diagonal.

Both our conditions are inspired by *sequential optimality conditions* [3, 8] which provide simple proofs for the facts that weak-nondegeneracy and sparse-nondegeneracy are CQs (the proof for sparse-nondegeneracy was not presented but it is left for the reader). Besides the simplicity of the approach, the convergence of an external penalty method to KKT points under these CQs is obtained automatically (see the discussion after Theorem 3.2), which is a direct application of the new CQs. Also, several other CQs for NLP have been recently (re)invented with sequential optimality conditions in mind. In particular, the so-called *constant rank constraint qualification* (CRCQ) by Janin [15], and the *constant positive linear dependence* (CPLD) of Qi and Wei [22], together with their weaker counterparts [6, 7, 17]. Previous attempts have been made to extend these CQs to the conic context, but they have turned out to be flawed [2] or incomplete [4], since the results in [4] are only relevant for multifold conic problems where at least one block of constraints is such that the zero eigenvalue is simple. The approach we present in this paper gives the proper tools for providing the extension of all mentioned CQs to the context of general NSDPs and, more generally, to optimization over symmetric cones, also extending the global convergence results to more practical algorithms. For instance, in NLP, it is known that the convergence theory of a safeguarded augmented Lagrangian method can be built around CPLD [1], which will also be the case for its NSDP variant [8]. A continuation of this paper will appear shortly with these results.

With this in mind, we believe that the concepts introduced in this paper are interesting enough to shed a new light to the classical theme of constraint nondegeneracy for conic programming, showing, in particular, how to redefine it in such a way that linear independence can be replaced by weaker notions. In this process, new and interesting challenging open questions have appeared which we believed should be addressed. In particular, new studies should be conducted to clarify the relationship between weak-nondegeneracy and sparse-nondegeneracy, together with the relationship between weak-Robinson's CQ and Robinson's CQ (see Figure 1).

References

[1] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Augmented lagragrian methods under the constant positive linear dependence constraint qualification. *Mathematical Programming*, 111(1-2):5–32, 2008.

[2] R. Andreani, E. H. Fukuda, G. Haeser, H. Ramírez C., D. O. Santos, P. J. S. Silva, and T. P. Silveira. Erratum to: New Constraint Qualifications and Optimality Conditions for Second Order Cone Programs. *Optimization Online*, 2020.

[3] R. Andreani, G. Haeser, and J. M. Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60:627–641, 2011.

[4] R. Andreani, G. Haeser, L. M. Mito, H. Ramírez, D. O. Santos, and T. P. Silveira. Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming. *Optimization Online*, 2020.

[5] R. Andreani, G. Haeser, L. M. Mito, A. Ramos, and L. D. Secchin. On the best achievable quality of limit points of augmented Lagrangian schemes. *Optimization Online*, 2020.

[6] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. Two new weak constraint qualifications and applications. *SIAM Journal on Optimization*, 22:1109–1135, 2012.

[7] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programming*, 135:255–273, 2012.

[8] R. Andreani, G. Haeser, and D. S. Viana. Optimality conditions and global convergence for nonlinear semidefinite programming. *Mathematical Programming*, 180(1-2):203–235, 2020.

[9] M. F. Anjos and J. B. Lassere (eds.). *Handbook on Semidefinite, Conic and Polynomial Optimization.* International Series in Operations Research and Management Science. Springer US. New York, 2012.

[10] J. F. Bonnans and A. Shapiro. *Pertubation Analysis of Optimization Problems.* Springer-Verlag. Berlin, 2000.

[11] D. Dorsch, W. Gómez, and V. Shikhman. Sufficient optimality conditions hold for almost all nonlinear semidefinite programs. *Mathematical Programming*, 158(1-2):77–97, 2016.

[12] A. Forsgren. Optimality conditions for nonconvex semidefinite programming. *Mathematical Programming*, 88:105–128, 2000.

[13] P. Fusek. On metric regularity for weakly almost piecewise smooth functions and some applications in nonlinear semidefinite programming. *SIAM Journal on Optimization*, 23(2):1041–1061, 2013.

[14] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.

[15] R. Janin. Direction derivative of the marginal function in nonlinear programming. *Mathematical Programming Studies*, 21:127–138, 1984.

[16] D. Klatte and B. Kummer. Aubin property and uniqueness of solutions in cone constrained optimization. *Mathematical Methods of Operations Research*, 77(3):291–304, 2013.

[17] L. Minchenko and S. Stakhovski. On relaxed constant rank regularity condition in mathematical programming. *Optimization*, 60(4):429–440, 2011.

[18] B. S. Mordukhovich, T. T. A. Nghia, and R. T. Rockafellar. Full stability in finite-dimensional optimization. *Math. Oper. Res.*, 40(1):226–252, 2015.

[19] B. S. Mordukhovich, J. V. Outrata, and H. Ramírez. Graphical derivatives and stability analysis for parameterized equilibria with conic constraints. *Set-Valued Analysis and Applications*, pages 1–18, 2015.

[20] B. S. Mordukhovich, J. V. Outrata, and H. Ramírez. Second-order variational analysis in conic programming with applications to optimality and stability. *SIAM Journal on Optimization*, 25(1):76–101, 2015.

[21] G. Pataki. Strong duality in conic linear programming: facial reduction and extended duals. In *D. Bailey et al. (eds.) Computational and Analytical Mathematics*, volume 50, pages 613–634. Springer Proceedings in Mathematics & Statistics, New York, NY, 2013.

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

[23] S. M. Robinson. First-order conditions for general nonlinear optimization. *SIAM Journal on Applied Mathematics*, 30(4):597–610, 1976.

[24] A. Shapiro. First and second order analysis of nonlinear semidefinite programs. *Mathematical Programming, Series B*, 77(2):301–320, 1997.

[25] A. Shapiro. On uniqueness of Lagrange multipliers in optimization problems subject to cone constraints. *SIAM Journal on Optimization*, 7(2):508–518, 1997.

[26] A. Shapiro and M. K. H. Fan. On Eigenvalue optimization. *SIAM Journal on Optimization*, 5(3):552–569, 1995.

[27] D. Sun. The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Mathematics of Operations Research*, 31(4):761–776, 2006.

[28] G. Wachsmuth. On LICQ and the uniqueness of Lagrange multipliers. *Operations Research Letters*, 41(1):78–80, 2013.

[29] H. Wolkowicz, R. Saigal, and L. (eds.) Vandenberghe. *Handbook of Semidefinite Programming: Theory, Algorithms, an Applications*. International Series in Operations Research and Management Science. Springer US. New York, 2000.

[30] H. Yamashita and H. Yabe. A survey of numerical methods for nonlinear semidefinite programming. *Journal of the Operations Research Society of Japan*, 58(1):24–60, 2015.

Appendix G

External reference VII

Title: Sequential constant rank constraint qualifications for nonlinear semidefinite programming with applications.

Status: Under review.

What is attached: A preprint that has *not* been peer reviewed yet.

Sequential constant rank constraint qualifications for nonlinear semidefinite programming with applications

Roberto Andreani * Gabriel Haeser † Leonardo M. Mito † Héctor Ramírez C. ‡

February 3, 2022

Abstract

We present new constraint qualification conditions for nonlinear semidefinite programming that extend some of the constant rank-type conditions from nonlinear programming. As an application of these conditions, we provide a unified global convergence proof of a class of algorithms to stationary points without assuming neither uniqueness of the Lagrange multiplier nor boundedness of the Lagrange multipliers set. This class of algorithm includes, for instance, general forms of augmented Lagrangian, sequential quadratic programming, and interior point methods. We also compare these new conditions with some of the existing ones, including the nondegeneracy condition, Robinson's constraint qualification, and the metric subregularity constraint qualification.

Keywords: Constant rank, Constraint qualifications, Semidefinite programming, Algorithms, Global convergence.

1 Introduction

Constraint qualification (CQ) conditions play a crucial role in optimization. They permit to establish first- and second-order necessary optimality conditions for local minima and support the convergence theory of many practical algorithms (see, for instance, a unified convergence analysis for a whole class of algorithms by Andreani et al. [8, Thm. 6]). Some of the well-known CQs in nonlinear programming (NLP) are the *constant-rank constraint qualification* (CRCQ), introduced by Janin [22], and the *constant positive linear dependence* (CPLD) condition. The latter was first conceptualized by Qi and Wei [27], and then proved to be a constraint qualification by Andreani et al. [12]. Moreover, it has been a source of inspiration for other authors to define even weaker constraint qualifications for NLP, such as the *constant rank of the subspace component* (CRSC) [9], and the relaxed versions of CRCQ [24] and CPLD [8]. Our interest in constant rank-type conditions is motivated, mainly, by their applications towards obtaining global convergence results of iterative algorithms to stationary points without relying on boundedness or uniqueness of Lagrange multipliers. However, several other applications that we do not pursue in this paper may be expected to be extended to the conic context, such as the computation of the derivative of the value function [22, 25] and the validity of strong second-order necessary optimality conditions that do not rely on the whole set of Lagrange multipliers [1]. Besides, their ability of dealing with redundant constraints, up to some extent, gives modellers some degree of freedom without losing regularity or convergence guarantees on algorithms. For instance, the standard NLP trick of replacing one nondegenerate equality constraint by two inequalities of opposite sign does not violate CRCQ, while violating the standard *Mangasarian-Fromovitz CQ* (MFCQ).

Constant-rank type CQs have been proposed in conic programming only very recently. The first extension of CRCQ to *nonlinear second-order cone programming* (NSOCP) appeared in [33], but it was shown to be incorrect in [2]. A second proposal [7], which encompasses also *nonlinear semidefinite programming* (NSDP) problems, consists of transforming some of the conic constraints into NLP constraints via a reduction function, whenever it was possible, and then demanding constant linear dependence of the reduced constraints, locally. This was considered by the authors a *naive* extension, since it basically avoids the main difficulties that are expected from a conic framework. What both these works have in common is that they somehow neglected the conic structure of the problem.

The authors received financial support from FAPESP (grants 2017/18308-2, 2017/17840-2, and 2018/24293-0), CNPq (grants 301888/2017-5, 303427/2018-3 and 404656/2018-8), and ANID (FONDECYT grant 1201982 and Basal Program CMM ANID PIA AFB170001).

*Department of Applied Mathematics, University of Campinas, Campinas, SP, Brazil. Email: andreani@unicamp.br

†Department of Applied Mathematics, University of São Paulo, São Paulo, SP, Brazil. Emails: ghaeser@ime.usp.br, leokoto@ime.usp.br

‡Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático (AFB170001 - CNRS IRL2807), Universidad de Chile, Santiago, Chile. Email: hramirez@dim.uchile.cl

In a recent article [6], we introduced weak notions of regularity for *nonlinear semidefinite programming* (NSDP) that were defined in terms of the eigenvectors of the constraints – therein called *weak-nondegeneracy* and *weak-Robinson's CQ*. These conditions take into consideration only the diagonal entries of some particular transformation of the matrix constraint. Noteworthy, weak-nondegeneracy happens to be equivalent to the *linear independence CQ* (LICQ) when an NLP constraint is modeled as a structurally diagonal matrix constraint, unlike the standard *nondegeneracy* condition [31], which in turn is considered the usual extension of LICQ to NSDP. Moreover, the proof technique we employed in [6] induces a direct application in the convergence theory of an external penalty method. In this paper, we use these conditions to derive our extension proposals for CRCQ and CPLD to NSDP, which also recover their counterparts in NLP when it is modelled as a structurally diagonal matrix constraint. These CQs are called, in this paper, as weak-CRCQ and weak-CPLD, respectively.

However, to provide support for algorithms other than the external penalty method, we present stronger variants of these conditions, called sequential-CRCQ and sequential-CPLD (abbreviated seq-CRCQ and seq-CPLD, respectively), by incorporating perturbations in their definitions. This makes them robust and easily connectible with algorithms that keep track of approximate Lagrange multipliers, but also more exigent. Nevertheless, seq-CRCQ is still strictly weaker than nondegeneracy, and independent of Robinson's CQ, while seq-CPLD is strictly weaker than Robinson's CQ. On the other hand, weak-CRCQ is strictly weaker than seq-CRCQ, while weak-CPLD is strictly weaker than weak-CRCQ and seq-CPLD. Moreover, we show that seq-CPLD implies the *metric subregularity CQ*.

The content of this paper is organized as follows: Section 2 introduces notation and some well-known theorems and definitions that will be useful in the sequel. Our main results for NSDP are presented in Sections 3 and 4. Indeed, Section 3 is devoted to the study of weak-CRCQ and weak-CPLD and their properties, which in turn need to invoke weak-nondegeneracy and weak-Robinson's CQ as a motivation. Section 4 studies seq-CRCQ and seq-CPLD – the main CQs of this paper – and some algorithms supported by them. In Section 5, we discuss the relationship between seq-CPLD and the metric subregularity CQ. Lastly, some final remarks are given in Section 6.

2 A nonlinear semidefinite programming review

In this section, $\mathbb{S}^m$ denotes the linear space of all $m \times m$ real symmetric matrices equipped with the inner product defined as $\langle M, N\rangle \doteq \operatorname{trace}(MN) = \sum_{i,j=1}^m M_{ij}N_{ij}$ for all $M, N \in \mathbb{S}^m$, and $\mathbb{S}^m_+$ is the cone of all positive semidefinite matrices in $\mathbb{S}^m$. Additionally, for every $M \in \mathbb{S}^m$ and every $\tau > 0$, we denote by $B(M, \tau) \doteq \{Z \in \mathbb{S}^m : \|M - Z\| < \tau\}$ the open ball centered at M with radius τ with respect to the Frobenius norm $\|M\| \doteq \sqrt{\langle M, M\rangle}$, and its closure will be denoted by $\overline{B}(M, \tau)$.

We consider the NSDP problem in standard (dual) form:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & G(x) \succeq 0, \end{aligned} \tag{NSDP}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{S}^m$ are continuously differentiable functions, and $\succeq$ is the partial order induced by $\mathbb{S}^m_+$; that is, $M \succeq N$ if, and only if, $M - N \in \mathbb{S}^m_+$.

Equality constraints are omitted in (NSDP) for simplicity of notation, but our definitions and results are flexible regarding inclusion of such constraints, which should be done in the same way as in [7]. Moreover, throughout the whole paper, we will denote the feasible set of (NSDP) by $\mathcal{F}$.

Let us recall that the orthogonal projection of an element $M \in \mathbb{S}^m$ onto $\mathbb{S}^m_+$, which is defined as

$$\Pi_{\mathbb{S}^m_+}(M) \doteq \underset{N\in\mathbb{S}^m_+}{\operatorname{argmin}} \|M - N\|,$$

is a convex continuous function of M since $\mathbb{S}^m_+$ is nonempty, closed, and convex. Furthermore, since $\mathbb{S}^m_+$ is self-dual, every $M \in \mathbb{S}^m$ has a *Moreau decomposition* [26, Prop. 1] in the form

$$M = \Pi_{\mathbb{S}^m_+}(M) - \Pi_{\mathbb{S}^m_+}(-M)$$

with $\langle \Pi_{\mathbb{S}^m_+}(M), \Pi_{\mathbb{S}^m_+}(-M)\rangle = 0$, and a *spectral decomposition* in the form

$$M = \lambda_1(M)u_1(M)u_1(M)^\top + \ldots + \lambda_m(M)u_m(M)u_m(M)^\top, \tag{1}$$

where $u_1(M), \ldots, u_m(M) \in \mathbb{R}^m$ are arbitrarily chosen orthonormal eigenvectors associated with the eigenvalues $\lambda_1(M), \ldots, \lambda_m(M)$, respectively. In turn, these eigenvalues are assumed to be arranged in non-increasing order. Equivalently, we can write (1) as $M = U\mathcal{D}U^\top$, where U is an orthogonal matrix whose i-th column is $u_i(M)$, and $\mathcal{D} \doteq \operatorname{Diag}(\lambda_1(M), \ldots, \lambda_m(M))$ is a matrix whose diagonal entries are $\lambda_1(M), \ldots, \lambda_m(M)$ and the remaining entries are zero.

A convenient property of the orthogonal projection onto $\mathbb{S}^m_+$ is that, for every $M \in \mathbb{S}^m$, we have

$$\Pi_{\mathbb{S}^m_+}(M) = [\lambda_1(M)]_+ u_1(M) u_1(M)^\top + \ldots + [\lambda_m(M)]_+ u_m(M) u_m(M)^\top,$$

where $[\,\cdot\,]_+ \doteq \max\{\,\cdot\,, 0\}$.

Given a sequence of sets $\{S^k\}_{k\in\mathbb{N}}$, recall its *outer limit* (or *upper limit*) in the sense of Painlevé-Kuratowski (cf. [29, Def. 4.1] or [15, Def. 2.52]), defined as

$$\operatorname{Lim\,sup}_{k\in\mathbb{N}} S^k \doteq \left\{ y \colon \exists I \subseteq_\infty \mathbb{N},\ \exists \{y^k\}_{k\in I} \to y,\ \forall k \in I,\ y^k \in S^k \right\},$$

which is the collection of all cluster points of sequences $\{y^k\}_{k\in\mathbb{N}}$ such that $y^k \in S^k$ for every $k \in \mathbb{N}$. The notation $I \subseteq_\infty \mathbb{N}$ means that I is an infinite subset of the set of natural numbers $\mathbb{N}$.

We denote the *Jacobian* of G at a given point $x \in \mathbb{R}^n$ by $DG(x)$, and the *adjoint* operator of $DG(x)$ will be denoted by $DG(x)^*$. Moreover, the i-th partial derivative of G at x will be denoted by $D_{x_i}G(x)$, and the *gradient* of f at x will be written as $\nabla f(x)$, for every $x \in \mathbb{R}^n$.

2.1 Classical optimality conditions and constraint qualifications

As usual in continuous optimization, we drive our attention towards local solutions of (NSDP) that satisfy the so-called *Karush-Kuhn-Tucker* (KKT) conditions, defined as follows:

Definition 2.1. *We say that the* Karush-Kuhn-Tucker *conditions hold at* $\overline{x} \in \mathcal{F}$ *when there exists some* $\overline{Y} \succeq 0$ *such that*

$$\nabla_x L(\overline{x}, \overline{Y}) = 0 \quad \text{and} \quad \langle G(\overline{x}), \overline{Y} \rangle = 0,$$

where $L(x, Y) \doteq f(x) - \langle G(x), Y \rangle$ *is the Lagrangian function of* (NSDP). *The matrix* $\overline{Y}$ *is called a* Lagrange multiplier *associated with* $\overline{x}$*, and the set of all Lagrange multipliers associated with* $\overline{x}$ *will be denoted by* $\Lambda(\overline{x})$*.*

Of course, not every local minimizer satisfies KKT in the absence of a CQ. In order to recall some classical CQs, it is necessary to use the *(Bouligand) tangent cone* to $\mathbb{S}^m_+$ at a point $M \succeq 0$. This object can be characterized in terms of any matrix $E \in \mathbb{R}^{m\times(m-r)}$, whose columns form an orthonormal basis of $\operatorname{Ker} M$, as follows (e.g., [15, Ex. 2.65]):

$$T_{\mathbb{S}^m_+}(M) = \{N \in \mathbb{S}^m \colon E^\top N E \succeq 0\}, \tag{2}$$

where r denotes the rank of M. So, its *lineality space*, defined as the largest linear space contained in $T_{\mathbb{S}^m_+}(M)$, is computed as follows:

$$\operatorname{lin}(T_{\mathbb{S}^m_+}(M)) = \{N \in \mathbb{S}^m \colon E^\top N E = 0\}. \tag{3}$$

The latter is a direct consequence of the identity $\operatorname{lin}(C) = C \cap (-C)$, satisfied for any closed convex cone C.

One of the most recognized constraint qualifications in NSDP is the *nondegeneracy* (or *transversality*) condition introduced by Shapiro and Fan [31], which can be characterized [15, Eq. 4.172] at a point $\overline{x} \in \mathcal{F}$ when the following relation is satisfied:

$$\operatorname{Im} DG(\overline{x}) + \operatorname{lin}(T_{\mathbb{S}^m_+}(G(\overline{x}))) = \mathbb{S}^m.$$

If $\overline{x}$ is a local solution of (NSDP) that satisfies nondegeneracy, then $\Lambda(\overline{x})$ is a singleton, but the converse is not necessarily true unless $\operatorname{rank}(G(\overline{x})) + \operatorname{rank}(\overline{Y}) = m$ holds for some $\overline{Y} \in \Lambda(\overline{x})$ [15, Prop. 4.75]. This last condition is known as *strict complementarity* in this NSDP framework. By (2) it is possible to characterize nondegeneracy at $\overline{x}$ by means of any given matrix $\overline{E}$ with orthonormal columns that span $\operatorname{Ker} G(\overline{x})$. Indeed, following [15, Sec. 4.6.1], nondegeneracy holds at $\overline{x}$ if, and only if, either $\operatorname{Ker} G(\overline{x}) = \{0\}$ or the linear mapping $\psi_{\overline{x}} \colon \mathbb{R}^n \to \mathbb{S}^{m-r}$ given by

$$\psi_{\overline{x}}(\,\cdot\,) \doteq \overline{E}^\top DG(\overline{x})[\,\cdot\,]\overline{E} \tag{4}$$

is surjective, which is in turn equivalent to saying that the vectors

$$v_{ij}(\overline{x}, \overline{E}) \doteq \left[\overline{e}_i^\top D_{x_1} G(\overline{x}) \overline{e}_j, \ldots, \overline{e}_i^\top D_{x_n} G(\overline{x}) \overline{e}_j \right]^\top, \quad 1 \le i \le j \le m - r, \tag{5}$$

are linearly independent [30, Prop. 6], where $\overline{e}_i$ denotes the i-th column of $\overline{E}$ and r is the rank of $G(\overline{x})$.

Another widespread constraint qualification is Robinson's CQ [28], which can be characterized at $\overline{x} \in \mathcal{F}$ by the existence of some $d \in \mathbb{R}^n$ such that

$$G(\overline{x}) + DG(\overline{x})[d] \in \operatorname{int} \mathbb{S}^m_+ \tag{6}$$

where $\operatorname{int} \mathbb{S}^m_+$ stands for the topological interior of $\mathbb{S}^m_+$. It is known (e.g., [15, Props. 3.9 and 3.17]) that when $\overline{x}$ is a local solution of (NSDP), then $\Lambda(\overline{x})$ is nonempty and compact if, and only if, Robinson's CQ holds at $\overline{x}$.

Given the properties and characterizations recalled above, the nondegeneracy condition is typically considered the natural extension of LICQ from NLP to NSDP, while Robinson's CQ is considered the extension of MFCQ.

2.2 A sequential optimality condition connected to the external penalty method

If we do not assume any CQ, every local minimizer of (NSDP) can still be proved to satisfy at least a *sequential* type of optimality condition that is deeply connected to the classical external penalty method. Namely:

Theorem 2.1. *Let $\overline{x}$ be a local minimizer of* (NSDP)*, and let $\{\rho_k\}_{k\in\mathbb{N}} \to +\infty$. Then, there exists some $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, such that for each $k \in \mathbb{N}$, x^k is a local minimizer of the regularized penalized function*

$$F(x) \doteq f(x) + \frac{1}{2}\|x - \overline{x}\|_2^2 + \frac{\rho_k}{2}\|\Pi_{\mathbb{S}^m_+}(-G(x))\|^2.$$

Proof. See [10, Thm. 2]. For a more general proof, see the first part of the proof of [4, Thm. 2]. ■

Note that Theorem 2.1 provides a sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ such that each x^k satisfies, with an error $\varepsilon^k \to 0^+$, the first-order optimality condition of the unconstrained minimization problem

$$\underset{x\in\mathbb{R}^n}{\text{Minimize}}\ f(x) + \frac{\rho_k}{2}\|\Pi_{\mathbb{S}^m_+}(-G(x))\|^2,$$

so $\{x^k\}_{k\in\mathbb{N}}$ characterizes an output sequence of an *external penalty method*. Moreover, the sequence $\{Y^k\}_{k\in\mathbb{N}} \subseteq \mathbb{S}^m_+$, where

$$Y^k \doteq \rho_k \Pi_{\mathbb{S}^m_+}(-G(x^k))$$

for every $k \in \mathbb{N}$, consists of *approximate Lagrange multipliers* for $\overline{x}$, in the sense that $\nabla_x L(x^k, Y^k) \to 0$ and complementarity and feasibility are approximately fulfilled, in view of Moreau's decomposition – indeed, note that $\langle G(x^k) + \Delta^k, Y^k\rangle = 0$ and $G(x^k) + \Delta^k \succeq 0$, with $\Delta^k = \Pi_{\mathbb{S}^m_+}(-G(x^k)) \to 0$, for every $k \in \mathbb{N}$.

These sequences will suffice to obtain the results of the first part of this paper (Section 3), but in order to extend their scope to a larger class of iterative algorithms, in Section 4, we will need a more general sequential optimality condition, which will be presented later on.

2.3 Reviewing constant rank-type constraint qualifications for NLP

This section is meant to be a brief review of the main results regarding the classical nonlinear programming problem:

$$\begin{aligned} &\underset{x\in\mathbb{R}^n}{\text{Minimize}} && f(x), \\ &\text{subject to} && g_1(x) \geq 0, \ldots, g_m(x) \geq 0, \end{aligned} \tag{NLP}$$

where $f, g_1, \ldots, g_m \colon \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable functions.

As far as we know, the first constant rank-type constraint qualification was introduced by Janin [22], to obtain directional derivatives for the optimal value function of a perturbed NLP problem. Janin's condition is defined as follows:

Definition 2.2. *Let $\overline{x} \in \mathcal{F}$. The* constant rank constraint qualification *for* (NLP) *(CRCQ) holds at $\overline{x}$ if there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that, for every subset $J \subseteq \{i \in \{1, \ldots, m\} \colon g_i(\overline{x}) = 0\}$, the rank of the family $\{\nabla g_i(x)\}_{i\in J}$ remains constant for all $x \in \mathcal{V}$.*

As noticed by Qi and Wei [27] it is possible to rephrase Definition 2.2 in terms of the "constant linear dependence" of $\{\nabla g_i(x)\}_{i\in J}$ for every J. That is, CRCQ holds at $\overline{x}$ if, and only if, there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that, for every $J \subseteq \{i \in \{1, \ldots, m\} \colon g_i(\overline{x}) = 0\}$, if $\{\nabla g_i(\overline{x})\}_{i\in J}$ is linearly dependent, then $\{\nabla g_i(x)\}_{i\in J}$ remains linearly dependent for every $x \in \mathcal{V}$. Based on this characterization, Qi and Wei proposed a relaxation of CRCQ, which they called *constant positive linear dependence* (CPLD) condition, but this was only proven to be a constraint qualification a few years later, in [12]. To properly define CPLD, recall that a family of vectors $\{z_i\}_{i\in J}$ of $\mathbb{R}^n$ is said to be *positively linearly independent* when

$$\sum_{i\in J} z_i\alpha_i = 0, \quad \alpha_i \geq 0, \ \forall i \in J \quad \Rightarrow \quad \alpha_i = 0, \ \forall i \in J.$$

Next, we recall the CPLD constraint qualification:

Definition 2.3. *Let $\overline{x} \in \mathcal{F}$. The* constant positive linear dependence *condition for* (NLP) *(CPLD) holds at $\overline{x}$ if there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that, for every $J \subseteq \{i \in \{1, \ldots, m\} \colon g_i(\overline{x}) = 0\}$, if the family $\{\nabla g_i(\overline{x})\}_{i\in J}$ is positively linearly dependent, then $\{\nabla g_i(x)\}_{i\in J}$ remains linearly dependent for all $x \in \mathcal{V}$.*

Clearly, CPLD is implied by CRCQ, which is in turn implied by LICQ and is independent of MFCQ. Moreover, CPLD is implied by MFCQ, and all those implications are strict [12, 22]. To show that our extensions of CRCQ and CPLD to NSDP are indeed constraint qualifications (Theorem 3.1), we shall take inspiration in [8], where the authors employ Theorem 2.1 together with the well-known *Carathéodory's Lemma*:

Lemma 2.1 (Exercise B.1.7 of [13])**.** *Let* $z_1, \ldots, z_p \in \mathbb{R}^n$*, and let* $\alpha_1, \ldots, \alpha_p \in \mathbb{R}$ *be arbitrary. Then, there exist some* $J \subseteq \{1, \ldots, p\}$ *and some scalars* $\tilde{\alpha}_i$ *with* $i \in J$*, such that* $\{z_i\}_{i \in J}$ *is linearly independent,*

$$\sum_{i=1}^{p} \alpha_i z_i = \sum_{i \in J} \tilde{\alpha}_i z_i,$$

and $\alpha_i \tilde{\alpha}_i > 0$*, for all* $i \in J$*.*

If one considers equality constraints in (NSDP) separately, one should employ an adapted version of Carathéodory's Lemma that fixes a particular subset of vectors, which can be found in [8, Lem. 2]. In our current setting, Lemma 2.1 will suffice as is.

3 Constant rank constraint qualifications for NSDP

Based on the relationship between LICQ and CRCQ, the most natural candidate for an extension of CRCQ to NSDP is to demand every subset of

$$\{v_{ij}(x, \overline{E}) \colon 1 \leq i \leq j \leq m - r\}$$

to remain with constant rank (or constant linear dependence) in a neighborhood of $\overline{x}$. However, this candidate cannot be a CQ, as shown in the following counterexample, adapted from [2, Eq. 2]:

Example 3.1. *Consider the problem to minimize* $f(x) \doteq -x$ *subject to*

$$G(x) \doteq \begin{bmatrix} x & x + x^2 \\ x + x^2 & x \end{bmatrix} \succeq 0.$$

For this problem, $\overline{x} \doteq 0$ *is the only feasible point and, therefore, the unique global minimizer of the problem. Since* $G(\overline{x}) = 0$*, the columns of the matrix* $\overline{E} \doteq \mathbb{I}_2$ *form an orthonormal basis of* $\mathrm{Ker}\, G(\overline{x})$ *(the whole space* $\mathbb{R}^2$*). For this choice of* $\overline{E}$*, we have*

$$v_{11}(x, \overline{E}) = v_{22}(x, \overline{E}) = 1 \quad \text{and} \quad v_{12}(x, \overline{E}) = 1 + 2x.$$

Since they are all bounded away from zero, the rank of every subset of $\{v_{ij}(x, \overline{E}) \colon 1 \leq i \leq j \leq 2\}$ *remains constant for every* x *around* $\overline{x}$*. However, Note that* $\overline{x}$ *does not satisfy the KKT conditions because any* $\overline{Y} \doteq \begin{bmatrix} \overline{Y}_{11} & \overline{Y}_{12} \\ \overline{Y}_{12} & \overline{Y}_{22} \end{bmatrix} \in \Lambda(\overline{x})$ *would necessarily be a solution of the system*

$$\begin{aligned} &\overline{Y}_{11} \geq 0, \\ &\overline{Y}_{22} \geq 0, \\ &\overline{Y}_{11}\overline{Y}_{22} - \overline{Y}_{12}^2 \geq 0, \\ &\overline{Y}_{11} + 2\overline{Y}_{12} + \overline{Y}_{22} = -1, \end{aligned}$$

which has no solution.

Besides, it is well known that even if G is affine, not all local minimizers of (NSDP) satisfy KKT, but in this case every subfamily of $\{v_{ij}(x, \overline{E}) \colon 1 \leq i \leq j \leq m - r\}$ remains with constant rank for every $x \in \mathbb{R}^n$.

What Example 3.1 tells us is that $\overline{E} = \mathbb{I}_2$ may be a bad choice of $\overline{E}$. In fact, let us choose a different $\overline{E}$, namely, denote the columns of $\overline{E}$ by $\overline{e}_1 \doteq [a, b]^\top$ and $\overline{e}_2 \doteq [c, d]^\top$, and take $a = -1/\sqrt{2}$ and $b = c = d = 1/\sqrt{2}$. This election of $\overline{E}$ happens to diagonalize $G(x)$ for every x, but it follows that

$$\begin{aligned} v_{11}(x, \overline{E}) &= 1 + 2ab(1 + 2x) = -2x; \\ v_{22}(x, \overline{E}) &= 1 + 2cd(1 + 2x) = 2(1 + x); \\ v_{12}(x, \overline{E}) &= (ad + bc)(1 + 2x) = 0, \end{aligned}$$

and the rank of $\{v_{11}(x, \overline{E})\}$ does not remain constant in a neighborhood of $\overline{x} = 0$.

In light of our previous work [6], the situation presented above is not surprising. Therein, we already noted that identifying the "good" matrices $\overline{E}$ allows us to obtain relaxed versions of nondegeneracy and Robinson's CQ for NSDP. This identification can also be used to extend constant-rank type conditions to NSDP and is the starting point for the results we will present in the current manuscript.

For the sake of completeness, let us quickly summarize a discussion raised in [6] before presenting the results of this paper. Consider a feasible point $\overline{x} \in \mathcal{F}$ and denote by r the rank of $G(\overline{x})$. Observe that $\lambda_r(M) > \lambda_{r+1}(M)$ for every $M \in \mathbb{S}^m$ close enough to $G(\overline{x})$. Thus, when $r < m$, define the set

$$\mathcal{E}_r(M) \doteq \left\{ E \in \mathbb{R}^{m \times (m-r)} \colon \begin{array}{l} ME = E\mathrm{Diag}(\lambda_{r+1}(M), \ldots, \lambda_m(M)) \\ E^\top E = \mathbb{I}_{m-r} \end{array} \right\}, \tag{7}$$

which consists of all matrices whose columns are orthonormal eigenvectors associated with the $m-r$ smallest eigenvalues of M, which is well defined whenever $\lambda_r(M) > \lambda_{r+1}(M)$. In (7), $\mathrm{Diag}(\lambda_{r+1}(M), \ldots, \lambda_m(M))$ denotes the diagonal matrix whose diagonal entries are $\lambda_{r+1}(M), \ldots, \lambda_m(M)$. By convention, $\mathcal{E}_r(M) \doteq \emptyset$ when $r = m$. By construction, $\mathcal{E}_r(M)$ is nonempty provided $r < m$ and M is close enough to $G(\overline{x})$. In particular, in this situation, $\mathcal{E}_r(G(\overline{x}))$ is the set of all matrices with orthonormal columns that span $\mathrm{Ker}\, G(\overline{x})$.

We showed, in [6, Prop. 3.2], that nondegeneracy can be equivalently stated as the linear independence of the smaller family, $\{v_{ii}(\overline{x}, \overline{E})\}_{i\in\{1,\ldots,m-r\}}$, as long as this holds for all $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ instead of a fixed one. Similarly, Robinson's CQ can be translated as the positive linear independence of the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i\in\{1,\ldots,m-r\}}$ for every $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ [6, Prop. 5.1]. This characterization suggested a weak form of nondegeneracy (and Robinson's CQ) that takes into account only a particular subset of $\mathcal{E}_r(G(\overline{x}))$ instead of the whole set, which reads as follows:

Definition 3.1 (Def. 3.2 and Def. 5.1 of [6])**.** *Let $\overline{x} \in \mathcal{F}$ and let r be the rank of $G(\overline{x})$. We say that $\overline{x}$ satisfies:*

- Weak-nondegeneracy condition *for NSDP if either $r = m$ or, for each sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there exists some $\overline{E} \in \mathrm{Lim\,sup}_{k\in\mathbb{N}}\mathcal{E}_r(G(x^k))$ such that the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i\in\{1,\ldots,m-r\}}$ is linearly independent;*
- Weak-Robinson's CQ condition *for NSDP if either $r = m$ or, for each sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there exists some $\overline{E} \in \mathrm{Lim\,sup}_{k\in\mathbb{N}}\mathcal{E}_r(G(x^k))$ such that the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i\in\{1,\ldots,m-r\}}$ is positively linearly independent.*

Note that, in general, $\mathrm{Lim\,sup}_{k\in\mathbb{N}}\mathcal{E}_r(G(x^k)) \subseteq \mathcal{E}_r(G(\overline{x}))$, but the reverse inclusion is not always true, meaning $\mathcal{E}_r(G(x))$ is not necessarily continuous at $\overline{x}$ as a set-valued mapping. It then follows that weak-nondegeneracy is indeed a strictly weaker CQ than nondegeneracy [6, Ex. 3.1]. Moreover, in contrast with nondegeneracy, weak-nondegeneracy happens to fully recover LICQ when $G(x)$ is a structurally diagonal matrix constraint in the form $G(x) \doteq \mathrm{Diag}(g_1(x), \ldots, g_m(x))$ [6, Prop. 3.3]. Similarly, weak-Robinson's CQ is implied by Robinson's CQ and coincides with MFCQ when $G(x)$ is diagonal.

3.1 Weak constant rank CQs for NSDP

A straightforward relaxation of weak-nondegeneracy and weak-Robinson's CQ, likewise NLP, leads to our first extension proposal of CRCQ and CPLD to NSDP:

Definition 3.2 (weak-CRCQ and weak-CPLD)**.** *Let $\overline{x} \in \mathcal{F}$ and let r be the rank of $G(\overline{x})$. We say that $\overline{x}$ satisfies the:*

- Weak constant rank constraint qualification *for NSDP (weak-CRCQ) if either $r = m$ or, for each sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there exists some $\overline{E} \in \mathrm{Lim\,sup}_{k\in\mathbb{N}}\mathcal{E}_r(G(x^k))$ such that, for every subset $J \subseteq \{1, \ldots, m-r\}$: if the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i\in J}$ is linearly dependent, then $\{v_{ii}(x^k, E^k)\}_{i\in J}$ remains linearly dependent, for all $k \in I$ large enough.*
- Weak constant positive linear dependence constraint qualification *for NSDP (weak-CPLD) if either $r = m$ or, for each sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, there exists some $\overline{E} \in \mathrm{Lim\,sup}_{k\in\mathbb{N}}\mathcal{E}_r(G(x^k))$ such that, for every subset $J \subseteq \{1, \ldots, m-r\}$: if the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i\in J}$ is positively linearly dependent, then $\{v_{ii}(x^k, E^k)\}_{i\in J}$ remains linearly dependent, for all $k \in I$ large enough.*

For both definitions, $I \subseteq_\infty \mathbb{N}$, and $\{E^k\}_{k\in I}$ is a sequence converging to $\overline{E}$ and such that $E^k \in \mathcal{E}_r(G(x^k))$ for every $k \in I$, as required by the Painlevé-Kuratowski outer limit.

Observe that if weak-nondegeneracy holds at $\overline{x}$, then weak-CRCQ is vacuously true, and weak-CRCQ in turn implies weak-CPLD. Similarly, the condition weak-Robinson's CQ implies weak-CPLD as well. However, Robinson's CQ and its weak variant are both independent of weak-CRCQ. In fact, the next example shows that weak-CRCQ is not implied by either (weak-)Robinson's CQ or weak-CPLD.

Example 3.2. *Let us consider the constraint*

$$G(x) \doteq \begin{bmatrix} 2x_1 + x_2^2 & -x_2^2 \\ -x_2^2 & 2x_1 + x_2^2 \end{bmatrix} \succeq 0$$

and note that, for every orthogonal matrix E in the form

$$E \doteq \begin{bmatrix} a & c \\ b & d \end{bmatrix},$$

we have

$$v_{11}(x, E) = \begin{bmatrix} 2 \\ 2(a-b)^2 x_2 \end{bmatrix} \quad \text{and} \quad v_{22}(x, E) = \begin{bmatrix} 2 \\ 2(c-d)^2 x_2 \end{bmatrix}.$$

Then, at $\overline{x} = 0$, we have $v_{11}(\overline{x}, \overline{E}) = v_{22}(\overline{x}, \overline{E}) = [2, 0]^\top$, so they are linearly dependent, but positively linearly independent for all $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$. However, choosing any sequence $\{x^k\}_{k\in\mathbb{N}} \to 0$ such that $x_2^k \neq 0$ for all k, it follows that the eigenvalues of $G(x_k)$:

$$\lambda_1(G(x^k)) = 2(x_1 + x_2^2) \quad \text{and} \quad \lambda_2(G(x^k)) = 2x_1,$$

are simple, with associated orthonormal eigenvectors

$$u_1(G(x^k)) = \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right] \quad \text{and} \quad u_2(G(x^k)) = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right],$$

respectively, for every $k \in \mathbb{N}$. Then, the only sequence $\{E^k\}_{k\in\mathbb{N}}$ such that $E^k \in \mathcal{E}_r(G(x^k))$ for every k, up to sign, is given by $a = -1/\sqrt{2}$ and $b = c = d = 1/\sqrt{2}$. However, keep in mind that $v_{ii}(x, E)$, $i \in \{1, 2\}$, is invariant to the sign of the columns of E, so $v_{22}(x^k, E^k) = [2, 0]^\top$ and $v_{11}(x^k, E^k) = [2, 4x_2^k]^\top$ are linearly independent for all large k. Therefore, we conclude that (weak-)Robinson's CQ holds at $\overline{x}$, and consequently weak-CPLD also holds, but weak-CRCQ does not hold at $\overline{x}$.

Conversely, we show with another counterexample, that weak-CRCQ does not imply (weak-)Robinson's CQ, and neither does weak-CPLD.

Example 3.3. *Let us consider the constraint*

$$G(x) \doteq \begin{bmatrix} x & x^2 \\ x^2 & -x \end{bmatrix} \succeq 0$$

and the point $\overline{x} = 0$. Take any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ such that $x^k \neq 0$ for every k, and consider two subsequences of it, indexed by I_+ and I_-, such that $x^k > 0$ for every $k \in I_+$, and $x^k < 0$ for every $k \in I_-$. Then, for every $k \in I_+$, we have that:

$$\lambda_1(G(x^k)) = x^k\sqrt{(x^k)^2 + 1} \quad \text{and} \quad \lambda_2(G(x^k)) = -x^k\sqrt{(x^k)^2 + 1},$$

are simple, with associated orthonormal eigenvectors uniquely determined (up to sign) by

$$u_1(G(x^k)) = \frac{1}{\eta_1^k}\left[\frac{1 + \sqrt{(x^k)^2 + 1}}{x^k}, 1\right] \quad \text{and} \quad u_2(G(x^k)) = \frac{1}{\eta_2^k}\left[\frac{1 - \sqrt{(x^k)^2 + 1}}{x^k}, 1\right],$$

where

$$\eta_1^k \doteq \sqrt{\left(\frac{1 + \sqrt{(x^k)^2 + 1}}{x^k}\right)^2 + 1} \quad \text{and} \quad \eta_2^k \doteq \sqrt{\left(\frac{1 - \sqrt{(x^k)^2 + 1}}{x^k}\right)^2 + 1}.$$

Moreover, one can verify that whenever I_+ is an infinite set,

$$\lim_{k\in I_+} u_1(G(x^k)) = [1, 0] \quad \text{and} \quad \lim_{k\in I_+} u_2(G(x^k)) = [0, 1].$$

Then, we have that for all $\overline{E} \in \operatorname{Lim\,sup}_{k\in I_+} \mathcal{E}_r(G(x^k))$, the vectors

$$v_{11}(\overline{x}, \overline{E}) = 1 \quad \text{and} \quad v_{22}(\overline{x}, \overline{E}) = -1$$

are positively linearly dependent. In addition, since $\eta_1^k \to \infty$ and $\eta_2^k \to 1$, the vectors

$$v_{11}(x^k, E^k) = \frac{(\eta_1^k)^2 + 4\sqrt{(x^k)^2 + 1} - 2}{(\eta_1^k)^2} \quad \text{and} \quad v_{22}(x^k, E^k) = \frac{(\eta_2^k)^2 - 4\sqrt{(x^k)^2 + 1} - 2}{(\eta_2^k)^2}$$

are nonzero and have opposite signs; and thus, remain positively linearly dependent, for all large $k \in I_+$.

For the indices $k \in I_-$ the order of $\lambda_1(G(x^k))$ and $\lambda_2(G(x^k))$ is swapped, together with their respective eigenvectors, and we have $\lim_{k\in I_-} u_1(G(x^k)) = [0, 1]$ and $\lim_{k\in I_-} u_2(G(x^k)) = [-1, 0]$. Hence, for all $\overline{E} \in \operatorname{Lim\,sup}_{k\in I_-} \mathcal{E}_r(G(x^k))$, the vectors

$$v_{11}(\overline{x}, \overline{E}) = -1 \quad \text{and} \quad v_{22}(\overline{x}, \overline{E}) = 1$$

are also positively linearly dependent. The order of $v_{11}(x^k, E^k)$ and $v_{22}(x^k, E^k)$ is also swapped, so they remain positively linearly dependent for all large $k \in I_-$.

By the above reasoning, observe that any sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, such that $x^k \neq 0$ for every $k \in \mathbb{N}$, shows that (weak-)Robinson's CQ fails at $\overline{x}$. Moreover, if $x^k = 0$ for infinitely many indices, we may simply take $E^k = \overline{E} = \mathbb{I}_2$ for every k, and then $v_{11}(x^k, E^k) = v_{11}(\overline{x}, \overline{E}) = 1$ and $v_{22}(x^k, E^k) = v_{22}(\overline{x}, \overline{E}) = -1$ are positively linearly dependent for every $k \in \mathbb{N}$. This completes checking that weak-CPLD and weak-CRCQ both hold at $\overline{x}$, while (weak-)Robinson's CQ does not.

Just as it happens in NLP, the weak-CPLD condition is strictly weaker than (weak-)Robinson's CQ, and also weaker than weak-CRCQ, which are in turn, independent. Furthermore, let us establish a formal relationship between weak-CRCQ and weak-CPLD, and their NLP counterparts:

Proposition 3.1. *Let $G(x) \doteq \mathrm{Diag}(g_1(x)\ldots, g_m(x))$ be a structurally diagonal constraint and let $\overline{x}$ be such that $g_1(\overline{x}) \geq 0, \ldots, g_m(\overline{x}) \geq 0$. Then, the following statements are equivalent:*

1. *weak-CRCQ holds at $\overline{x}$;*
2. *For every $J \subseteq \mathcal{A}(\overline{x})$, if the set $\{\nabla g_i(\overline{x})\colon i \in J\}$ is linearly dependent, then $\{\nabla g_i(x)\colon i \in J\}$ is also linearly dependent, for every x close enough to $\overline{x}$;*

where $\mathcal{A}(\overline{x}) \doteq \{i \in \{1, \ldots, m\}\colon g_i(\overline{x}) = 0\}$ is the set of active indices *at $\overline{x}$.*

Proof. Let $r \doteq \mathrm{rank}(G(\overline{x}))$, and note that the result follows trivially if $m = r$. Hence, we will assume that $r < m$. For simplicity, we will also assume that $\mathcal{A}(\overline{x}) = \{1, \ldots, m - r\}$.

- **1 ⇒ 2:** By contradiction, suppose that there is some $J \subseteq \mathcal{A}(\overline{x})$ and a sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ such that $\{\nabla g_i(x^k)\colon i \in J\}$ is linearly independent for every k, but $\{\nabla g_i(\overline{x})\colon i \in J\}$ is not. Let $\{E^k\}_{k\in\mathbb{N}}$ and $\overline{E}$ be the sequence and its limit point described in Definition 3.2, for this particular $\{x^k\}_{k\in\mathbb{N}}$. Note that any other set J' that contains J such that $\{\nabla g_i(x^k)\colon i \in J'\}$ is linearly independent also fits this description, so let us assume that J is maximal.

  Since $G(x^k)$ is diagonal, every eigenvector v^k associated with an eigenvalue λ^k must satisfy $G_{jj}(x^k)v_j^k = \lambda^k v_j^k$ for every $j \in \{1, \ldots, m\}$, which implies $\lambda^k = G_{jj}(x^k)$ or $v_j^k = 0$. Moreover, since G is continuous, the $m - r$ smallest eigenvalues of $G(x^k)$ converge to zero, and consequently, they are bounded from above by
  $$\alpha \doteq \frac{1}{2}\min\{G_{ii}(\overline{x})\colon i \in \{m - r + 1, \ldots, m\}\}$$
  for k large enough. On the other hand, by continuity of G again, the r largest eigenvalues of $G(x^k)$ are bounded from below by α for all k large enough. Hence, it necessarily holds that $v_j^k = 0$ for all $j \in \{m - r + 1, ..., m\}$ and for all k large enough. That is, E^k has the form
  $$E^k = \begin{bmatrix} Q^k \\ 0 \end{bmatrix}, \text{ where } Q^k \in \mathbb{R}^{(m-r)\times(m-r)} \text{ is orthogonal,} \tag{8}$$
  for every k large enough. A simple computation shows us that
  $$v_{ii}(x^k, E^k) = \sum_{j=1}^{m-r} \nabla g_j(x^k)(Q_{ji}^k)^2 \quad \text{and} \quad v_{ii}(\overline{x}, \overline{E}) = \sum_{j=1}^{m-r} \nabla g_j(\overline{x})\overline{Q}_{ji}^2 \tag{9}$$
  for every $i \in \{1, \ldots, m - r\}$, where $\overline{Q}$ is the top $(m - r) \times (m - r)$ submatrix of $\overline{E}$ similarly to (8). Observe that
  $$\mathrm{span}(\{\nabla g_i(x^k)\colon i \in J\}) = \mathrm{span}(\{\nabla g_i(x^k)\colon i \in \{1, \ldots, m - r\}\}),$$
  for all k large enough; otherwise, there would be a subsequence $\{x^k\}_{k\in I} \subseteq \{x^k\}_{k\in\mathbb{N}}$ and another index $j' \in \mathcal{A}(\overline{x}) \setminus J$ such that $\{\nabla g_i(x^k)\colon i \in J \cup \{j'\}\}$ is linearly independent for every $k \in I$, contradicting the maximality of J. Hence, for every $S \subseteq \{1, \ldots, m - r\}$ we have
  $$\mathrm{span}(\{v_{ii}(x^k, E^k)\colon i \in S\}) \subseteq \mathrm{span}(\{\nabla g_i(x^k)\colon i \in J\}) \tag{10}$$
  for every large enough k. In particular, there exists some $S' \subseteq \{1, \ldots, m-r\}$ with the same cardinality as J, such that (10) holds with equality for every large k because J. On the other hand, it follows from (9) that
  $$\mathrm{span}(\{v_{ii}(\overline{x}, \overline{E})\colon i \in S'\}) \subseteq \mathrm{span}(\{\nabla g_i(\overline{x})\colon i \in J\}),$$
  and this implies $\{v_{ii}(\overline{x}, \overline{E})\colon i \in S'\}$ is a linearly dependent set because the rank of $\{\nabla g_i(\overline{x})\colon i \in J\}$ is assumed to be smaller than $|S'| = |J|$. However, knowing that $\{v_{ii}(x^k, E^k)\colon i \in S'\}$ is linearly independent for all k, by weak-CRCQ, we obtain a contradiction.
- **2 ⇒ 1:** Take $Q^k = \mathbb{I}_{m-r}$ and E^k as in (8), so we have $v_{ii}(x^k, E^k) = \nabla g_i(x^k)$ for every $i \in \{1, \ldots, m-r\}$ and every $k \in \mathbb{N}$, and the result follows immediately.

■

Using analogous arguments to the proposition above, we can also prove the following:

Corollary 3.1. *Under the same hypotheses of the previous proposition, the following are equivalent:*

1. *weak-CPLD holds at $\overline{x}$;*
2. *For every $J \subseteq \mathcal{A}(\overline{x})$, if the set $\{\nabla g_i(\overline{x})\colon i \in J\}$ is positively linearly dependent, then $\{\nabla g_i(x)\colon i \in J\}$ is linearly dependent, for every x close enough to $\overline{x}$.*

Proof. Note, in (9), that $v_{ii}(x^k, E^k)$ is generated by a nonnegative linear combination of $\nabla g_i(x^k)$, $i \in \{1, \ldots, m-r\}$. Therefore, every argument in the proof of Proposition 3.1 can be adapted to prove Corollary 3.1. It suffices to consider positive linear independence, instead of linear independence; and the smallest cone generated by $\{v_{ii}(x^k, E^k)\}_{i \in S}$, instead of the smallest subspace. ∎

Advancing to the main result of this section, which is to prove that weak-CPLD (and therefore, weak-CRCQ) guarantees the existence of Lagrange multipliers at all local solutions of (NSDP), we get inspiration in the proof of [12, Thm. 3.1] for NLP, and the proof of [6, Thm. 3.2]. That is, we analyse the sequence from Theorem 2.1 in terms of the spectral decomposition of its approximate Lagrange multiplier candidates, under weak-CPLD. Then, we use Carathéodory's Lemma 2.1 to construct a bounded sequence from it, that converges to a Lagrange multiplier. As an intermediary step, we also obtain a convergence result of the external penalty method to KKT points under weak-CPLD, a fact that is emphasized in the statement of the next theorem.

Theorem 3.1. *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*

$$\nabla_x L\left(x^k, \rho_k \Pi_{\mathbb{S}^m_+}(-G(x^k))\right) \to 0.$$

If $\overline{x}$ satisfies weak-CPLD, then $\overline{x}$ satisfies the KKT conditions. In particular, every local minimizer of (NSDP) that satisfies weak-CPLD also satisfies KKT.

Proof. Let $Y^k \doteq \rho_k \Pi_{\mathbb{S}^m_+}(-G(x^k))$, for every $k \in \mathbb{N}$. Recall that we assume $\lambda_1(-G(x^k)) \geq \ldots \geq \lambda_m(-G(x^k))$, for every k, and denote by r the rank of $\operatorname{Ker} G(\overline{x})$. Note that when k is large enough, say greater than some k_0, we necessarily have $\lambda_i(-G(x^k)) = -\lambda_{m-i+1}(G(x^k)) < 0$ for all $i \in \{m-r+1, \ldots, m\}$. Let $I \subseteq_\infty \mathbb{N}$, and $\{E^k\}_{k\in I} \to \overline{E}$ be such that $E^k \in \mathcal{E}_r(G(x^k))$ for every $k \in I$, as described in Definition 3.2. Then, for each $k \in I$ greater than k_0, the spectral decomposition of Y^k is given by

$$Y^k = \sum_{i=1}^{m-r} \alpha_i^k e_i^k (e_i^k)^\top,$$

where $\alpha_i^k \doteq [\rho_k \lambda_i(-G(x^k))]_+ \geq 0$ and e_i^k denotes the i-th column of E^k, for every $i \in \{1, \ldots, m-r\}$. Since $\nabla_x L(x^k, Y^k) \to 0$, we have

$$\nabla f(x^k) - \sum_{i=1}^{m-r} \alpha_i^k DG(x^k)^* \left[e_i^k (e_i^k)^\top\right] \to 0, \tag{11}$$

but note that

$$DG(x^k)^* \left[e_i^k (e_i^k)^\top\right] = \begin{bmatrix} \langle D_{x_1} G(x^k), e_i^k (e_i^k)^\top \rangle \\ \vdots \\ \langle D_{x_n} G(x^k), e_i^k (e_i^k)^\top \rangle \end{bmatrix} = \begin{bmatrix} (e_i^k)^\top D_{x_1} G(x^k) e_i^k \\ \vdots \\ (e_i^k)^\top D_{x_n} G(x^k) e_i^k \end{bmatrix} = v_{ii}(x^k, E^k),$$

so we can rewrite (11) as

$$\nabla f(x^k) - \sum_{i=1}^{m-r} \alpha_i^k v_{ii}(x^k, E^k) \to 0.$$

Using Carathéodory's Lemma 2.1 for the family $\{v_{ii}(x^k, E^k)\}_{i\in\{1,\ldots,m-r\}}$, for each fixed $k \in I$, we obtain some $J^k \subseteq \{1, \ldots, m-r\}$ such that $\{v_{ii}(x^k, E^k)\}_{i\in J^k}$ is linearly independent and

$$\nabla f(x^k) - \sum_{i=1}^{m-r} \alpha_i^k v_{ii}(x^k, E^k) = \nabla f(x^k) - \sum_{i\in J} \tilde{\alpha}_i^k v_{ii}(x^k, E^k), \tag{12}$$

where $\tilde{\alpha}_i^k \geq 0$ for every $k \in I$ and every $i \in J^k$. By the pigeonhole principle, we can assume J^k is the same, say equal to J, for all $k \in I$ large enough. We claim that the sequences $\{\tilde{\alpha}_i^k\}_{k\in I}$ are all bounded. In order to prove this, suppose that

$$m^k \doteq \max_{i\in J}\{\tilde{\alpha}_i^k\}$$

is unbounded with $k \in I$, divide (12) by m^k and note that $m^k \to \infty$ on a subsequence implies that the vectors $v_{ii}(\overline{x}, \overline{E})$, $i \in J$, are positively linearly dependent. On the other hand, the vectors $v_{ii}(x^k, E^k)$, $i \in J$, are linearly independent for all large k, which contradicts weak-CPLD. Finally, note that every collection of limit points $\{\overline{\alpha}_i : i \in J\}$ of their respective sequences $\{\tilde{\alpha}_i^k\}_{k\in\mathbb{N}}$, $i \in J$, generates a Lagrange multiplier associated with $\overline{x}$, which is $\overline{Y} \doteq \sum_{i\in J} \overline{\alpha}_i u_i(G(\overline{x})) u_i(G(\overline{x}))^\top$. Here, $u_i(G(\overline{x}))$ is the eigenvector associated with the i-th eigenvalue of $G(\overline{x})$ (see (1)). Thus, $\overline{x}$ is a KKT point.

The second part of the statement of the theorem follows from Theorem 2.1. ∎

Back to Example 3.1, observe that weak-CPLD does not hold at $\overline{x} = 0$, as expected. Indeed, for any sequence $\{x^k\}_{k\in\mathbb{N}} \to 0$ such that $x^k < 0$ for all k, the matrix $G(x^k)$ has only simple eigenvalues, for all large k, so $E^k \in \mathcal{E}_r(G(x^k))$ is unique up to sign. Without loss of generality, we can assume

$$E^k \doteq \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix},$$

and then we have $v_{11}(x^k, E^k) = -2x^k > 0$, which is linearly independent for all k while $v_{11}(\overline{x}, \overline{E}) = 0$ is positively linearly dependent. Thus Definition 3.2 is not satisfied.

Remark 3.1. *In [7], we presented a different extension proposal of CRCQ (and CPLD) to NSDP problems with multiple constraints, which is weaker than nondegeneracy (respectively, Robinson's CQ) for a single constraint as in (NSDP) only when the zero eigenvalue of $G(\overline{x})$ is simple. We called this definition the "naive extension of CRCQ (and CPLD)". We remark that Definition 3.2 coincides with the naive extension of CRCQ (and CPLD) when zero is a simple eigenvalue of $G(\overline{x})$, which makes Definition 3.2 an improvement of it, or a "non-naive variant" of it.*

The phrasing of Theorem 3.1 was chosen to draw the reader's attention to the fact that it is, essentially, a convergence proof of the external penalty method to KKT points, under weak-CPLD. To obtain a more general convergence result, in the next section we introduce new constant rank-type CQs for NSDP that support every algorithm that converges with a more general type of sequential optimality condition. Then, we prove some properties of these new conditions, and we compare them with weak-CPLD and weak-CRCQ.

4 Stronger sequential-type constant rank CQs for NSDP and global convergence of algorithms

A more general sequential optimality condition, which was brought from NLP to NSDP by Andreani et al. [10], is the so-called *Approximate Karush-Kuhn-Tucker* (AKKT) condition. Such a notion has been recently refined and generalized to optimization problems in Banach spaces [16] and also problems with geometric constraints [23]. Let us recall one (the most convenient for this paper) of its many characterizations[1].

Definition 4.1 (Def. 4 of [4])**.** *We say that a point $\overline{x} \in \mathcal{F}$ satisfies the AKKT condition when there exist sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and $\{Y^k\}_{k\in\mathbb{N}} \subseteq \mathbb{S}^m_+$, and perturbation sequences $\{\delta^k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^n$ and $\{\Delta^k\}_{k\in\mathbb{N}} \subseteq \mathbb{S}^m$, such that:*

1. $\nabla_x L(x^k, Y^k) = \delta^k$*, for every* $k \in \mathbb{N}$*;*
2. $G(x^k) + \Delta^k \succeq 0$ *and* $\langle G(x^k) + \Delta^k, Y^k\rangle = 0$*, for every* $k \in \mathbb{N}$*;*
3. $\Delta^k \to 0$ *and* $\delta^k \to 0$*.*

Note that $\{Y^k\}_{k\in\mathbb{N}}$ is a sequence of approximate Lagrange multipliers of $\overline{x}$, in the sense that Y^k is an exact Lagrange multiplier, at $x = x^k$, for the perturbed problem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x) + \langle \overline{x} - x, \delta^k\rangle, \\ \text{subject to} \quad & G(x) + \Delta^k \succeq 0. \end{aligned}$$

The main goal in enlarging the class of approximate Lagrange multipliers Y^k and perturbations Δ^k as in Definition 4.1 instead of considering only the ones given by Theorem 2.1, is to capture the output sequences of a larger class of iterative algorithms. In the next two subsections, we illustrate the previous statement. What is remarkable is that the proof of Theorem 3.1 can still be somewhat conducted considering this more general class of sequences, arriving at strong global convergence results for such algorithms (Theorem 4.2).

4.1 Example 1: A safeguarded augmented Lagrangian method

Let us briefly recall a variant of the *Powell-Hestenes-Rockafellar* augmented Lagrangian algorithm that employs a *safeguarding* technique, which is the direct generalization of the one studied in [14]. The variant we use is also a generalization of [8, Pg. 13] and [10, Alg. 1], for instance.

For an arbitrary penalty parameter $\rho > 0$ and a *safeguarded multiplier* $\tilde{Y} \succeq 0$, we define $L_{\rho,\tilde{Y}} : \mathbb{R}^n \to \mathbb{R}$ as the *Augmented Lagrangian function* of (NSDP), which is given by

$$L_{\rho,\tilde{Y}}(x) \doteq f(x) + \frac{\rho}{2}\left\|\Pi_{\mathbb{S}^m_+}\left(-G(x) + \frac{\tilde{Y}}{\rho}\right)\right\|^2 - \frac{1}{2\rho}\left\|\tilde{Y}\right\|^2.$$

[1]Definition 4.1 coincides with the AKKT condition presented in [10, Def. 3.1]. See, for instance, [4, Prop. 4].

Since it will be useful in the convergence proof, we compute the gradient of $L_{\rho,\tilde{Y}}$ at x below:

$$\nabla L_{\rho,\tilde{Y}}(x) = \nabla f(x) - DG(x)^* \left[\rho \Pi_{\mathbb{S}^m_+}\left(-G(x) + \frac{\tilde{Y}}{\rho}\right)\right]. \tag{13}$$

Now, we state the algorithm:

Algorithm 1 Safeguarded augmented Lagrangian method

Input: A sequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$ of positive scalars such that $\varepsilon_k \to 0$; a nonempty convex compact set $\mathcal{B} \subset \mathbb{S}^m_+$; real parameters $\tau > 1$, $\sigma \in (0,1)$, and $\rho_1 > 0$; and initial points $(x^0, \tilde{Y}^1) \in \mathbb{R}^n \times \mathcal{B}$. Also, define $\|V^0\| = \infty$.

Initialize $k \leftarrow 1$. Then:

Step 1 *(Solving the subproblem):* Compute an approximate stationary point x^k of $L_{\rho_k,\tilde{Y}^k}(x)$, that is, a point x^k such that

$$\|\nabla L_{\rho_k,\tilde{Y}^k}(x^k)\| \leq \varepsilon_k;$$

Step 2 *(Updating the penalty parameter):* Calculate

$$V^k \doteq \Pi_{\mathbb{S}^m_+}\left(-G(x^k) + \frac{\tilde{Y}^k}{\rho_k}\right) - \frac{\tilde{Y}^k}{\rho_k}; \tag{14}$$

Then,

a. If $k = 1$ or $\|V^k\| \leq \tau\|V^{k-1}\|$, set $\rho_{k+1} \doteq \rho_k$;

b. Otherwise, take ρ_{k+1} such that $\rho_{k+1} \geq \gamma\rho_k$.

Step 3 *(Estimating a new safeguarded multiplier):* Choose any $\tilde{Y}^{k+1} \in \mathcal{B}$, set $k \leftarrow k+1$ and go to Step 1.

By the definition of projection we have that $\tilde{Y}^k = \Pi_{\mathbb{S}^m_+}(\tilde{Y}^k - \rho_k G(x^k))$ if, and only if, $\tilde{Y}^k, G(x^k) \in \mathbb{S}^m_+$ and $\langle \tilde{Y}^k, G(x^k)\rangle = 0$, which means that $V^k = 0$ if, and only if, the pair $(x^k, \tilde{Y}^k)$ is primal-dual feasible and complementary. Moreover, note that Algorithm 1 does not necessarily keep a record of the approximate multiplier sequence associated with $\{x^k\}_{k\in\mathbb{N}}$, which is

$$Y^k \doteq \rho_k \Pi_{\mathbb{S}^m_+}\left(-G(x^k) + \frac{\tilde{Y}^k}{\rho_k}\right). \tag{15}$$

These are usually computed, however, in several practical implementations of it, where $\tilde{Y}^{k+1}$ is chosen as the projection of Y^k onto $\mathcal{B}$. Also, with these multipliers at hand, it is very easy to prove that any feasible limit point $\overline{x}$ of $\{x^k\}_{k\in\mathbb{N}}$ must satisfy AKKT:

Theorem 4.1. *Fix any choice of parameters in Algorithm 1 and let $\{x^k\}_{k\in\mathbb{N}}$ be the output sequence generated by it. If $\{x^k\}_{k\in\mathbb{N}}$ has a convergent subsequence $\{x^k\}_{k\in I} \to \overline{x}$, then:*

1. *The point $\overline{x}$ is stationary for the problem of minimizing $\frac{1}{2}\|\Pi_{\mathbb{S}^m_+}(-G(x))\|^2$;*
2. *If $\overline{x}$ is feasible, then $\overline{x}$ satisfies AKKT.*

Proof. Let $\{\varepsilon_k\}_{k\in\mathbb{N}} \to 0$, $\{\tilde{Y}^k\}_{k\in\mathbb{N}} \subset \mathcal{B} \subset \mathbb{S}^m_+$, $\tau > 1$, $\sigma \in (0,1)$, and $\rho_1 > 0$ be the fixed input parameters of Algorithm 1. Moreover, let $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{V^k\}_{k\in\mathbb{N}}$ computed as in Step 2. For simplicity, let us also assume that $I = \mathbb{N}$.

1. This part of the proof is standard; see, for instance, [4, Prop. 4.3];
2. Define $\{Y^k\}_{k\in\mathbb{N}}$ as in (15) and take $\Delta^k \doteq V^k$ for all $k \in \mathbb{N}$, where V^k is as given in (14). Then, it follows from Step 1 that $\nabla_x L(x^k, Y^k) = \nabla L_{\rho_k,\tilde{Y}^k}(x^k) \to 0$. We also have

$$G(x^k) + \Delta^k = \Pi_{\mathbb{S}^m_+}\left(G(x^k) - \frac{\tilde{Y}^k}{\rho_k}\right)$$

for every $k \in \mathbb{N}$, which yields $\langle Y^k, G(x^k) + \Delta^k\rangle = 0$ for every k. If $\rho_k \to \infty$, then recall that $\{\tilde{Y}^k\}_{k\in\mathbb{N}}$ remains bounded and thus $V^k \to \Pi_{\mathbb{S}^m_+}(-G(\overline{x}))$ by definition and $\Pi_{\mathbb{S}^m_+}(-G(\overline{x})) = 0$ because $\overline{x}$ is assumed to be feasible; on the other hand, if ρ_k remains bounded, then $V^k \to 0$ due to Step 2-a. Therefore, $\Delta^k \to 0$ and $\overline{x}$ satisfies AKKT.

■

Note that when $\tilde{Y}^k$ is set as zero for every k, then Algorithm 1 reduces to the external penalty method, meaning Theorem 4.1 also covers this method.

4.2 Example 2: A sequential quadratic programming method

Next, we recall Correa and Ramírez's [18] *sequential quadratic programming* (SQP) method:

Algorithm 2 General SQP method

Input: A real parameter $\tau > 1$, a pair of initial points $(x^1, Y^1) \in \mathbb{R}^n \times \mathbb{S}^m_+$, and an approximation of $\nabla^2_x L(x^1, Y^1)$ given by H^1.

Initialize $k \leftarrow 1$. Then:

Step 1 *(Solving the subproblem):* Compute a solution d^k, together with its Lagrange multiplier Y^{k+1}, of the problem

$$\begin{aligned} \underset{d\in\mathbb{R}^n}{\text{Minimize}} \quad & d^\top H^k d + \nabla f(x^k)^\top d, \\ \text{subject to} \quad & G(x^k) + DG(x^k)d \in \mathbb{S}^m_+, \end{aligned} \tag{Lin-QP}$$

and if $d^k = 0$, stop;

Step 2 *(Step corrections):* Perform line search to find a steplength $\alpha^k \in (0,1)$ satisfying *Armijo's rule*

$$f(x^k + \alpha^k d^k) - f(x^k) \leq \tau \alpha^k \nabla f(x^k)^\top d^k. \tag{16}$$

Step 3 *(Approximating the Hessian):* Set $x^{k+1} \leftarrow x^k + \alpha^k d^k$, compute a positive definite approximation H^{k+1} of $\nabla^2_x L(x^{k+1}, Y^{k+1})$, set $k \leftarrow k+1$, and go to Step 1.

The SQP algorithm generates AKKT sequences as well, as it can be seen in the following proposition:

Proposition 4.1. *Assume that Step 1 of Algorithm 2 is always well defined. If there is an infinite subset $I \subseteq_\infty \mathbb{N}$ such that $\lim_{k\in I} d^k = 0$ and $\{\|H^k\|\}_{k\in I}$ is bounded, then any limit point $\overline{x}$ of $\{x^k\}_{k\in I}$ satisfies AKKT.*

Proof. By the KKT conditions for (Lin-QP), there exists some $Y^k \succeq 0$ such that

$$\nabla f(x^k) + H^k d^k - DG(x^k)^*[Y^k] = 0 \tag{17}$$

$$\langle G(x^k) + DG(x^k)d^k, Y^k \rangle = 0. \tag{18}$$

Set $\Delta^k \doteq DG(x^k)d^k$ for every $k \in I$ and since $d^k \to 0$, we obtain that $\lim_{k\in I} H^k d^k = 0$ and $\lim_{k\in I} \Delta^k = 0$. Moreover, since d^k is feasible, $G(x^k) + \Delta^k \succeq 0$. Thus, $\overline{x}$ satisfies AKKT. ■

The hypothesis on the convergence of a subsequence of $\{d^k\}_{k\in\mathbb{N}}$ to zero, directly or indirectly, is somewhat common regarding some types of SQP methods, as well as the boundedness of H^k – see, for instance, [9, 18, 27].

4.3 Sequential constant rank CQs for NSDP

Inspired by AKKT, we are led to introduce a small perturbation in weak-CPLD and weak-CRCQ, which makes it stronger, but also brings some useful properties in return. At first, we present it in a form that most resembles Definition 3.2, for comparison purposes. Later, for convenience, we will provide a characterization of it without sequences.

Definition 4.2 (seq-CRCQ and seq-CPLD)**.** *Let $\overline{x} \in \mathcal{F}$ and let r be the rank of $G(\overline{x})$. We say that $\overline{x}$ satisfies the*

1. Sequential CRCQ condition for NSDP *(seq-CRCQ) if $r = m$ or, for all sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and $\{\Delta^k\}_{k\in\mathbb{N}} \subseteq \mathbb{S}^m$ with $\Delta^k \to 0$, there exists $\overline{E} \in \operatorname{Lim\,sup}_{k\in\mathbb{N}} \mathcal{E}_r(G(x^k) + \Delta^k)$ such that, for every subset $J \subseteq \{1, \ldots, m-r\}$: if the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i\in J}$ is linearly dependent, then $\{v_{ii}(x^k, E^k)\}_{i\in J}$ remains linearly dependent, for all $k \in I$ large enough.*

2. Sequential CPLD condition for NSDP *(seq-CPLD) if $r = m$ or, for all sequences $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and $\{\Delta^k\}_{k\in\mathbb{N}} \subseteq \mathbb{S}^m$ with $\Delta^k \to 0$, there exists $\overline{E} \in \operatorname{Lim\,sup}_{k\in\mathbb{N}} \mathcal{E}_r(G(x^k)+\Delta^k)$ such that, for every subset $J \subseteq \{1,\dots,m-r\}$: if $\{v_{ii}(\overline{x},\overline{E})\}_{i\in J}$ is positively linearly dependent, then $\{v_{ii}(x^k,E^k)\}_{i\in J}$ remains linearly dependent, for all $k \in I$ large enough.*

For both definitions, $I \subseteq_\infty \mathbb{N}$, and $\{E^k\}_{k\in I}$ is a sequence converging to $\overline{E}$ and such that $E^k \in \mathcal{E}_r(G(x^k)+\Delta^k)$ for every $k \in I$, as required by the Painlevé-Kuratowski outer limit.

Note that the only difference between Definitions 3.2 and 4.2 is the perturbation matrix $\Delta^k \to 0$. In particular, set $\Delta^k \doteq 0$ for every k to see that seq-CRCQ and seq-CPLD imply weak-CRCQ and weak-CPLD, respectively. Moreover, both implications are strict, as we can see in the following example:

Example 4.1. *Consider the constraint*

$$G(x) \doteq \begin{bmatrix} x & 0 \\ 0 & -x \end{bmatrix} \succeq 0$$

at the point $\overline{x} = 0$, so in this case $r = 0$. For every sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, we have (up to sign)

$$\mathcal{E}_r(G(x^k)) = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\},$$

for every $k \in \mathbb{N}$ such that $x^k \neq \overline{x}$, whereas if $x^k = \overline{x}$, then $\mathcal{E}_r(G(x^k))$ is the set of all orthogonal 2×2 matrices. Take $E^k = \mathbb{I}_2$ for every $k \in \mathbb{N}$ to see that both, weak-CRCQ and weak-CPLD, hold at $\overline{x}$, since

$$v_{11}(x^k,E^k) = 1 \quad \text{and} \quad v_{22}(x^k,E^k) = -1$$

are nonzero and (positively) linearly dependent for every $k \in \mathbb{N}$.

On the other hand, take

$$\Delta^k \doteq \begin{bmatrix} 2x^k(x^k+1)^2 & x^k(x^k+1) \\ x^k(x^k+1) & 3x^k + x^k(x^k+1)^2 \end{bmatrix},$$

and note that the eigenvectors of

$$\begin{aligned} G(x^k) + \Delta^k &= x^k \begin{bmatrix} 2(x^k)^2 + 4x^k + 3 & x^k+1 \\ x^k+1 & (x^k)^2 + 2x^k + 3 \end{bmatrix} \\ &= \begin{bmatrix} -1 & x^k+1 \\ x^k+1 & 1 \end{bmatrix} \begin{bmatrix} x^k & 0 \\ 0 & 2x^k \end{bmatrix} \begin{bmatrix} -1 & x^k+1 \\ x^k+1 & 1 \end{bmatrix} \end{aligned}$$

are uniquely determined up to sign whenever $x^k \neq 0$. Then, because $v_{ii}(x,E)$, $i \in \{1,2\}$, is invariant to the sign of the columns of E, we can assume without loss of generality that any $E^k \in \mathcal{E}_r(G(x^k)+\Delta^k)$ has the form

$$E^k = \frac{1}{\sqrt{1+(x^k+1)^2}} \begin{bmatrix} -1 & x^k+1 \\ x^k+1 & 1 \end{bmatrix}$$

for every $k \in \mathbb{N}$, if $x^k \neq 0$. Then, for any sequence $\{E^k\}_{k\in\mathbb{N}}$ such that $E^k \in \mathcal{E}_r(G(x^k)+\Delta^k)$ for every k, we have

$$v_{11}(x^k,E^k) = 1 - (x^k+1)^2 \quad \text{and} \quad v_{22}(x^k,E^k) = (x^k+1)^2 - 1,$$

which are both nonzero whenever $x^k \neq 0$, but if $\overline{E}$ is a limit point of $\{E^k\}_{k\in\mathbb{N}}$, then $v_{11}(\overline{x},\overline{E}) = v_{22}(\overline{x},\overline{E}) = 0$. Thus, neither seq-CRCQ nor seq-CPLD hold at $\overline{x}$.

Furthermore, since nondegeneracy can be characterized as the linear independence of $v_{ii}(\overline{x},\overline{E})$, $i \in \{1,\dots,m-r\}$, for every $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$ [6, Prop. 3.2], we observe that it implies seq-CRCQ (see also Remark 4.1 at the end of this section), but this implication is also strict. Let us show this with a counterexample.

Example 4.2. *We analyse the constraint*

$$G(x) \doteq \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix} \succeq 0$$

at the point $\overline{x} \doteq 0$. For any $x \in \mathbb{R}$ and any arbitrary orthogonal matrix $E \in \mathbb{R}^{2\times 2}$, note that E has the form

$$E = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \text{ if } \det(E) = 1 \quad \text{or} \quad E = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}, \text{ if } \det(E) = -1 \tag{19}$$

where $a^2 + b^2 = 1$. In both cases, we have

$$v_{11}(x,E) = v_{22}(x,E) = a^2 + b^2 = 1.$$

That is, $v_{11}(x,E)$ and $v_{22}(x,E)$ are nonzero and linearly dependent, regardless of x and E. Thus, seq-CRCQ holds at $\overline{x}$, although nondegeneracy does not. Note that weak-nondegeneracy also fails at $\overline{x}$, in this example.

By Example 3.2, we verify that Robinson's CQ does not imply seq-CRCQ; because otherwise, it would also imply weak-CRCQ, contradicting the example. As for the converse, the counterexample below shows that seq-CRCQ does not imply Robinson's CQ either.

Example 4.3. *Consider the constraint*

$$G(x) \doteq \begin{bmatrix} x_1 & x_2 \\ x_2 & -x_1 \end{bmatrix} \succeq 0.$$

Clearly, the only feasible point is $\overline{x} = 0$. Then, due to the linearity of G, it is immediate to see that Robinson's CQ does not hold at $\overline{x} = 0$. On the other hand, for any $x \in \mathbb{R}^2$ and any orthogonal matrix $E \in \mathbb{R}^{2\times 2}$, note that regardless of the form of E as in (19), we have $v_{11}(x, E) \neq 0$, $v_{22}(x, E) \neq 0$, and

$$v_{11}(x, E) = -v_{22}(x, E).$$

Thus, seq-CRCQ holds at $\overline{x} = 0$; see also the characterization of Proposition 4.2.

The same argument that is used to verify that nondegeneracy implies seq-CRCQ can also be adapted to show that Robinson's CQ implies seq-CPLD in view of [6, Prop. 5.1]. Then, Example 4.3 tells us that seq-CPLD is actually strictly weaker than Robinson's CQ.

Next, we will show that seq-CPLD (and, consequently, seq-CRCQ) is enough to establish equivalence between AKKT and KKT with a small adaptation of the proof of Theorem 3.1. Note that in view of Theorem 2.1, any condition that establishes that an AKKT point is also a KKT point is, in particular, a CQ; in addition, such a CQ necessarily supports the global convergence of any algorithm supported by AKKT to KKT points. This includes the algorithms presented in Subsections 4.1 and 4.2, and Yamashita, Yabe, and Harada's primal-dual interior point method for NSDP [32] – for details on the latter, see [3]. We should also stress that this convergence result neither assumes compactness of the Lagrange multiplier set nor that it is a singleton.

Theorem 4.2. *Let $\overline{x} \in \mathcal{F}$ be an AKKT point that satisfies seq-CPLD. Then, $\overline{x}$ satisfies KKT.*

Proof. Let $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$, $\{Y^k\}_{k\in\mathbb{N}} \subseteq \mathbb{S}^m_+$, and $\{\tilde{\Delta}^k\}_{k\in\mathbb{N}} \to 0$ be the AKKT sequences from Definition 4.1. Let r is the rank of $G(\overline{x})$ and recall that $\langle G(x^k) + \tilde{\Delta}^k, Y^k\rangle = 0$ implies that $G(x^k) + \tilde{\Delta}^k$ and Y^k are simultaneously diagonalizable by some orthogonal matrix $U^k \doteq [u_1^k, \ldots, u_m^k] \in \mathbb{R}^{m\times m}$, such that

$$\begin{aligned} 0 = \langle G(x^k) + \tilde{\Delta}^k, Y^k\rangle &= \left\langle (U^k)^\top (G(x^k) + \tilde{\Delta}^k) U^k,\ (U^k)^\top Y^k U^k \right\rangle \\ &= \sum_{i=1}^m \lambda_i(G(x^k) + \tilde{\Delta}^k)\lambda_{m-i+1}(Y^k) \end{aligned}$$

and because all terms above are nonnegative it follows that, for every $i \in \{1, \ldots, m\}$, the product $\lambda_i(G(x^k) + \tilde{\Delta}^k)\lambda_{m-i+1}(Y^k)$ is equal to zero. Also, because $\lambda_i(G(x^k)) > 0$ for every $i \in \{1, \ldots, r\}$, then $\lambda_i(G(x^k) + \tilde{\Delta}^k) > 0$ and $\lambda_{m-i+1}(Y^k) = 0$ for every such i and all k large enough.

Hence, assuming without loss of generality that the columns of U^k are ordered accordingly to the eigenvalues of Y^k, the spectral decomposition of Y^k can be represented in the format

$$Y^k = \sum_{i=1}^{m-r} \lambda_i(Y^k) u_i^k (u_i^k)^\top$$

Defining $E^k = [u_1^k, \ldots, u_{m-r}^k]$ for every k, we obtain

$$\nabla_x L(x^k, Y^k) = \nabla f(x^k) - \sum_{i=1}^{m-r} \lambda_i(Y^k) v_{ii}(x^k, E^k) \to 0.$$

Also, for each $k \in \mathbb{N}$, construct

$$M^k \doteq U^k \left[\begin{array}{c|c} \mathrm{Diag}(\lambda_1(G(x^k)), \ldots, \lambda_r(G(x^k))) & 0 \\ \hline 0 & \mathrm{Diag}((r+1)\|x^k - \overline{x}\|, \ldots, m\|x^k - \overline{x}\|) \end{array}\right] (U^k)^\top. \tag{20}$$

Note that $M^k \to G(\overline{x})$ and that the $m - r$ smallest eigenvalues of M^k are simple, if $x^k \neq \overline{x}$, meaning their associated eigenvectors are unique up to sign, when k is large enough. Consequently, $v_{ii}(x^k, E^k)$ is invariant to the choice of $E^k \in \mathcal{E}_r(M^k)$, for all such k, and every $i \in \{1, \ldots, m - r\}$. Define $\Delta^k \doteq M^k - G(x^k)$ for every $k \in \mathbb{N}$ and the rest of this proof follows the exact same lines as the proof of Theorem 3.1. ■

Remark 4.1. *The "perturbed versions" of weak-nondegeneracy and weak-Robinson's CQ, in the sense of Definition 4.2, are nondegeneracy and Robinson's CQ, respectively. In other words, nondegeneracy (respectively, Robinson's CQ) holds at $\overline{x} \in \mathcal{F}$ if, and only if, for every sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ and every $\{\Delta^k\}_{k\in\mathbb{N}} \subseteq \mathbb{S}^m$ such that $\Delta^k \to 0$, there is some $\overline{E} \in \mathrm{Lim\,sup}_{k\in\mathbb{N}} \mathcal{E}_r(G(x^k) + \Delta^k)$ such that $\{v_{ii}(\overline{x}, \overline{E}) \colon i \in \{1, \ldots, m - r\}\}$ is (positively) linearly independent, where $r = \mathrm{rank}(G(\overline{x}))$. For more details, see [6, Rem. 3.1].*

We end this section with a characterization of seq-CRCQ and seq-CPLD without sequences, which may be better suited for some theory-focused applications.

Proposition 4.2. *Let $\overline{x} \in \mathcal{F}$ and let r be the rank of $G(\overline{x})$.*

- *seq-CRCQ holds at $\overline{x}$ if, and only if, $r = m$ or, for every $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$, there exists some neighborhood $\mathcal{V}$ of $(\overline{x}, \overline{E})$ such that for all $J \subseteq \{1, \dots, m-r\}$, we have that if the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is linearly dependent, then $\{v_{ii}(x, E)\}_{i \in J}$ remains linearly dependent for every $(x, E) \in \mathcal{V}$ such that E has orthonormal columns;*
- *seq-CPLD holds at $\overline{x}$ if, and only if, $r = m$ or, for every $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$, there exists some neighborhood $\mathcal{V}$ of $(\overline{x}, \overline{E})$ such that for all $J \subseteq \{1, \dots, m-r\}$, we have that if the family $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is positively linearly dependent, then $\{v_{ii}(x, E)\}_{i \in J}$ remains linearly dependent for every $(x, E) \in \mathcal{V}$ such that E has orthonormal columns.*

Proof. We will prove only item 1, since item 2 follows analogously. Let $\overline{x}$ satisfy seq-CRCQ; by contradiction: suppose that there exists some $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$, some $J \subseteq \{1, \dots, m-r\}$, and some sequence $\{(x^k, E^k)\}_{k \in \mathbb{N}} \to (\overline{x}, \overline{E})$ such that $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is linearly dependent, but $\{v_{ii}(x^k, E^k)\}_{i \in J}$ is linearly independent for every large $k \in \mathbb{N}$. Let $P^k \in \mathbb{R}^{m \times r}$ be a matrix whose columns are orthogonal eigenvectors associated with the r largest eigenvalues of $G(x^k)$, define $U^k \doteq [P^k, E^k]$, and consider M^k as in (20). Set $\Delta^k \doteq M^k - G(x^k)$ and note that $v_{ii}(x^k, E^k)$ is invariant to $E^k \in \mathcal{E}_r(\Delta^k + G(x^k))$ when k is large, provided that $x^k \neq \overline{x}$. This contradicts seq-CRCQ.

Conversely, let $\{x^k\}_{k \in \mathbb{N}} \to \overline{x}$ and $\Delta^k \to 0$ be any sequences, and let $J \subseteq \{1, \dots, m-r\}$ be any subset. For each k, pick any $E^k \in \mathcal{E}_r(G(x^k) + \Delta^k)$ and consider the sequence $\{E^k\}_{k \in \mathbb{N}}$, which is bounded[2]. Let $I \subseteq_\infty \mathbb{N}$ and $\overline{E}$ be arbitrary, as long as $\{E^k\}_{k \in I} \to \overline{E}$, so $\overline{E} \in \mathcal{E}_r(G(\overline{x}))$. Then, by hypothesis, there exists a neighborhood $\mathcal{V}$ of $(\overline{x}, \overline{E})$ such that if $\{v_{ii}(\overline{x}, \overline{E})\}_{i \in J}$ is linearly dependent, then $\{v_{ii}(x^k, E^k)\}_{i \in J}$ is also linearly dependent for all large enough $k \in I$, since $(x^k, E^k) \in \mathcal{V}$ for all such k. ■

5 Relationship with metric subregularity CQ

Besides convergence of algortihms, the CQs we present also have implications towards stability and error analysis. We make this link by means of establishing a relationship between seq-CPLD (and seq-CRCQ) and the so-called *metric subregularity CQ* (also known as the *error bound CQ* in NLP), defined in our SDP framework as follows:

Definition 5.1 (e.g., Def. 1.1 of [19])**.** *We say that a feasible point $\overline{x}$ of (NSDP) satisfies the* metric subregularity CQ *when there exists some $\gamma > 0$ and a neighborhood $\mathcal{V}$ of $\overline{x}$ such that*

$$\operatorname{dist}(x, \mathcal{F}) \leq \gamma \|\Pi_{\mathbb{S}^m_+}(-G(x))\|$$

for every $x \in \mathcal{V}$. That is, when the set-valued mapping $\mathcal{G} \colon \mathbb{R}^n \rightrightarrows \mathbb{S}^m$ that maps $x \mapsto G(x) - \mathbb{S}^m_+$ is metrically subregular at $(\overline{x}, 0) \in \operatorname{graph}(\mathcal{G})$. Here $\operatorname{dist}(x, \mathcal{F})$ denotes the distance between x and $\mathcal{F}$, and $\operatorname{graph}(\mathcal{G}) \subset \mathbb{R}^n \times \mathbb{S}^m$ is the graph of $\mathcal{G}$.

The metric subregularity CQ is implied by Robinson's CQ, which in turn coincides with a similar condition called *metric regularity CQ*, and it has relevant implications on the stability analysis of optimization problems – for details, we refer to Ioffe's survey [20, 21]. Besides, there are several works addressing the relationship between constant rank constraint qualifications and the metric subregularity CQ in NLP, such as Minchenko and Stakhovski [24], Andreani et al. [8], and others.

We will use a sufficient condition for metric subregularity CQ to hold, originally proposed by Minchenko and Stakhovski [24, Thm. 2] for NLP problems. We made a simple extension of it to NSDP, which seems not to have been done before in the literature. It is worth mentioning, nevertheless, that the proof we present is essentially the same as the original one, with some minor adaptations to the NSDP context via Moreau's decomposition.

Proposition 5.1. *Let $\overline{x} \in \mathcal{F}$ and assume that G is twice continuously differentiable around $\overline{x}$. For every given $x \in \mathbb{R}^n$, let $\Lambda_\Pi(x)$ denote the set of Lagrange multipliers of the problem of minimizing $\|z - x\|$ subject to $G(z) \succeq 0$, $z \in \mathbb{R}^n$. If there exist numbers $\tau > 0$ and $\delta > 0$ such that $\Lambda_\Pi(x) \cap \operatorname{cl}(B(0, \tau)) \neq \emptyset$ for every $x \in B(\overline{x}, \delta)$, then $\overline{x}$ satisfies metric subregularity CQ.*

Proof. Let τ and δ be as described in the hypothesis. Following the proof of [24, Thm. 2], note that if $\overline{x} \in \operatorname{int}\mathcal{F}$, then it trivially satisfies metric subregularity CQ, so we will assume that $\overline{x} \in \operatorname{bd}\mathcal{F}$. Let $\delta_0 \in (0, \delta)$ be such that

$$\frac{4}{\delta_0}\mathbb{I}_n - D^2G(z)^*[Y] \succeq 0$$

for all $z \in B(\overline{x}, \delta)$ and all $Y \in \operatorname{cl}(B(0, 2\tau))$. Let $x \in B(\overline{x}, \delta_0/2)$ be such that $x \notin \mathcal{F}$. Although $\Pi_{\mathcal{F}}(x)$ may not be well-defined as a function of x, we will use the notation $\Pi_{\mathcal{F}}(x)$ to denote an arbitrary minimizer

[2] One can consider, for instance, Frobenius norm for $\mathbb{R}^{m \times (m-r)}$ or a column-wise maximum norm.

of $\|z - x\|$ subject to $G(z) \succeq 0$. Then, by definition, we have that $\|\Pi_{\mathcal{F}}(x) - x\| \leq \|\overline{x} - x\| < \delta_0/2$, so $\Pi_{\mathcal{F}}(x) \in B(x, \delta_0/2)$ and, therefore, $\|\Pi_{\mathcal{F}}(x) - \overline{x}\| \leq \|\Pi_{\mathcal{F}}(x) - x\| + \|x - \overline{x}\| < \delta_0$. Let $h\colon \mathbb{R}^n \times \mathbb{S}^m \to \mathbb{R}$ be defined as

$$h(z, Y) \doteq \frac{\langle z - x, z - \Pi_{\mathcal{F}}(x)\rangle}{\|x - \Pi_{\mathcal{F}}(x)\|} - \langle G(z), Y\rangle$$

and note that

$$\nabla_z h(z, Y) = \frac{2z - x - \Pi_{\mathcal{F}}(x)}{\|x - \Pi_{\mathcal{F}}(x)\|} - DG(z)^*[Y]$$

and

$$\nabla_z^2 h(z, Y) = \frac{2}{\|x - \Pi_{\mathcal{F}}(x)\|}\mathbb{I}_n - D^2G(z)^*[Y] \ \succeq\ \frac{4}{\delta_0}\mathbb{I}_n - D^2G(z)^*[Y] \ \succeq\ 0$$

whenever $z \in B(\overline{x}, \delta)$ and $Y \in \operatorname{cl}(B(0, 2\tau))$. Thus, $h(z, Y)$ is convex with respect to its first variable $z \in B(\overline{x}, \delta)$, for every $Y \in \operatorname{cl}(B(0, 2\tau))$. Now let us fix an arbitrary $Y \in \Lambda_\Pi(x) \cap \operatorname{cl}(B(0, \tau))$, which is nonempty by hypothesis. Recall that, by definition of the set $\Lambda_\Pi(x)$, we have that Y is a Lagrange multiplier of the projection problem associated with the point $\Pi_{\mathcal{F}}(x)$. Hence, $2Y$ is a Lagrange multiplier of the problem:

$$\text{Minimize } \tilde{f}_x(z) \doteq \|z - x\| + \frac{\langle z - x, z - \Pi_{\mathcal{F}}(x)\rangle}{\|x - \Pi_{\mathcal{F}}(x)\|}, \quad \text{subject to } G(z) \succeq 0 \tag{21}$$

associated with the point $\Pi_{\mathcal{F}}(x)$, which is a local minimizer of $\tilde{f}_x$ since it is elementary to check that $\tilde{f}_x(\Pi_{\mathcal{F}}(x)) \leq \|z - x\|$ for every $z \in \mathcal{F}$, by the definition of projection, with equality at $\Pi_{\mathcal{F}}(x)$. Writing the KKT conditions for problem (21) at $\Pi_{\mathcal{F}}(x)$ with respect to the Lagrange multiplier $2Y \in \operatorname{cl}(B(0, 2\tau))$, we obtain

$$\frac{2(\Pi_{\mathcal{F}}(x) - x)}{\|x - \Pi_{\mathcal{F}}(x)\|} - DG(\Pi_{\mathcal{F}}(x))^*[2Y] = 0 \tag{22}$$

with $\langle G(\Pi_{\mathcal{F}}(x)), 2Y\rangle = 0$, which yields

$$\begin{aligned} \|x - \Pi_{\mathcal{F}}(x)\| \quad &= -\|x - \Pi_{\mathcal{F}}(x)\| - \langle DG(\Pi_{\mathcal{F}}(x))^*[2Y], x - \Pi_{\mathcal{F}}(x)\rangle \\ &\leq \langle G(\Pi_{\mathcal{F}}(x)) - G(x), 2Y\rangle \\ &= -\langle G(x), 2Y\rangle \end{aligned} \tag{23}$$

after taking inner products of both sides of (22) with $x - \Pi_{\mathcal{F}}(x)$. The middle inequality follows from the definition of adjoint and the convexity of $h(z, Y)$ in the first variable. Taking Moreau's decomposition for $G(x)$, we obtain from (23) that

$$\|x - \Pi_{\mathcal{F}}(x)\| \leq -\langle \Pi_{\mathbb{S}^m_+}(G(x)), 2Y\rangle + \langle \Pi_{\mathbb{S}^m_+}(-G(x)), 2Y\rangle \leq \langle \Pi_{\mathbb{S}^m_+}(-G(x)), 2Y\rangle,$$

because $Y \succeq 0$, which is self-dual, so $\langle \Pi_{\mathbb{S}^m_+}(G(x)), 2Y\rangle \geq 0$; then

$$\operatorname{dist}(x, \mathcal{F}) = \|x - \Pi_{\mathcal{F}}(x)\| \leq \|2Y\|\|\Pi_{\mathbb{S}^m_+}(-G(x))\| \leq 2\tau\|\Pi_{\mathbb{S}^m_+}(-G(x))\|.$$

Since x was chosen arbitrarily, set $\gamma \doteq 2\tau$ and we are done. ■

Now, to compare metric subregularity CQ with seq-CRCQ and seq-CPLD, we first need to show that they are robust, in the sense they are preserved in a neighborhood of the point of interest.

Proposition 5.2. *If seq-CPLD holds at $\overline{x}$, then there exists a neighborhood $\mathcal{V}$ of $\overline{x}$ such that seq-CPLD also holds for every $x \in \mathcal{V} \cap \mathcal{F}$. Moreover, the same property holds for seq-CRCQ.*

Proof. If $\operatorname{Ker} G(\overline{x}) = \{0\}$, the result follows trivially, so let us assume otherwise. Suppose that the statement of the theorem is false, then there exists a sequence $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ such that seq-CPLD does not hold at any x^k. Let us denote by r^k the rank of $G(x^k)$, and note that we can assume, without loss of generality, that $r^k = r$ for all k (by the pigeonhole principle). If seq-CPLD does not hold at x^k, there exists some $E^k \in \mathcal{E}_r(G(x^k))$, some $J^k \subseteq \{1, \ldots, m - r\}$, and some sequence $\{(x^k_\ell, E^k_\ell)\}_{\ell\in\mathbb{N}} \to (x^k, E^k)$ such that $\{v_{ii}(x^k, E^k)\}_{i\in J^k}$ is positively linearly dependent, but $\{v_{ii}(x^k_\ell, E^k_\ell)\}_{i\in J^k}$ is linearly independent for every large $\ell \in \mathbb{N}$. We can also assume that J^k is the same, say J, for every k, by the infinite pigeonhole principle. Then, for each k, let $\ell(k)$ be such that $\lim_{k\to\infty} \ell(k) = \infty$ and that $\{v_{ii}(x^k_\ell, E^k_\ell)\}_{i\in J}$ is linearly independent for every $\ell \geq \ell(k)$. Note that $x^k_{\ell(k)} \to \overline{x}$ and $E^k_{\ell(k)} \to \overline{E}$, and since $\{v_{ii}(x^k, E^k)\}_{i\in J}$ is positively linearly dependent for every large k, the same holds for $\{v_{ii}(\overline{x}, \overline{E})\}_{i\in J}$. This contradicts seq-CPLD. The proof for seq-CRCQ is completely analogous. ■

Now, using Proposition 5.2, it is possible to prove that seq-CPLD (and seq-CRCQ) implies metric subregularity CQ. We shall do this in the same style as Andreani et al. [8]:

Theorem 5.1. *If seq-CPLD holds at $\overline{x}$ and G is twice differentiable around $\overline{x}$, then $\overline{x}$ satisfies metric subregularity CQ.*

Proof. Suppose that metric subregularity CQ does not hold at $\overline{x}$. In view of Proposition 5.1, there exist sequences $\{\tau^k\}_{k\in\mathbb{N}} \to \infty$ and $\{y^k\}_{k\in\mathbb{N}} \to \overline{x}$ such that $\Lambda(y^k) \cap \text{cl}(B(0, \tau^k)) = \emptyset$ for every $k \in \mathbb{N}$.

Now let $\{z^k\}_{k\in\mathbb{N}}$ be such that $z^k = \Pi_{\mathcal{F}}(y^k)$ for each k and note that $z^k \to \overline{x}$. Indeed, even if $\Pi_{\mathcal{F}}(y^k)$ is not uniquely defined, we have that $\text{dist}(y^k, \mathcal{F}) = \|z^k - y^k\| \to 0$ because $y^k \to \overline{x} \in \mathcal{F}$ and, thus, $\|z^k - \overline{x}\| \leqslant \|z^k - y^k\| + \|y^k - \overline{x}\| \to 0$. By the previous proposition, z^k satisfies seq-CPLD for all k large enough. Consequently, there exists a sequence $\{Y^k\}_{k\in\mathbb{N}} \subseteq \mathbb{S}^m_+$ such that

$$\frac{z^k - y^k}{\|z^k - y^k\|} - DG(z^k)^*[Y^k] = 0$$

and $\langle G(z^k), Y^k\rangle = 0$ for every k, which implies that $\lambda_i(Y^k) = 0$ for every $i \in \{m - r + 1, \dots, m\}$ and every $k \in \mathbb{N}$. Let U^k be an arbitrary matrix that diagonalizes Y^k and let E^k be the part of it that corresponds to the $m - r$ smallest eigenvalues of $G(z^k)$. So

$$\frac{z^k - y^k}{\|z^k - y^k\|} - \sum_{i=1}^{m-r} \lambda_i(Y^k) v_{ii}(z^k, E^k) = 0. \tag{24}$$

Again, by Caratheodory' lemma (cf. Lemma 2.1) and the infinite pigeonhole principle, we obtain a set $J \subseteq \{1, \dots, m - r\}$ such that $\{v_{ii}(z^k, E^k) \colon i \in J\}$ is linearly independent and $\sum_{i=1}^{m-r} \lambda_i(Y^k) v_{ii}(z^k, E^k) = \sum_{i\in J} \alpha_i^k v_{ii}(z^k, E^k)$ for every k where $\alpha_i^k \lambda_i(Y^k) > 0$ for all $i \in J$. Then, recall from the definition that $Y^k \in \Lambda(y^k)$, so $\|Y^k\| > \tau^k \to \infty$. Let $m^k \doteq \max\{\alpha_i^k \colon i \in J\}$ and divide (24) by m^k to obtain that $\{v_{ii}(\overline{x}, \overline{E}) \colon i \in J\}$ is linearly dependent for every limit point $\overline{E}$ of $\{E^k\}_{k\in\mathbb{N}}$, which contradicts seq-CPLD at $\overline{x}$. ■

6 Conclusion

There are few constraint qualifications available for NSDP, and as far as we know, the use of CQs in the global convergence of algorithms is somewhat limited to nondegeneracy and Robinson's CQ. In contrast, several constraint qualifications have been defined for NLP over the past decades, mostly improving the global convergence of algorithms beyond the case when the set of Lagrange multipliers is bounded. We are in a path to extend these CQs to conic contexts, such as NSDP, that started in [7]. In fact, the results of this paper can be considered a significant improvement of [7] based on our previous developments in [6]. We introduced two weak constant rank CQs for NSDP, called weak-CRCQ and weak-CPLD, which are essentially "diagonal extensions" of their NLP counterparts, in the sense of Proposition 3.1. Namely, one can embed an NLP problem using a structurally diagonal semidefinite constraint and both conditions are preserved. This is a fairly unusual property as this approach usually induces a degenerate NSDP problem; we however believe that this, in some sense, provides a sound mathematical consistency to our approach. These conditions were used to prove convergence of an external penalty method to stationary points, but any application beyond that, besides the mere existence of Lagrange multipliers, is still a subject for investigation. However, they were the starting points for introducing stronger constant rank CQs, called seq-CRCQ and seq-CPLD, with more interesting properties, such as the convergence theory of a larger class of algorithms such as augmented Lagrangians, sequential quadratic programming, and interior point methods, and a property related with the ability to compute error bounds under these conditions. We believe that several other applications of constant rank CQs will appear in the literature, such as the computation of the derivative of the value function of a parameterized NSDP problem and the computation of second-order necessary optimality conditions. In NLP, constant rank CQs are used to define a strong second-order necessary optimality condition that depends on a single Lagrange multiplier, rather than on the full set of Lagrange multipliers, which we believe will be the case for conic problems as well. It is also the case that constant rank conditions provide the adequate assumptions for guaranteeing global convergence of algorithms to second-order stationary points, which has not been considered yet in the conic programming literature.

This paper leaves several interesting open questions that can be addressed in future works, such as the use of weak-CRCQ and weak-CPLD in algorithms other than external penalty methods, and the analysis of some stability properties under the conditions introduced in this manuscript. It is also worth recalling that although our conditions were defined by means of sequences, which seems appropriate when talking about convergence of algorithms, we also provided characterizations of them without sequences, in a more classical way, which should foster new applications.

The relationship among the CQs we presented in this paper, and existing ones, is summarized in the following diagram, where (solid) arrows represent (strict) implications, existing CQs are in blue boxes, and new CQs are in green boxes.

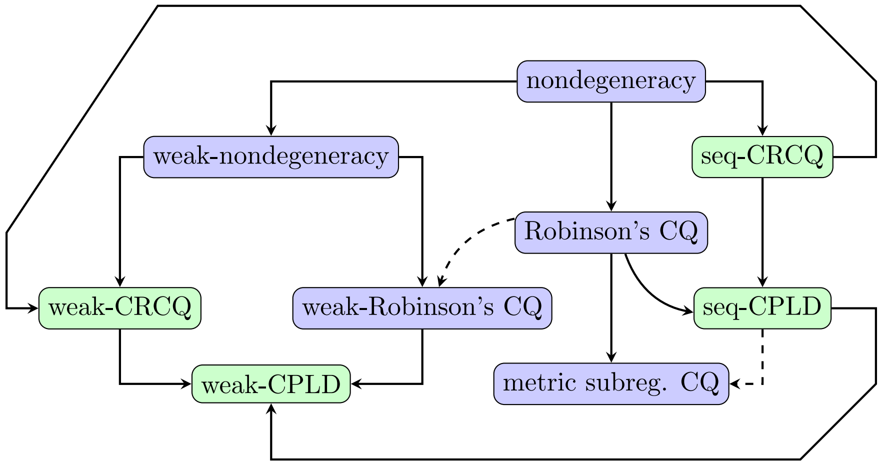


Figure 1: Relationship among the new constraint qualifications and some of the existing ones.

References

[1] R. Andreani, C. E. Echagüe, and M. L. Schuverdt, *Constant-rank condition and second-order constraint qualification*, Journal of Optimization Theory and Applications, 146 (2010), pp. 255–266, https://doi.org/10.1007/s10957-010-9671-8.

[2] R. Andreani, E. H. Fukuda, G. Haeser, H. Ramírez C., D. O. Santos, P. J. S. Silva, and T. P. Silveira, *Erratum to: New constraint qualifications and optimality conditions for second order cone programs*, To appear in Set-Valued and Variational Analysis, (2021), https://doi.org/10.1007/s11228-021-00573-5.

[3] R. Andreani, E. H. Fukuda, G. Haeser, D. O. Santos, and L. D. Secchin, *On the use of Jordan algebras for improving global convergence of an augmented Lagrangian method in nonlinear semidefinite programming*, To appear in Computational Optimization and Applications, (2021), https://doi.org/10.1007/s10589-021-00281-8.

[4] R. Andreani, W. Gómez, G. Haeser, L. M. Mito, and A. Ramos, *On optimality conditions for nonlinear conic programming*, To appear in Mathematics of Operations Research, 2022, DOI: 10.1287/moor.2021.1203.

[5] R. Andreani, G. Haeser, and J. M. Martínez, *On sequential optimality conditions for smooth constrained optimization*, Optimization, 60 (2011), pp. 627–641, http://dx.doi.org/10.1080/02331930903578700.

[6] R. Andreani, G. Haeser, L. M. Mito, and H. Ramírez, *Weak notions of nondegeneracy in nonlinear semidefinite programming*, tech. report, 2020, https://arxiv.org/abs/2012.14810v1 (accessed 2021/02/16).

[7] R. Andreani, G. Haeser, L. M. Mito, H. Ramírez, D. O. Santos, and T. P. Silveira, *Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming*, To appear in Optimization Letters, (2021), https://doi.org/10.1007/s11590-021-01737-w (accessed 2021/02/16).

[8] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva, *A relaxed constant positive linear dependence constraint qualification and applications*, Mathematical Programming, Series A, 135 (2012), pp. 255–273, https://doi.org/10.1007/s10107-011-0456-0.

[9] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva, *Two new weak constraint qualifications and applications*, SIAM Journal on Optimization, 22 (2012), pp. 1109–1135, http://dx.doi.org/10.1137/110843939.

[10] R. Andreani, G. Haeser, and D. S. Viana, *Optimality conditions and global convergence for nonlinear semidefinite programming*, Mathematical Programming, Series A, 180 (2020), pp. 203–235, http://dx.doi.org/10.1007/s10107-018-1354-5.

[11] R. Andreani, J. Martínez, A. Ramos, and P. J. S. Silva, *Strict constraint qualifications and sequential optimality conditions for constrained optimization*, Mathematics of Operations Research, 43 (2018), pp. 693–717, https://doi.org/10.1287/moor.2017.0879.

[12] R. Andreani, J. M. Martínez, and M. L. Schuverdt, *On the relation between constant positive linear dependence condition and quasinormality constraint qualification*, Journal of Optimization Theory and Applications, 125 (2005), pp. 473–485, https://doi.org/10.1007/s10957-004-1861-9.

[13] D. P. Bertsekas, *Nonlinear Programming*, Athenas Scientific. Belmont, Mass, 1999.

[14] E. Birgin and J. M. Martínez, *Practical Augmented Lagrangian Methods for Constrained Optimization*, SIAM Publications. Philadelphia, 2014.

[15] J. F. Bonnans and A. Shapiro, *Pertubation Analysis of Optimization Problems*, Springer-Verlag. Berlin, 2000.

[16] E. Börgens, C. Kanzow, P. Mehlitz, and G. Wachsmuth, *New constraint qualifications for optimization problems in Banach spaces based on asymptotic KKT conditions*, SIAM Journal on Optimization, 30 (2020), pp. 2956–2982, https://doi.org/10.1137/19M1306804.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[18] R. Correa and H. Ramírez C., *A global algorithm for nonlinear semidefinite programming*, SIAM Journal on Optimization, 15 (2004), pp. 303–318, https://doi.org/10.1137/S1052623402417298.

[19] H. Gfrerer, *First order and second order characterizations of metric subregularity and calmness of constraint set mappings*, SIAM Journal on Optimization, 21 (2011), pp. 1439–1474, https://doi.org/10.1137/100813415.

[20] A. D. Ioffe, *Metric regularity - A survey. Part I. Theory*, Journal of the Australian Mathematical Society, 101 (2016), pp. 188–243, https://doi.org/10.1017/S1446788715000701.

[21] A. D. Ioffe, *Metric regularity - A survey. Part II. Applications*, Journal of the Australian Mathematical Society, 101 (2016), pp. 376–417, https://doi.org/10.1017/S1446788715000695.

[22] R. Janin, *Directional derivative of the marginal function in nonlinear programming*, Mathematical Programming Studies, 21 (1984), pp. 127–138, https://doi.org/10.1007/BFb0121214.

[23] P. Mehlitz, *Asymptotic stationarity and regularity for nonsmooth optimization problems*, Journal of Nonsmooth Analysis and Optimization (2020), https://doi.org/10.46298/jnsao-2020-6575.

[24] L. Minchenko and S. Stakhovski, *On relaxed constant rank regularity condition in mathematical programming*, Optimization, 60 (2011), pp. 429–440, https://doi.org/10.1080/02331930902971377.

[25] L. Minchenko and S. Stakhovski, *Parametric nonlinear programming problems under the relaxed constant rank condition*, SIAM Journal on Optimization, 21 (2011), pp. 314–332, https://doi.org/10.1137/090761318.

[26] J. J. Moreau, *Décomposition orthogonale d'un espace hilbertien selon deux cones mutuellement polaires*, Comptes Rendus de l'Academie des Sciences de Paris, 255 (1962), pp. 238–240.

[27] L. Qi and Z. Wei, *On the constant positive linear dependence conditions and its application to SQP methods*, SIAM Journal on Optimization, 10 (2000), pp. 963–981, https://doi.org/10.1137/S1052623497326629.

[28] S. M. Robinson, *First-order conditions for general nonlinear optimization*, SIAM Journal on Applied Mathematics, 30 (1976), pp. 597–610, https://doi.org/10.1137/0130053.

[29] R. T. Rockafellar and R. Wets, *Variational Analysis*, Grundlehren der mathematischen Wissenschaften, v. 317. Springer-Verlag Berlin Heidelberg. Berlin, 2009.

[30] A. Shapiro, *First and second order analysis of nonlinear semidefinite programs*, Mathematical Programming, Series B, 77 (1997), pp. 301–320, https://doi.org/10.1007/BF02614439.

[31] A. Shapiro and M. K. H. Fan, *On eigenvalue optimization*, SIAM Journal on Optimization, 5 (1995), pp. 552–569, https://doi.org/10.1137/0805028.

[32] H. Yamashita, H. Yabe, and K. Harada, *A primal-dual interior point method for nonlinear semidefinite programming*, Mathematical Programming, Series A, 135 (2012), pp. 89–121, https://doi.org/10.1007/s10107-011-0449-z.

[33] Y. Zhang and L. Zhang, *New constraint qualifications and optimality conditions for second order cone programs*, Set-Valued and Variational Analysis, 27 (2019), pp. 693–712, https://doi.org/10.1007/s11228-018-0487-2.

Appendix H

External reference VIII

Title: Global convergence of algorithms under constant rank conditions for nonlinear second-order cone programming.

Status: Under review.

What is attached: A preprint that has *not* been peer reviewed yet.

Global Convergence of Algorithms Under Constant Rank Conditions for Nonlinear Second-Order Cone Programming

Roberto Andreani * Gabriel Haeser † Leonardo M. Mito † Héctor Ramírez C. ‡ Thiago P. Silveira †

October 26, 2021

Abstract

In [R. Andreani, G. Haeser, L. M. Mito, H. Ramírez C., Weak notions of nondegeneracy in nonlinear semidefinite programming, arXiv:2012.14810, 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 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.

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 wheren known under the *Mangasarian-Fromovitz constraint qualification* (MFCQ) [24, 44] and the *linear independence constraint qualification* (LICQ) [24].

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 [1, 25, 27], strong second-order necessary optimality conditions [5], global convergence of algorithms [4], 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 are discussed respectively in [34] and [14].

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 and 404656/2018-8), PRONEX - CNPq/FAPERJ (grant E-26/010.001247/2016), and ANID (FONDECYT grant 1201982, Program ANID ACE210010 and Basal Program CMM ANID AFB210005).

*Department of Applied Mathematics, University of Campinas, Campinas, SP, Brazil. Email: andreani@unicamp.br

†Department of Applied Mathematics, University of São Paulo, São Paulo, SP, Brazil. Emails: {ghaeser,leokoto,thiagops}@ime.usp.br

‡Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático (CNRS IRL2807), Universidad de Chile, Santiago, Chile. Email: hramirez@dim.uchile.cl

Although constraint qualifications with applications towards 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 [3], hydroacoustic classification of fishes [20], and others [32]. 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 counter-example 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 non-negative 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 [5]. However, no application towards 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 modelled 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 Section 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 Section 5 we introduce perturbed versions of the constant rank conditions of Section 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 Section 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 j-th 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 $\operatorname{cl}(B(y, \delta))$.

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 non-empty 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

$$\mathcal{P}_C(y) = \big(\mathcal{P}_{C_1}(y_1), \ldots, \mathcal{P}_{C_q}(y_q)\big).$$

To relate our results with 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 non-zero $v \in C$ 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*

$$C = \bigcup_{w\in S} \operatorname{cone}(\mathcal{E}_w), \tag{1}$$

where $\operatorname{cone}(\mathcal{E}_w)$ 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 $\tilde{v} \in \operatorname{cone}(\mathcal{E}_w)$; 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 $\mathbb{R}^2$:

$$\mathbb{R}^2 = \{(r\cos(w), r\sin(w))\colon w \in [0, 2\pi], r \geq 0\} = \bigcup_{w\in[0,2\pi]} \operatorname{cone}((\cos(w), \sin(w))) \tag{3}$$

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

$$\mathbb{L}_m = \bigcup_{\substack{w\in\mathbb{R}^{m-1}\\ \|w\|=1}} \operatorname{cone}(\{(1, -w), (1, w)\}),$$

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 $\text{span}(\{(1,-w),(1,w)\})$ 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 2 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 $y_1,\dots,y_p \in \mathbb{R}^n$, and let $\alpha_1,\dots,\alpha_p \in \mathbb{R}$ be arbitrary. Then, there exists some $J \subseteq \{1,\dots,p\}$ 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:

$$\begin{aligned} &\underset{x\in\mathbb{R}^n}{\text{Minimize}} && f(x), \\ &\text{subject to} && g_j(x) \in \mathbb{L}_{m_j},\ \forall j \in \{1,\dots,q\}, \end{aligned} \tag{NSOCP}$$

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 $j \in \{1,\dots,q\}$, 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 $\text{int}\mathbb{L}_{m_j}$ and $\text{bd}_+\mathbb{L}_{m_j}$, respectively; and as usual in the study of NSOCP, for any $x \in \mathcal{F}$ we partition $\{1,\dots,q\}$ as follows:

$$\begin{aligned} I_0(x) &:= \{j \in \{1,\dots,q\}\colon g_j(x) = 0\}, \\ I_B(x) &:= \{j \in \{1,\dots,q\}\colon g_j(x) \in \text{bd}_+\mathbb{L}_{m_j}\}, \\ I_{\text{int}}(x) &:= \{j \in \{1,\dots,q\}\colon g_j(x) \in \text{int}\mathbb{L}_{m_j}\}. \end{aligned} \tag{6}$$

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

$$\lambda_i(y) := y_0 + (-1)^i\|\widehat{y}\| \quad \text{and} \quad u_i(y) := \begin{cases} \dfrac{1}{2}\left(1, (-1)^i \dfrac{\widehat{y}}{\|\widehat{y}\|}\right), & \text{if } \widehat{y} \neq 0, \\ \dfrac{1}{2}\left(1, (-1)^i w\right), & \text{otherwise,} \end{cases} \tag{7}$$

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

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

where $[\,\cdot\,]_+ := \max\{\,\cdot\,, 0\}$.

Remark 2.2. *From this point onwards, we will assume that $m_j > 1$ for every $j \in \{1,\dots,q\}$. 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

$$\begin{aligned} \nabla_x L(\overline{x}, \overline{\mu}_1, \ldots, \overline{\mu}_q) = 0, \\ \langle \overline{\mu}_j, g_j(\overline{x}) \rangle = 0, \quad \forall j \in \{1, \ldots, q\}, \end{aligned} \tag{8}$$

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*

$$\left\{ Dg_j(\overline{x})^\top \Gamma_j g_j(\overline{x}) \right\}_{j \in I_B(\overline{x})} \bigcup \left\{ Dg_j(\overline{x})^\top \right\}_{j \in I_0(\overline{x})} \tag{9}$$

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

where

$$\Gamma_j := \begin{bmatrix} 1 & 0 \\ 0 & -\mathbb{I}_{m_j - 1} \end{bmatrix} \tag{10}$$

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 $m_1 = \ldots = m_q = 1$, 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 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 [33] – meaning 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

$$\operatorname{Arw}(g_0(x), \widehat{g}(x)) \in \mathbb{S}_+^m := \{ M \in \mathbb{R}^{m \times m} \colon M = M^\top;\ \forall d \in \mathbb{R}^m, d^\top M d \geq 0 \}$$

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.

[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].

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 i-th 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*

$$\left\{Dg_j(\overline{x})^\top u_1(g_j(\overline{x}))\right\}_{j\in I_B(\overline{x})} \bigcup \left\{Dg_j(\overline{x})^\top(1, -\overline{w}_j),\ Dg_j(\overline{x})^\top(1, \overline{w}_j)\right\}_{j\in I_0(\overline{x})} \tag{11}$$

   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 (11) 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 non-zero, 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 parametrizations, 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 analogue 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

$$\tilde{g}(x) := (g_0(x), g_1(x), 0, \ldots, 0) \in \mathbb{L}_m$$

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*

$$Dg(\overline{x})^\top(1, -\overline{w}) = \nabla g_0(\overline{x}) - \overline{w}_1 \nabla g_1(\overline{x}) \quad \text{and} \quad Dg(\overline{x})^\top(1, \overline{w}) = \nabla g_0(\overline{x}) + \overline{w}_1 \nabla g_1(\overline{x})$$

for every $\overline{w} = (\overline{w}_1, \ldots, \overline{w}_{m-1})$ *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*

$$u_1(\tilde{g}(x)) = \frac{1}{2}\left(1, -\frac{g_1(x)}{|g_1(x)|}, 0, \ldots, 0\right) \quad \text{and} \quad u_2(\tilde{g}(x)) = \frac{1}{2}\left(1, \frac{g_1(x)}{|g_1(x)|}, 0, \ldots, 0\right).$$

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 [10], 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*

$$Dg(\overline{x})^\top(1, -\overline{w}) = \nabla g_0(\overline{x}) \mp \nabla g_1(\overline{x}) \quad \text{and} \quad Dg(\overline{x})^\top(1, \overline{w}) = \nabla g_0(\overline{x}) \pm \nabla g_1(\overline{x})$$

are linearly independent.

Example 3.1 suggests that demanding linear independence of (11) 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* $\{u_2(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1, \overline{w}_j)$*, for every* $j \in I_0(\overline{x})$*, such that* (11) *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* (11) *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. *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 (9) 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.

Proposition 3.1. *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 (9) 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 (11) 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 non-zero 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 (11) 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. *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*

$$\{\nabla g_{j,0}(\overline{x})\}_{j \in I_0(\overline{x})} \bigcup \left\{ Dg_j(\overline{x})^\top u_1(g_j(\overline{x})) \right\}_{j \in I_B(\overline{x})} \tag{12}$$

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 (11) is positively linearly independent; and, by contradiction, suppose that (12) 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

$$\sum_{j \in I_B(\overline{x}) \cup I_0(\overline{x})} \eta_j \nabla g_{j,0}(\overline{x}) - \sum_{j \in I_B(\overline{x})} \eta_j D\widehat{g}_j(\overline{x})^\top \frac{\widehat{g}_j(\overline{x})}{\|\widehat{g}_j(\overline{x})\|} = 0. \tag{13}$$

Now set

$$\alpha_j = \beta_j = \frac{\eta_j}{2}$$

for every $j \in I_0(\overline{x})$ and (13) can be rewritten as

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

which implies (11) 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 $g_j(x) \in \mathbb{L}_{m_j}$, $j \in \{1, \ldots, q\}$, 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 $d_j \in \mathbb{R}^n$, $j \in \{1, \ldots, q\}$, 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 $d_1 = d_2 = \ldots = d_q$ 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 $j \in \{1, \ldots, q\}$ 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 $\ell \in \{1, \ldots, q\}$ 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 non-zero, 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

$$\nabla g_{\ell,0}(\overline{x}) = -D\widehat{g}_\ell(\overline{x})^\top \tilde{w}. \tag{14}$$

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 exists 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, $u_1(g_\ell(x^k))$ 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

$$\lim_{k \to \infty} u_1(g_\ell(x^k)) = \frac{1}{2}\left(1, -\frac{\tilde{w} - v}{\|\tilde{w} - v\|}\right) \quad \text{and} \quad \lim_{k \to \infty} u_2(g_\ell(x^k)) = \frac{1}{2}\left(1, \frac{\tilde{w} - v}{\|\tilde{w} - v\|}\right)$$

and, by weak-Robinson's CQ, the vectors $Dg_\ell(\overline{x})^\top\left(1,-\frac{\tilde{w}-v}{\|\tilde{w}-v\|}\right)$ and $Dg_\ell(\overline{x})^\top\left(1,\frac{\tilde{w}-v}{\|\tilde{w}-v\|}\right)$ are positively linearly independent. However, the following system in the variables a and b:

$$\begin{aligned} 0 &= aDg_\ell(\overline{x})^\top\left(1,\frac{\tilde{w}-v}{\|\tilde{w}-v\|}\right)+bDg_\ell(\overline{x})^\top\left(1,-\frac{\tilde{w}-v}{\|\tilde{w}-v\|}\right) \\ &= a\nabla g_{\ell,0}(\overline{x})+\frac{a}{\|\tilde{w}-v\|}D\widehat{g}_\ell(\overline{x})^\top\tilde{w}+b\nabla g_{\ell,0}(\overline{x})-\frac{b}{\|\tilde{w}-v\|}D\widehat{g}_\ell(\overline{x})^\top\tilde{w} \\ &= \left[a\left(\frac{1}{\|\tilde{w}-v\|}-1\right)-b\left(\frac{1}{\|\tilde{w}-v\|}+1\right)\right]D\widehat{g}_\ell(\overline{x})^\top\tilde{w} \end{aligned}$$

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 $D\widehat{g}_\ell(\overline{x})^\top v=0$; and in the last equality, we used (14). ■

Remark 3.1. *The same strategy of the previous proof actually allows proving a slightly stronger result: if a feasible point $\overline{x}$ satisfies weak-Robinson's CQ, then for each index $j\in I_0(\overline{x})$ the constraint*

$$g_\ell(x)\in\mathbb{L}_{m_\ell},\ \ \forall\ell\in I_B(\overline{x})\cup\{j\}$$

satisfies Robinson's CQ at $\overline{x}$. In particular, if $I_0(\overline{x})$ 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 $I_0(\overline{x})$ is a singleton has been previously considered, for instance, in [36, 40]. In the general case we were not able to prove nor to 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 $m_1=\ldots=m_q=1$, whose constraints take the form $g_1(x)\geq 0,\ldots,g_q(x)\geq 0$, and let $\overline{x}\in\mathcal{F}$. We recall that the nondegeneracy condition in this case is equivalent to LICQ, which holds when the family of vectors

$$\{\nabla g_j(\overline{x})\}_{j\in I_0(\overline{x})} \tag{15}$$

has full rank. The constant rank constraint qualification (CRCQ) condition can be considered a relaxation of LICQ, since it allows the rank of (15) to be incomplete, as long as the rank of the family

$$\{\nabla g_j(x)\}_{j\in J_0} \tag{16}$$

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 (16) 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 (15), 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 (16) 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 [9] 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 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 (11) 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

$$\mathcal{D}_{J_B,J_-,J_+}(x,w):=\left\{Dg_j(x)^\top u_1(g_j(x))\right\}_{j\in J_B}\bigcup\left\{Dg_j(x)^\top(1,-w_j)\right\}_{j\in J_-}\bigcup\left\{Dg_j(x)^\top(1,w_j)\right\}_{j\in J_+} \tag{17}$$

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 (11).

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*
$$\{u_1(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1,-\overline{w}_j) \quad \text{and} \quad \{u_2(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1,\overline{w}_j),$$
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* $w^k = [w_j^k]_{j\in J_-\cup J_+}$ *satisfies*
$$u_1(g_j(x^k)) = \frac{1}{2}(1,-w_j^k) \quad \text{and} \quad u_2(g_j(x^k)) = \frac{1}{2}(1,w_j^k) \tag{18}$$
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*
$$\{u_1(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1,-\overline{w}_j) \quad \text{and} \quad \{u_2(g_j(x^k))\}_{k\in I} \to \frac{1}{2}(1,\overline{w}_j),$$
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 $j \in \{1,\dots,q\}$ – 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 below 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:*

- $\mathcal{N}_0 := \{k \in \mathbb{N} \colon \widehat{g}(x^k) = 0\}$*. For* $k \in \mathcal{N}_0$*, we can choose*
$$u_1(g(x^k)) = \frac{1}{2}\left(1,-w^k\right) \qquad \text{and} \qquad u_2(g(x^k)) = \frac{1}{2}\left(1,w^k\right),$$
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* $k \in I$*.*
- $\mathcal{N}_1 := \{k \in \mathbb{N} \colon \widehat{g}(x^k) \neq 0\}$*. 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* $g(\overline{x}) \in \text{bd}_+\mathbb{L}_m$*, 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 non-zero 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 below.*

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*
$$\begin{aligned} &\underset{x\in\mathbb{R}}{\text{Minimize}} && -x, \\ &\text{subject to} && g(x) := (x, x+x^2) \in \mathbb{L}_2. \end{aligned} \tag{19}$$
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 are 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. *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 analyse 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,*

$$Dg(x^k)^\top(1, (-1)^i w^k) = Dg(\overline{x})^\top(1, -\overline{w}) = \left(-1 + (-1)^i \frac{2}{\sqrt{2}}\right) < 0$$

for each $i \in \{1, 2\}$. *Then, the family (11) will have the same sign, 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. *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* $x^k \neq 0$ *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*

$$f(x) + \frac{1}{2}\|x - \overline{x}\|_2^2 + \frac{\rho_k}{2}\left(\sum_{j=1}^{q} \|\mathcal{P}_{\mathbb{L}_{m_j}}(-g_j(x))\|^2\right). \tag{20}$$

Proof. The proof of this theorem is contained in the proof of [7, Theorem 3.1]. ■

Observe that the gradient of (20) 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. *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

$$\nabla f(x^k) - \sum_{j=1}^{q} Dg_j(x^k)^\top \mu_j^k \to 0. \tag{21}$$

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 $\beta_j^k = [-\rho_k \lambda_2(g_j(x^k))]_+ \geq 0$. Define

$$\Psi^k := \sum_{j\in I_B(\overline{x})\cup I_0(\overline{x})} \alpha_j^k Dg_j(x^k)^\top u_1(g_j(x^k)) + \sum_{j\in I_0(\overline{x})} \beta_j^k Dg_j(x^k)^\top u_2(g_j(x^k)) \tag{22}$$

and note that (21) 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

$$\left\{Dg_j(x^k)^\top u_1(g_j(x^k))\right\}_{j\in J_B^k\cup J_-^k} \bigcup \left\{Dg_j(x^k)^\top u_2(g_j(x^k))\right\}_{j\in J_+^k} \tag{23}$$

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

$$m^k := \max\{\max\{\tilde{\alpha}_j^k : j \in J_B \cup J_-\},\ \max\{\tilde{\beta}_j^k : j \in J_+\}\},$$

diverges. Dividing (21) 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 non-zero, 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 (23) with w^k defined as in (18), 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 so-called "naive extensions" of CRCQ (and CPLD) to NSOCP, which were obtained by replacing the conic constraints of (NSOCP) that satisfy $g_j(\overline{x}) \in \text{bd}_+\mathbb{L}_{m_j}$ 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 $\tilde{\Phi}_j$ characterize 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 (9) 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 $\{Dg_j(x)^\top u_1(g_j(x))\}_{j\in J_B}$ 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, then 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 towards 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 analogue 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 $\{u_1(g_j(x^k) + \Delta_j^k)\}_{k\in I} \to (1, -\overline{w}_j)$ and $\{u_2(g_j(x^k) + \Delta_j^k)\}_{k\in I} \to (1, \overline{w}_j)$, 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 every $k \in I$ 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*

$$u_1(g_j(x^k) + \Delta_j^k) = \frac{1}{2}(1, -w_j^k) \quad \text{and} \quad u_2(g_j(x^k) + \Delta_j^k) = \frac{1}{2}(1, w_j^k) \tag{24}$$

for each $j \in J_- \cup J_+$. Recall that $\mathcal{D}_{J_B,J_-,J_+}(x, w)$ was defined in (17).

- Sequential CPLD condition for NSOCP *(seq-CPLD) 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 $\{u_1(g_j(x^k) + \Delta_j^k)\}_{k\in I} \to (1, -\overline{w}_j)$ and $\{u_2(g_j(x^k) + \Delta_j^k)\}_{k\in I} \to (1, \overline{w}_j)$, 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 $\mathcal{D}_{J_B,J_-,J_+}(x^k, w^k)$ remains linearly dependent for all $k \in I$ large enough, where $\overline{w}$ 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. *Consider the constraint*

$$g(x) := (-x, x) \in \mathbb{L}_2$$

at the point $\overline{x} := 0$, which is the only feasible point of the problem. In order to verify that $\overline{x}$ satisfies seq-CPLD and seq-CRCQ, let $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ 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

$$u_1(g(x^k) + \Delta^k) = \frac{1}{2}(1, -1) \qquad \text{and} \qquad u_2(g(x^k) + \Delta^k) = \frac{1}{2}(1, 1),$$

which implies that $w^k = \overline{w} = 1$ for all $k \in I$. Hence, the vectors $Dg(\overline{x})^\top(1, -\overline{w}) = -2$ 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. *Consider the constraint*

$$g(x) := (x^2, x, 0) \in \mathbb{L}_3$$

at $\overline{x} := 0$*. Let us begin by showing that* $\overline{x}$ *satisfies both weak-CRCQ and weak-CPLD, so let* $\{x^k\}_{k\in\mathbb{N}} \to \overline{x}$ *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 (17).*

Proof. Suppose that there exists 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 $w^k := [w_j^k]_{j\in J_-\cup J_+}$ and $\|w_j^k\| = 1$, 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

$$\Delta_j^k := \begin{cases} \dfrac{1}{k}\left(1, w_j^k\right) - g_j(x^k), & \text{if } j \in J_- \cup J_+ \\ g_{j,0}(\overline{x})\left(1, \dfrac{\widehat{g}_j(x^k)}{\|\widehat{g}_j(x^k)\|}\right) - g_j(x^k), & \text{if } j \in J_B, \end{cases} \tag{25}$$

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 $\mathcal{D}_{J_B,J_-,J_+}(x^k, w^k)$ is 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 $\{\mu_j^k\}_{k\in\mathbb{N}} \subseteq \mathbb{L}_{m_j}$, $j \in \{1,\dots,q\}$, in the sense that

$$\nabla_x L(x^k, \mu_1^k, \dots, \mu_q^k) \to 0 \tag{26}$$

and for every $k \in \mathbb{N}$,

$$g_j(x^k) + \Delta_j^k \in \mathbb{L}_{m_j} \quad \text{and} \quad \langle g_j(x^k) + \Delta_j^k, \mu_j^k \rangle = 0 \tag{27}$$

for some sequences $\Delta_j^k \to 0$, $j \in \{1,\dots,q\}$. 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. *Let $\{x^k\}_{k\in\mathbb{N}}$ and $\{\mu_j^k\}_{k\in\mathbb{N}} \subseteq \mathbb{L}_{m_j}$, $j \in \{1,\dots,q\}$ satisfy (26) and (27), 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 (26) we obtain that

$$\nabla f(x^k) - \sum_{j=1}^{q} Dg_j(x^k)^\top \mu_j^k \to 0. \tag{28}$$

Now, by (27) 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 (10), 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 (27) 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 $j \in \{1,\dots,q\}$. 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 (28) 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 (26) and (27) 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 [28, 41, 43] 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 (26) and (27). For completeness, we state their algorithm below:

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 $\alpha^k \geq \max\{|\mu_{j,0}^k| \colon j \in \{1,\dots,q\}\}$, then $\alpha^{k+1} := \alpha^k$; otherwise, $\alpha^{k+1} := \max\{\alpha^k, |\mu_{j,0}^k| \colon j \in \{1,\dots,q\}\} + \tau$;

Step 3: Compute some scalar $t^k \in (0,1]$ satisfying

$$\Phi_{\alpha^{k+1}}(x^k) - \Phi_{\alpha^{k+1}}(x^k + t^k d^k) \leq \sigma t^k (d^k)^\top M^k d^k; \tag{29}$$

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.

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 towards convergence. Instead of A3, an assumption over the problem (and not the method), for instance the fulfilment 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.

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* (26) *and* (27).

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:

$$\begin{aligned} \nabla f(x^k) + M^k d^k - \sum_{j=1}^{q} Dg_j(x^k)^\top \mu_j^k = 0, \\ \langle \mu_j^k, g_j(x^k) + Dg_j(x^k)d^k \rangle = 0, \quad \forall j \in \{1,\dots,q\}, \\ g_j(x^k) + Dg_j(x^k)d^k \in \mathbb{L}_{m_j}, \quad \forall j \in \{1,\dots,q\}. \end{aligned}$$

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. *Let $\overline{x}$ be a feasible point of (NSOCP) and let $g(x) := (g_1(x), \dots, g_q(x))$. We say that $\overline{x}$ satisfies the* metric subregularity CQ *when there exists some $\gamma > 0$ and a neighborhood $\mathcal{V}$ of $\overline{x}$ such that*

$$\mathrm{dist}(x, \mathcal{F}) \leq \gamma \mathrm{dist}(g(x), \Pi_{j=1}^q \mathbb{L}_{m_j})$$

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 metric subregularity CQ. 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 $j \in \{1, \dots, q\}$. 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 $g_j(z) \in \mathbb{L}_{m_j}$, $j \in \{1, \dots, q\}$, $z \in \mathbb{R}^n$. If there exist numbers $\tau > 0$ and $\delta > 0$ such that $\Lambda_x(y) \cap \mathrm{cl}(B(0, \tau)) \neq \emptyset$ for every $x \in B(\overline{x}, \delta)$, then $\overline{x}$ satisfies metric subregularity CQ.*

Then, we shall prove that seq-CPLD and seq-CRCQ are robust, and this, together with Proposition 5.3, is enough to show that that they imply metric subregularity CQ.

Theorem 5.2. *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. *metric subregularity CQ 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 (25), 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 $g_j(x_{\ell(k)}^k) + \Delta_j^k \in \mathrm{bd}_+\mathbb{L}_{m_j}$ 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. ∎

6 Conclusion

In our previous work [10], we studied two ways of incorporating some structural features of the semidefinite cone into the nondegeneracy condition of Shapiro and Fan [45]; 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 non-zero 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 [23]; the study of the boundedness of Lagrange multipliers estimates and the use of scaled stopping criteria [13]; stability analysis of parametric optimization problems [18, 26, 35, 37, 38, 39, 40]; and necessary optimality conditions for some extended classes of bilevel optimization problems with conic constraints [21, 22, 46].

References

[1] L. Adam, R. Henrion, and J. Outrata. On M-stationarity conditions in MPECs and the associated qualification conditions. *Mathematical Programming*, 168:229–259, 2018.

[2] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming Series B*, 95:3–51, 2003.

[3] F. Alvarez, J. López, and H. Ramírez C. Interior proximal algorithm with variable metric for second-order cone programming: applications to structural optimization and support vector machines. *Optimization Methods and Software*, 25(6):859–881, 2010.

[4] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Augmented Lagragrian methods under the constant positive linear dependence constraint qualification. *Mathematical Programming*, 111(1-2):5–32, 2008.

[5] R. Andreani, C. E. Echagüe, and M. L. Schuverdt. Constant-rank condition and second-order constraint qualification. *Journal of Optimization theory and Applications*, 146:255–266, 2010.

[6] R. Andreani, E. H. Fukuda, G. Haeser, H. Ramírez C., D. O. Santos, P. J. S. Silva, and T. P. Silveira. Erratum to: New constraint qualifications and optimality conditions for second order cone programs. *To appear in Set-Valued and Variational Analysis*, 2021.

[7] R. Andreani, E. H. Fukuda, G. Haeser, D. O. Santos, and L. D. Secchin. Optimality conditions for nonlinear second-order cone programming and symmetric cone programming. Technical report, 2019.

[8] R. Andreani, W. Gómez, G. Haeser, L. M. Mito, and A. Ramos. On optimality conditions for nonlinear conic programming. *To appear in Mathematics of Operations Research*, 2021.

[9] R. Andreani, G. Haeser, L. M. Mito, and H. Ramírez C. Sequential constant rank constraint qualifications for nonlinear semidefinite programming with applications. Technical report, 2020. Available at arXiv:2106.00775.

[10] R. Andreani, G. Haeser, L. M. Mito, and H. Ramírez C. Weak notions of nondegeneracy in nonlinear semidefinite programming. Technical report, 2020. Available at arXiv:2012.14810.

[11] R. Andreani, G. Haeser, L. M. Mito, H. Ramírez C., D. O. Santos, and T. P. Silveira. Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming. *To appear in Optimization Letters*, 2021.

[12] R. Andreani, G. Haeser, L. M. Mito, H. Ramírez C., and T. P. Silveira. First- and second-order optimality conditions for second-order cone and semidefinite programming under a constant rank condition. Technical report, 2021. Available at arXiv:2107.04693.

[13] R. Andreani, G. Haeser, M. L. Schuverdt, L. D. Secchin, and P. J. S. Silva. On scaled stopping criteria for a safeguarded augmented Lagrangian method with theoretical guarantees. *To appear in Mathematical Programming Computation*, 2021.

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

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

[16] R. Andreani, J. M. Martínez, and M. L. Schuverdt. On the relation between constant positive linear dependence condition and quasinormality constraint qualification. *Journal of Optimization Theory and Applications*, 125(2):473–485, 2005.

[17] D. P. Bertsekas. *Nonlinear Programming.* Athenas Scientific. Belmont, Mass, 1999.

[18] J. F. Bonnans and H. Ramírez C. Perturbation analysis of second-order cone programming problems. *Mathematical Programming, Series B*, 104:205–227, 2005.

[19] J. F. Bonnans and A. Shapiro. *Pertubation Analysis of Optimization Problems.* Springer-Verlag. Berlin, 2000.

[20] P. Bosch, J. López, H. Ramírez C., and H. Robotham. Support vector machine under uncertainty: An application for hydroacoustic classification of fish-schools in Chile. *Expert Systems with Applications*, 40(10):4029–4034, 2013.

[21] J.-S. Chen, J. J. Ye, J. Zhang, and J. Zhou. Exact formula for the second-order tangent set of the second-order cone complementarity set. *SIAM Journal on Optimization*, 29(4):2986–3011, 2019.

[22] X. Chi, Z. Wan, and Z. Hao. Second order sufficient conditions for a class of bilevel programs with lower level second-order cone programming problem. *Journal of Industrial & Management Optimization*, 11(4):1111–1125, 2015.

[23] E. H. Fukuda, G. Haeser, and L. M. Mito. On the weak second-order optimality condition for nonlinear semidefinite and second-order cone programming. Technical report, 2020. Available at Optmization Online.

[24] J. Gauvin and F. Dubeau. Differential properties of the marginal function in mathematical programming. *Mathematical Programming Study*, 19:101–119, 1982.

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

[26] H. Gfrerer and 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.

[27] R. Henrion, A. Y. Kruger, and J. V. Outrata. Some remarks on stability of generalized equations. *Journal of Optimization Theory and Applications*, 159:681–697, 2013.

[28] M. R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4:303–320, 1969.

[29] R. Janin. Directional derivative of the marginal function in nonlinear programming. *Mathematical Programming Study*, 21:127–138, 1984.

[30] H. Kato and M. Fukushima. An SQP-type algorithm for nonlinear second-order cone programs. *Optimization Letters*, 1(2):129–144, 2007.

[31] M. Kočvara and M. Stingl. PENNON: A code for convex nonlinear and semidefinite programming. *Optimization Methods and software*, 18(3):317–333, 2003.

[32] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and Applications*, 284:193–228, 1998.

[33] J. P. Luna, C. Sagastizábal, and P. J. S. Silva. A discussion on electricity prices, or the two sides of the coin. *Philosophical Transactions of the Royal Society A*, 379: 20190428, 2021.

[34] L. Minchenko and S. Stakhovski. On relaxed constant rank regularity condition in mathematical programming. *Optimization*, 60(4):429–440, 2011.

[35] L. Minchenko and S. Stakhovski. Parametric nonlinear programming problems under the relaxed constant rank condition. *SIAM Journal on Optimization*, 21:314–332, 2011.

[36] B. S. Mordukhovich, N. T. V. Hang, and M. E. Sarabi. Augmented Lagrangian method for second-order conic programs under second-order sufficiency. *To appear in Journal of Global Optimization*, 2021.

[37] B. S. Mordukhovich, J. V. Outrata, and H. Ramírez C. Graphical derivatives and stability analysis for parameterized equilibria with conic constraints. *Set-Valued and Variational Analysis*, 23:687–704, 2015.

[38] B. S. Mordukhovich, J. V. Outrata, and H. Ramírez C. Second-order variational analysis in conic programming with applications to optimality and stability. *SIAM Journal on Optimization*, 25(1):76–101, 2015.

[39] B. S. Mordukhovich, J. V. Outrata, and M. E. Sarabi. Full stability of locally optimal solutions in second-order cone programs. *SIAM Journal on Optimization*, 24(4):1581–1613, 2014.

[40] J. V. Outrata and H. Ramírez C. On the Aubin property of critical points to perturbed second-order cone programs. *SIAM Journal on Optimization*, 21(3):798–823, 2011.

[41] M. J. D. Powell. *A method for nonlinear constraints in minimization problems*. Academic Press. New York, 1982.

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

[43] R. T. Rockafellar. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control and Optimization*, 12:268–285, 1974.

[44] R. T. Rockafellar. Directional differentiability of the optimal value function in a nonlinear programming problem. *Mathematical Programming Study*, 21:213–226, 1984.

[45] A. Shapiro and M. K. H. Fan. On eigenvalue optimization. *SIAM Journal on Optimization*, 5(3):552–569, 1995.

[46] J. J. Ye and D. L. Zhu. New necessary optimality conditions for bilevel programs by combining the MPEC and value function approaches. *SIAM Journal on Optimization*, 20:1885–1905, 2010.

[47] Y. Zhang and L. Zhang. New constraint qualifications and optimality conditions for second order cone programs. *Set-Valued and Variational Analysis*, 27:693–712, 2019.