Condições de otimalidade para otimização cônica

Daiana dos Santos Viana

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
DOUTOR EM CIÊNCIAS

Programa: Matemática Aplicada
Orientador: Prof. Dr. Gabriel Haeser

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES

São Paulo, fevereiro de 2019

Condições de otimalidade para otimização cônica

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 27/02/2019. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Ernesto G. Birgin - IME-USP
- Prof. Dr. Antoine Laurain - IME-USP
- Prof. Dr. José Mario Martínez - Unicamp
- Prof. Dr. Roberto Andreani - Unicamp
- Prof. Dr. Douglas Soares Gonçalves - UFSC

Agradecimentos

- À Deus por todas as graças concebidas em minha vida.
- À minha família pelo apoio e compreensão em especial ao meu filho Fágner.
- Aos membros da banca examinadora pelas sugestões dadas e melhoria desta tese.
- Aos professores do Programa de Pós-Graduação em Matemática Aplicada que contribuíram para minha formação acadêmica.
- Aos colegas do doutorado pelo apoio constante.
- À UFAC pela liberação durante o doutorado e colegas de trabalho.
- À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES pelo apoio financeiro.

Resumo

VIANA, D.S. **Condições de otimalidade para otimização cônica**. 2019. 120 f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2019.

Neste trabalho, realizamos uma extensão da chamada condição Aproximadamente Karush-Kuhn-Tucker (AKKT), inicialmente introduzida em programação não linear [AHM11], para os problemas de otimização sob cones simétricos não linear. Uma condição nova, a qual chamamos Trace AKKT (TAKKT), também foi apresentada para o problema de programação semidefinida não linear. TAKKT se mostrou mais prática que AKKT para programação semidefinida não linear. Provamos que, tanto a condição AKKT como a condição TAKKT são condições de otimalidade. Resultados de convergência global para o método de Lagrangiano aumentado foram obtidos. Condições de qualificação estritas foram introduzidas para medir a força dos resultados de convergência global apresentados. Através destas condições de qualificação estritas, foi possível verificar que nossos resultados de convergência global se mostraram melhores do que os conhecidos na literatura. Também apresentamos uma prova para um caso particular da conjectura feita em [AMS07].
Palavras-chave: condições sequenciais de otimalidade, programação semidefinida não linear, programação sob cones simétricos não linear, condições de qualificação estritas.

Abstract

VIANA, D.S. **Optimality conditions for conical optimization**. 2019. 120 f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2019.

In this work, we perform an extension of the so-called Approximate Karush-Kuhn-Tucker (AKKT) condition, initially introduced in nonlinear programming [AHM11], for nonlinear symmetric cone programming. A new condition, which we call Trace AKKT (TAKKT), was also presented for the nonlinear semidefinite programming problem. TAKKT proved to be more practical than AKKT for nonlinear semidefinite programming. We prove that both the AKKT condition and the TAKKT condition are optimality conditions. Results of global convergence for the augmented Lagrangian method were obtained. Strict qualification conditions were introduced to measure the strength of the overall convergence results presented. Through these strict qualification conditions, it was possible to verify that our results of global convergence proved to be better than those known in the literature. We also present a proof for a particular case of the conjecture made in [AMS07].

Keywords: Sequential optimality conditions, nonlinear semidefinite programming, nonlinear symmetric cone programming, strict qualification conditions.

Sumário

Lista de Abreviaturas

NLP	Nonlinear programming
NSDP	Nonlinear semidefinite programming
NSCP	Nonlinear symmetric cone programming
NSOCP	Nonlinear second-order cone programming
CQ	Constraint qualification
MFCQ	Mangasarian-Fromovitz constraint qualification
LICQ	Linear independence constraint qualification
KKT	Karush-Kuhn-Tucker
AKKT	Approximate KKT
TAKKT	Trace aproximadamente KKT
CAKKT	Complementarity aproximadamente KKT
CRCQ	Constant rank constraint qualification
RCRCQ	Relaxed constant rank constraint qualification
CRSC	Constant rank of the subspace component
WSOC	Weak second-order condition

Lista de Símbolos

$\mathbb{N}$	Conjunto dos números naturais
$\mathbb{R}$	Conjunto dos números reais
$\mathbb{R}^n$	Espaço vetorial $n-$dimensional sobre o conjunto dos números reais
$\mathbb{S}^m$	Conjunto das matrizes simétricas $m \times m$ com entradas reais
$\mathbb{S}^m_+$	Conjunto das matrizes simétricas $m \times m$ semidefinidas positivas com entradas reais
$\mathbb{S}^m_-$	Conjunto das matrizes simétricas $m \times m$ semidefinidas negativas com entradas reais
$A_{i,j}$	Matriz com entrada na $i-$ésima linha e $j-$ésima coluna
$\|A\|$	Norma de Frobenius de uma matriz A
$A \preceq 0$	Matriz semidefinida negativa
$A \succeq 0$	Matriz semidefinida positiva
$\lambda_k(A)$	k-ésimo autovalor da matriz A em ordem crescente
$\lambda_k^U(A)$	k-ésimo autovalor da matriz A com ordenação dada pela matriz ortogonal U
$\text{tr}(A)$	Traço da matriz A
$\langle A, B \rangle$	Produto interno matricial
$[A]_+$	Projeção da matriz A sobre $\mathbb{S}^m$
∇f	Gradiente de uma função $f : \mathbb{R}^n \to \mathbb{R}$
DG	Derivada de uma função $G : \mathbb{R}^n \to \mathbb{S}^m$
G_i	Derivada parcial da função $G : \mathbb{R}^n \to \mathbb{S}^m$ com respeito a x_i
DG^*	Adjunta da derivada de uma função $G : \mathbb{R}^n \to \mathbb{S}^m$
$\mathscr{F}$	Conjunto viável do problema de NLSDP

Capítulo 1

Introdução

O problema de otimização consiste em encontrar valores o melhor possível para uma função objetivo respeitando restrições que possam ser impostas às variáveis.

Um dos campos de estudo dessa área é a otimização não linear, também conhecida como *programação não linear* (NLP)[1]. Em NLP, estamos interessados em minimizar uma função objetivo não linear sujeito à restrições de igualdade e desigualdade não lineares. Entretanto, alguns problemas matemáticos podem ser naturalmente formulados de forma diferente de um NLP. Sendo assim, faz-se necessário estender essa classe de problemas permitindo-se assim contemplar um maior número de problemas matemáticos com a teoria desenvolvida.

Algumas classes de problemas em otimização que estendem um NLP foram propostas nos últimos anos. A primeira classe de problemas que podemos citar é formada por problemas de otimização onde as restrições são de pertinência a um cone chamado *cone de segunda-ordem* (ou *cone de Lorentz*). A segunda classe de problemas, que possuem o problema de NLP como um caso particular, são aqueles problemas cujas restrições são matrizes simétricas que pertencem ao cone das matrizes semidefinidas não negativas. Todos esses problemas citados possuem algo em comum: seus pontos viáveis pertencem a cones simétricos. Desta forma, pretendemos realizar um estudo sobre problemas de otimização cujas restrições incluem a pertinência a um cone simétrico geral.

Todos os problemas citados anteriormente têm sido bastante estudados nos últimos anos. Sobre o problema de otimização sob o cone de segunda-ordem, muitas aplicações podem ser citadas para o caso linear como: design de filtros, design de treliças, otimização em robótica dentre outros. Em programação semidefinida não linear temos uma grande quantidade de aplicações nos mais diferentes campos de estudo, tal como: teoria de controle, otimização estrutural, otimização de materiais, problemas de autovalores e outros.[2]

Em programação não linear, um conceito que tem se mostrado bastante útil é a noção de condição sequencial de otimalidade introduzido inicialmente por Qi e Wei em [QW00] e posteriormente abordado por Andreani, Haeser e Martínez em [AHM11]. As condições sequenciais de otimalidade surgiram a partir da observação prática de que algoritmos param satisfazendo aproximadamente uma condição de otimalidade pontual. Condições sequenciais são uma formalização deste conceito onde prova-se que são genuínas condições de otimalidade, uma vez que são satisfeitas independentemente de condição de qualificação tal como: *condição de independência linear* (LICQ), *condição de Mangasarian Fromovitz* (MFCQ), *condição de dependência linear positiva constante* (CPLD) dentre outras.

Além disso, as sequências geradas por várias classes de algoritmos (como Lagrangiano aumentado, métodos de ponto interior, programação quadrática sequencial e métodos de restauração inexata, veja [AHSS12b]), são precisamente as sequências necessárias para verificar a condição sequencial de otimalidade. Esta propriedade torna as condições sequenciais de otimalidade ferramentas úteis para fornecer naturalmente uma condição de otimalidade perturbada, que é adequada para a definição de critérios de parada e análise de complexidade para vários algoritmos. Além disso, um estudo cuidadoso da relação das condições sequenciais de otimalidade com as medidas clássicas de estacionaridade sob uma condição de qualificação, pode produzir resultados de convergência global sob hipóteses fracas, como podemos ver em [AFSS19, AMRS16,

[1] Do inglês: nonlinear programming

[2] Para mais detalhes, consulte [FAN01, FNA02, FJV07, KT06, KKW03, QS06, ScL09, VBW98] e suas referências.

AMRS18].

Tendo em vista o quão poderosas são as condições sequenciais de otimalidade, pretendemos estender esse conceito para *programação não linear sob cones simétricos*. Embora esta classe de problemas seja, dentro do nosso ambiente de cones simétricos, o problema mais geral que podemos considerar, iniciamos nossa investigação trabalhando com o problema de programação semidefinida não linear e posteriormente estendemos os resultados para problemas com cones simétricos gerais.

Além da extensão da condição AKKT para problemas sob cones simétricos também iremos apresentar uma prova para um caso particular da conjectura feita no artigo [AMS07]. Essa conjectura determina que se em qualquer minimizador local de um problema de programação não linear, satisfazendo MFCQ, o posto do conjunto de gradientes das restrições ativas aumenta em no máximo uma unidade na vizinhança do ponto então uma condição de otimalidade de segunda-ordem que depende de um único multiplicador de Lagrange é satisfeita.

1.1 Objetivos

O objetivo central desta tese é estender as condições sequenciais de otimalidade conhecidas em programação não linear para programação não linear sob cones simétricos.

1.2 Contribuições

As principais contribuições deste trabalho são as seguintes:

1. Estendemos o conceito de condição sequencial de otimalidade para programação semidefinida não linear. A condição de otimalidade mais conhecida e útil em NLP chamada Approximate-Karush-Kuhn-Tucker (AKKT) foi estendida. Embora a extensão de AKKT de NLP para NSDP não tenha sido imediata conseguimos obter os mesmos resultados teoricos esperados: provamos que AKKT é uma condição de otimalidade, apresentamos um critério de parada para algoritmos baseado em AKKT e provamos que o Algoritmo de Lagrangiano aumentado baseado em penalização quadrática gera sequências AKKT.

2. Durante nossa investigação sobre condições sequenciais conseguimos definir uma condição nova tanto em NLP como em NSDP. Essa condição sequencial de otimalidade nova foi chamada Trace-AKKT (TAKKT). Conseguimos verificar que essa condição é realmente uma condição de otimalidade e que sob uma hipótese adicional o algoritmo de Lagrangiano aumentado baseado em penalização quadrática gera sequências TAKKT.

3. Estendemos os conceitos que apresentamos nessa tese para uma classe de problemas ainda maior chamada problemas de programação não linear sob cones simétricos com uma abordagem mais detalhada para o problema de programação não linear sob o cone de segunda-ordem.

4. Apresentamos uma prova para um caso particular da conjectura feita em [AMS07]. Essa conjectura trata de uma condição de otimalidade de segunda-ordem para o problema de programação não linear.

1.3 Organização do Trabalho

Esta tese é organizada no formato de coletânea de artigos, isto é, cada capítulo é um artigo desenvolvido durante o doutorado. Visando facilitar a leitura iniciamos cada capítulo com uma breve apresentação sobre o que trata o artigo.

No capítulo 2, apresentamos o artigo: *Optimality conditions and global convergence for nonlinear semidefinite programming*, desenvolvido em conjunto com os professores Roberto Andreani (Universidade de Campinas) e Gabriel Haeser (Universidade de São Paulo), publicado na revista Mathematical Programming, DOI:10.1007/s10107-018-1354-5. O objetivo deste artigo é estender a condição AKKT para programação semidefinida não linear. Além da conhecida condição AKKT apresentamos uma nova condição de otimalidade à qual chamamos Trace-AKKT (TAKKT). Provamos que AKKT e TAKKT são genuínas condições de otimalidade, que o algoritmo de Lagrangiano aumentado gera sequências AKKT e TAKKT sob uma hipótese de suavidade. Algumas discussões são feitas acerca da conhecida condição de complementariedade-AKKT (CAKKT) em NSDP.

No capítulo 3, apresentamos o artigo: *Optimality conditions for nonlinear symmetric cone programming*, desenvolvido em conjunto com os professores Ellen Hidemi Fukuda (Universidade de de Kyoto) e Gabriel Haeser (Universidade de São Paulo), em fase de conclusão. O objetivo deste artigo é estender os resultados apresentados no artigo do capítulo anterior, mais especificamente a condição AKKT, para problemas de otimização onde as restrições pertencem a um cone simétrico geral. Neste artigo um destaque é dado ao problema de otimização sob cones de segunda-ordem.

No capítulo 4, apresentamos o artigo: *On a conjecture in second-order optimality conditions*, desenvolvido em conjunto com os professores Roger Behling (Universidade Federal de Santa Catarina), Gabriel Haeser (Universidade de São Paulo) e Alberto Ramos (Universidade Federal do Paraná), publicado na revista Journal of Optimization Theory and Applications, DOI: 10.1007/s10957-018-1229-1. O objetivo deste artigo é apresentar uma prova para um caso particular da conjectura feita por Andreani, Martínez e Schuverdt no artigo [AMS07].

No capítulo 5, apresentamos as conclusões do trabalho.

Capítulo 2

Optimality conditions and global convergence for nonlinear semidefinite programming

Neste capítulo vamos apresentar o artigo: *Optimality conditions and global convergence for nonlinear semidefinite programming*. O objetivo deste artigo é introduzir o conceito de condição sequencial de otimalidade, bem conhecido em programação não linear, para programação semidefinida não linear. Para clarificar algumas ideias e discussões apresentadas no artigo, vamos realizar uma breve revisão desse conceito em NLP. Vamos começar considerando o seguinte problema de programação não linear:

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{Minimizar}} \quad & f(x), \\ \text{sujeito a} \quad & g(x) \in \mathbb{R}^m_+, \end{aligned} \tag{NLP}$$

onde $f\colon \mathbb{R}^n \to \mathbb{R}$ e $g\colon \mathbb{R}^n \to \mathbb{R}^m$ são funções continuamente diferenciáveis. Um ponto $\bar{x} \in \mathbb{R}^n$ é um ponto de Karush-Kuhn-Tucker (KKT) se existir multiplicador de Lagrange $\mu \in \mathbb{R}^m_+$ com $\mu_j = 0$ para $g_j(\bar{x}) < 0$ tal que

$$\nabla f(\bar{x}) + \sum_{j=1}^{m} \mu_j \nabla g_j(\bar{x}) = 0.$$

É sabido que as condições KKT não são condições de otimalidade, sendo necessário que o ponto $\bar{x}$ satisfaça alguma *condição de qualificação* (CQ). Em NLP existem várias condições de qualificação que podem ser utilizadas para obter uma condição de otimalidade, sendo algumas dessas: LICQ, MFCQ, CRCQ, RCRCQ, Abadie e outras. Grande parte dos algoritmos em NLP são iterativos e incapazes de verificar exatamente a condição exigida "KKT ou não-CQ". O que acontece na prática é uma verificação de que o atual iterando obtido através do algoritmo em questão satisfaz aproximadamente a condição KKT. Dessa forma, as condições sequenciais de otimalidade são ferramentas teóricas desenvolvidas para justificar o que acontece na prática nos algoritmos para NLP.

Definição 1 *Um ponto $\bar{x} \in \mathbb{R}^n$ satisfaz Aproximadamente-KKT (AKKT) se existirem sequências $x^k \to \bar{x}$ e $\{\mu^k\} \subset \mathbb{R}^m_+$ com $\mu^k_j = 0$ para $g_j(\bar{x}) < 0$ tais que*

$$\nabla f(x^k) + \sum_{j=1}^{m} \mu^k_j \nabla g_j(x^k) \to 0.$$

Através da definição acima é possível provar que AKKT é de fato uma condição de otimalidade onde, em contraste com KKT não é necessário assumir condições de qualificação. Podemos definir um critério de parada para algoritmos usando a definição de AKKT e provar resultados de convergência global para algoritmos tal como o Algoritmo de Lagrangiano aumentado e outros.

Embora AKKT seja considerada uma das mais importantes condições sequenciais, Andreani, Martínez e Svaiter apresentam em [AMS10] uma nova condição de otimalidade mais forte que AKKT, à qual em certas situações pode ser mais conclusiva que AKKT.

Definição 2 *Um ponto* $\bar{x} \in \mathbb{R}^n$ *satisfaz Complementarity-AKKT (CAKKT) se existirem sequências* $x^k \to \bar{x}$ *e* $\{\mu^k\} \subset \mathbb{R}^m_+$ *tais que*

$$\nabla f(x^k) + \sum_{j=1}^{m} \mu_j^k \nabla g_j(x^k) \to 0,$$

$$\mu_j^k g_j(x^k) \to 0, \quad j = 1, \ldots, m.$$

Analisando as condições AKKT e CAKKT podemos notar que a principal diferença entre elas está no fato de que CAKKT fornece informações sobre multiplicadores de restrições ativas. Observe que em AKKT o multiplicador associado a uma restrição ativa no ponto pode divergir arbitrariamente rápido. Por outro lado, CAKKT impõe um limite na taxa de crescimento do multiplicador. Similar a AKKT, podemos apresentar resultados de convergência global para o Algoritmo de Lagrangiano aumentado utilizando uma hipótese de suavidade.

No artigo a seguir iremos definir AKKT para o problema de *programação semidefinida não linear* (NSDP)[1]. Em NSDP alguns cuidados foram necessários para definir AKKT. Primeiramente, o conceito de restrição inativa não é tão claro em NSDP como em NLP. Segundo, em NLP é muito fácil relacionar a restrição inativa com o multiplicador nulo, em NSDP essa relação exige um pouco mais de cuidado. No caso particular em que NLP e NSDP coincidem, as definições de AKKT para NLP e AKKT para NSDP coincidem, para tanto, algumas exigências foram necessárias. Uma interessante discussão sobre a razão de não definirmos CAKKT para NSDP da mesma maneira como definimos AKKT em NSDP também é apresentada. Embora não tenhamos definido CAKKT, definimos uma nova condição sequencial chamada *Trace*-AKKT (TAKKT) para NSDP. A condição TAKKT se mostrou mais prática e natural dentro do contexto de programação semidefinida não linear.

2.1 Article: Optimality conditions and global convergence for nonlinear semidefinite programming

Sequential optimality conditions have played a major role in unifying and extending global convergence results for several classes of algorithms for general nonlinear optimization. In this paper, we extend these concepts for nonlinear semidefinite programming. We define two sequential optimality conditions for nonlinear semidefinite programming. The first is a natural extension of the so-called Approximate-Karush-Kuhn-Tucker (AKKT), well known in nonlinear optimization. The second one, called Trace-AKKT (TAKKT), is more natural in the context of semidefinite programming as the computation of eigenvalues is avoided. We propose an Augmented Lagrangian algorithm that generates these types of sequences and new constraint qualifications are proposed, weaker than previously considered ones, which are sufficient for the global convergence of the algorithm to a stationary point.

Keywords: Nonlinear semidefinite programming, Optimality conditions, Constraint qualifications, Practical algorithms.

2.2 Introduction

Nonlinear semidefinite programming (NLSDP) is a generalization of the usual nonlinear programming problem, where the inequality constraints are replaced by a conic constraint defined by the matrix of constraints being negative semidefinite. The study of nonlinear semidefinite programming problems has grown a great deal in recent years from several application fields, such as control theory, structural optimization, material optimization, eigenvalue problems and others. See [FAN01, FNA02, FJV07, KT06, KKW03, QS06, ScL09, VBW98] and references therein. Theoretical issues such as optimality conditions, duality and nondegeneracy were also studied and several algorithms for solving NLSDPs have been proposed. The interested

[1] Do inglês: nonlinear semidefinite programming

reader may see, for instance, [ABMS07, BSS06, BS00, GR10, Jar12, LWC12, Sha97, SS04, YY15, ZZ14]. The particular case of (linear) semidefinite programming, which generalizes linear programming, has seen a variety of applications in several fields, and is an important tool in numerical analysis [cS07, For00, Tod03, VB96].

For nonlinear programming (NLP), a useful concept is the notion of sequential optimality conditions [AHM11]. These conditions are genuine necessary optimality conditions, independently of the fulfillment of any constraint qualification, such as the linear independence of the gradients of active constraints (LICQ), the Mangasarian-Fromovitz constraint qualification (MFCQ), or the constant positive linear dependence (CPLD). Besides that, the sequences generated by several classes of algorithms (as Augmented Lagrangians, interior point methods, sequential quadratic programming and inexact restoration methods; see [AHSS12b]), are precisely the sequences required for verifying the sequential optimality condition. This property makes sequential optimality conditions useful tools for naturally providing a perturbed optimality condition, which is suitable for the definition of stopping criteria and complexity analysis for several algorithms. Also, a careful study of the relation of sequential optimality conditions with classical stationarity measures under a constraint qualification, yields global convergence results under weak assumptions [AFSS19, AMRS16, AMRS18].

Sequential optimality conditions have been shown to be important tools for extending and unifying global convergence results for NLP algorithms. In NLSDP, most of the methods proposed need constraint qualifications to guarantee global convergence results. Both [SXA15] and [M.S05] present an Augmented Lagrangian method for solving NLSDPs. Although [SXA15] uses quadratic penalty, and [M.S05] uses modified barrier functions in their Augmented Lagrangian methods, both need constraint qualifications such as CPLD, MFCQ or Nondegeneracy. With sequential optimality conditions, we introduce a very natural optimality condition for NLSDP, without constraint qualifications, that is fulfilled by the limit points of an Augmented Lagrangian algorithm. With this approach, we have also been able to introduce a companion constraint qualification, weaker than the ones previously considered for NLSDP, in order to provide a classical global convergence result to a KKT point.

In this article we extend the Approximate-Karush-Kuhn-Tucker (AKKT) condition, known in nonlinear programming, to the context of NLSDP. We also introduce a new sequential optimality condition that we call Trace-AKKT (TAKKT). We will show that these conditions are genuine necessary optimality conditions for nonlinear semidefinite programming and we will show that an Augmented Lagrangian method produces AKKT and TAKKT sequences.

This paper is organized as follows: In Section 3.3, we formally define the nonlinear semidefinite programming, the notation, and we present some preliminary results for NLSDP needed in the remaining sections. In Section 2.4, we define the AKKT condition and we prove that it is a genuine optimality condition without constraint qualifications. In Section 2.5, we show that an Augmented Lagrangian method, based on the quadratic penalization, generates a sequence whose feasible limit points satisfy AKKT. In Section 2.6, we introduce a new sequential optimality condition called TAKKT, which is more suitable for nonlinear semidefinite programming. In Section 2.7, we introduce new constraint qualifications, weaker than the ones previously considered for global convergence results of nonlinear semidefinite programming algorithms, which are suitable for our global convergence analysis. In Section 2.8, we present our conclusions and final remarks.

2.3 Preliminaries

In this section we introduce the main notations adopted throughout the paper, together with known optimality conditions for nonlinear semidefinite programming. We will consider the following problem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & G(x) \preceq 0, \end{aligned} \tag{NLSDP}$$

where $f:\mathbb{R}^n \to \mathbb{R}$ and $G:\mathbb{R}^n \to \mathbb{S}^m$ are continuously differentiable functions. For simplicity, we do not take equality constraints into consideration. We denote by $\mathbb{S}^m$ the set of symmetric $m\times m$ matrices equipped with the inner product $\langle A,B\rangle := \mathrm{tr}(AB), A,B \in \mathbb{S}^m$, where $\mathrm{tr}(A)$ denotes the trace of the matrix A. The set $\mathscr{F}$ will denote the feasible set $\mathscr{F} := \{x \in \mathbb{R}^n \mid G(x) \preceq 0\}$. The notation $G(x) \preceq 0$ means that $G(x)$ is a symmetric negative semidefinite matrix. The set of symmetric positive semidefinite matrices will be denoted by $\mathbb{S}^m_+$, while the set of symmetric negative semidefinite matrices will be denoted by $\mathbb{S}^m_-$. Given a matrix $A \in \mathbb{S}^m$ and an orthogonal diagonalization $A = U\Lambda U^T$ of A, we denote by $\lambda^U_i(A)$ the eigenvalue of A at position i on the diagonal matrix Λ, that is, $\lambda^U_i(A) = \Lambda_{ii}, i = 1,\dots,m$. We omit U when dealing with a diagonalization that places the eigenvalues of A in ascending order in Λ, namely, $\lambda_1(A) \le \cdots \le \lambda_m(A)$.

The Frobenius norm of a matrix $A \in \mathbb{S}^m$ is the norm associated with the companion inner product, that is,

$$\|A\| := \sqrt{\langle A,A\rangle} = \sqrt{\sum_{i,j=1}^m A^2_{i,j}} = \sqrt{\sum_{i=1}^m \lambda_i(A)^2}.$$

Given $A \in \mathbb{S}^m$, we denote by $[A]_+$ the projection of A onto $\mathbb{S}^m_+$. Namely, if

$$A = U\mathrm{diag}(\lambda^U_1(A),\dots,\lambda^U_m(A))U^T$$

is an orthogonal decomposition of A, then

$$[A]_+ := U\mathrm{diag}([\lambda^U_1(A)]_+,\dots,[\lambda^U_m(A)]_+)U^T,$$

where $[v]_+ := \max\{0,v\}, v \in \mathbb{R}$. It is clear that $[A]_+$ does not depend on the choice of U. The following lemma provides an upper bound on the inner product of two matrices in terms of the inner product of their ordered eigenvalues. The subsequent lemma provides additional information when the matrices are positive/negative semidefinite.

Lemma 1 (von Neumann-Theobald [The75]) *Let $A,B \in \mathbb{S}^m$ be given. Then,*

$$\langle A,B\rangle \le \sum_{i=1}^m \lambda_i(A)\lambda_i(B),$$

where equality holds if and only if A and B are simultaneously diagonalizable with eigenvalues in ascending order.

Lemma 2 ([M.S05]) *Let $A \in \mathbb{S}^m_+$ and $B \in \mathbb{S}^m_-$ be given. Then, the following conditions are equivalent:*

a) $\langle A,B\rangle = 0$,

b) $AB = 0$,

c) A and B are simultaneously diagonalizable by an orthogonal matrix U and $\lambda^U_i(A)\lambda^U_i(B) = 0$ for all $i = 1,\dots,m$,

d) A and B are simultaneously diagonalizable in ascending order and $\lambda_i(A)\lambda_i(B) = 0$ for all $i = 1,\dots,m$.

Note that with $A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$, $\langle A,B\rangle \neq 0$, but $\lambda_i(A)\lambda_i(B) = 0, i = 1,2$, even though A and B are simultaneously diagonalizable by the identity matrix I. Note however that, as claimed by Lemma 2, it is not the case that $\lambda^I_i(A)\lambda^I_i(B) = 0$ for all $i = 1,2$. The notation λ^U_i corrects the statement of this lemma in [M.S05], with a similar proof, which we omit. Note also that these matrices are not simultaneously diagonalizable in ascending order.

Another useful result is the Weyl Lemma below. A lower and upper bound for the eigenvalues of the sum of two symmetric matrices are obtained, by means of the sum of the eigenvalues of these matrices.

Lemma 3 (Weyl [HJ91]) *Let $A,B \in \mathbb{S}^m$ be given matrices. For $k = 1,\dots,m$,*

$$\lambda_1(A) + \lambda_k(B) \le \lambda_k(A+B) \le \lambda_m(A) + \lambda_k(B).$$

We proceed to define the main concepts associated with problem (NLSDP). We define the Lagrangian function $L : \mathbb{R}^n \times \mathbb{S}^m \to \mathbb{R}$ associated with (NLSDP) by

$$L(x,\Omega) := f(x) + \langle G(x), \Omega \rangle.$$

The derivative of the mapping $G : \mathbb{R}^n \to \mathbb{S}^m$ at a point $x \in \mathbb{R}^n$ is given by $DG(x) : \mathbb{R}^n \to \mathbb{S}^m$, defined by

$$DG(x)h := \sum_{i=1}^{n} G_i(x)h_i, \ h \in \mathbb{R}^n,$$

where $G_i(x) := \dfrac{\partial G(x)}{\partial x_i} \in \mathbb{S}^m, i = 1,\dots,m$, are the partial derivative matrices with respect to x_i. Also, we define the adjoint operator $DG(x)^* : \mathbb{S}^m \to \mathbb{R}^n$ by

$$DG(x)^*\Omega := (\langle G_1(x),\Omega\rangle,\dots,\langle G_n(x),\Omega\rangle)^T, \ \Omega \in \mathbb{S}^m,$$

where it is easy to see that $\langle DG(x)h, \Omega\rangle = \langle h, DG(x)^*\Omega\rangle$ for all $h \in \mathbb{R}^n$, $x \in \mathbb{R}^n$ and $\Omega \in \mathbb{S}^m$.

Definition 1 *We say that $\Omega \in \mathbb{S}^m$ is a Lagrange multiplier associated with $\bar{x} \in \mathbb{R}^n$ if the following Karush-Kuhn-Tucker (KKT) conditions hold:*

$$\nabla_x L(\bar{x},\Omega) = \nabla f(\bar{x}) + DG(\bar{x})^*\Omega = 0, \quad (2.1)$$
$$\langle G(\bar{x}),\Omega\rangle = 0, \quad (2.2)$$
$$G(\bar{x}) \in \mathbb{S}^m_-, \quad (2.3)$$
$$\Omega \in \mathbb{S}^m_+. \quad (2.4)$$

When there exists a Lagrange multiplier associated with $\bar{x}$, we say that $\bar{x}$ is a KKT point. Condition (2.2) is known as the complementarity condition, and it can be equivalently replaced by $\lambda_i^U(G(\bar{x}))\lambda_i^U(\Omega) = 0$ for all $i = 1,\dots,m$, where $G(\bar{x})$ and Ω are simultaneously diagonalizable by U. See Lemma 2.

In order to prove the validity of the KKT conditions at a local minimizer of (NLSDP), a constraint qualification is needed. In general, more stringent assumptions are needed in order to prove global convergence results of an algorithm, that is, that every feasible accumulation point of a sequence generated by the algorithm is a KKT point. The most common assumptions of these kind are the nondegeneracy constraint qualification and the Mangasarian-Fromovitz constraint qualification (MFCQ) that we define below. More details can be found, e.g., in [Sha97, BS00].

Definition 2 *We say that the feasible point $\bar{x} \in \mathscr{F}$ satisfies the nondegeneracy constraint qualification if the n-dimensional vectors $(e_i^T G_1(\bar{x})e_j,\dots,e_i^T G_n(\bar{x})e_j)$, $i,j = 1,\dots,m-r$ are linearly independent, where $r = \operatorname{rank} G(\bar{x})$ and the vectors $e_1,\dots,e_{m-r}$ form a basis of the null space of the matrix $G(\bar{x})$.*

Definition 3 *We say that the feasible point $\bar{x} \in \mathscr{F}$ satisfies the Mangasarian-Fromovitz constraint qualification if there is a vector $h \in \mathbb{R}^n$ such that $G(\bar{x}) + DG(\bar{x})h$ is negative definite.*

As in the nonlinear programming case, the Mangasarian-Fromovitz constraint qualification is equivalent to the non-emptiness and boundedness of the set of Lagrange multipliers. Similarly, the more stringent nondegeneracy condition (also called transversality condition), implies the uniqueness of the Lagrange multiplier, so this condition is treated in the literature as an analogous of the linear independence constraint qualification for nonlinear programming. This analogy is not perfect, as it does not reduce to the classical linear independence constraint qualification when the matrix $G(x)$ is diagonal. However, Lourenço, Fukuda and Fukushima [LFF16], showed that if one reformulates the NLSDP such as an NLP using squared slack variables, where the conic constraint is replaced by the equality constraints $G(x) + Y^2 = 0$ with $Y \in \mathbb{S}^m$, if $(x,Y) \in \mathbb{R}^n \times \mathbb{S}^m$ satisfies the linear independence constraint qualification for this new NLP, then x satisfies the nondegeneracy condition. On the other hand, if x satisfies the nondegeneracy condition, then (x,Y) satisfies the linear independence constraint qualification for the NLP, where Y is the square root of the positive semidefinite matrix $-G(x)$. In this sense, the nondegeneracy condition is essentially the linear independence constraint qualification for NLSDP. Finally, we note that since $G(\bar{x})\Omega = 0$, rank $G(\bar{x})$ + rank $\Omega \leq m$ always holds at a KKT point. When the equality holds, we say that $\bar{x}$ and Ω satisfy the strict complementarity.

In the survey paper [YY15], the two types of Augmented Lagrangian methods are discussed: one with quadratic penalization (by [SXA15]) and another with a modified barrier function (by [M.S05]), where both need a strict complementarity assumption for the local convergence results. In [YY12] and [YYH12], the authors present an interior point method for NLSDP where for the convergence results, the first one adopts MFCQ while the second one assumes nondegeneracy and strict complementarity. In [CR04], the global convergence of a sequential SDP (SSDP), which is a method based on sequential quadratic programming (SQP) for NLP, has been proved under MFCQ. Another global convergent method presented in [GR10] applies a filter algorithm to SSDP where MFCQ is also assumed. In constrast with most of the literature, we will prove global convergence for an Augmented Lagrangian method without assuming strict complementarity and under a constraint qualification weaker than MFCQ, that we define in Section 2.5. In particular, our results allow the algorithm to generate an unbounded sequence of Lagrange multipliers, without hampering the global convergence results to a KKT point.

2.4 Sequential Optimality Condition for NLSDPs

In this section we will define the notion of an Approximate-KKT (AKKT) point for nonlinear semidefinite programming. This concept has been defined for nonlinear programming in [AHM11, QW00]. Many first- and second-order global convergence proofs of algorithms have been done, under weak assumptions, based on this and similar notions [ABMS08, AFSS19, AHRS17, AHSS12b, AHSS12c, AMRS16, AMRS18, AMS16, ASS18, BM14, BGM+16, BHR18, BKM17, DDTA13, Hae18, HM15, HS11, MS03, AHSS12a, Ram16, TYW17].

The Approximate-KKT (AKKT) condition is the most natural sequential optimality condition in nonlinear programming. Most algorithms generate this type of sequence [ABMS07, AHSS12b, AMS16, BM14, MP00, QW00]. The following is a natural generalization of AKKT to the context of NLSDP.

Definition 4 *We say that $\bar{x} \in \mathbb{R}^n$ satisfies the Approximate-KKT (AKKT) condition if $G(\bar{x}) \preceq 0$ and there exist sequences $x^k \to \bar{x}$ and $\{\Omega^k\} \subset \mathbb{S}^m_+$ such that*

$$\lim_{k\to\infty} \nabla f(x^k) + DG(x^k)^*\Omega^k = 0, \tag{2.5}$$

$$\lambda_i^U(G(\bar{x})) < 0 \Rightarrow \lambda_i^{S_k}(\Omega^k) = 0, \textit{ for all } i = 1,\dots,m \textit{ and sufficiently large } k, \tag{2.6}$$

where $G(\bar{x}) = U diag(\lambda_1^U(G(\bar{x})),\dots,\lambda_m^U(G(\bar{x})))U^T$, $\Omega^k = S_k diag(\lambda_1^{S_k}(\Omega^k),\dots,\lambda_m^{S_k}(\Omega^k))S_k^T$, where U and S_k are orthogonal matrices with $S_k \to U$.

All KKT points satisfy AKKT by considering constant sequences. We refer to Section 2.7 concerning the reciprocal implication.

Note that the definition of AKKT is independent of the choices of U and of the sequence $\{S_k\}$. Note also that given any orthogonal S_k that diagonalizes Ω^k, one can take a convergent subsequence. Hence, the definition above restricts the choice of $\{\Omega^k\}$, in such a way that the limit of its eigenvectors coincides with the eigenvectors of $G(\bar{x})$. Note however that the corresponding eigenvalues of $\{\Omega^k\}$ may be unbounded.

One of the reasons the extension of AKKT from NLP to NLSDP is not straightforward lies in the fact that in NLSDP the concept of active/inactive constraint is not apparent, which plays a key role in the definition of AKKT for NLP. In NLSDP, the eigenvalues less than zero play the role of the inactive constraints, hence, the “corresponding” Lagrange multiplier should vanish. However, it is not clear how to “correspond” an inactive eigenvalue of the constraint with an eigenvalue of the Lagrange multiplier matrix to vanish (in the definition of a KKT point, this correspondence is made by the assumption of simultaneous diagonalization). In this sense, the relation given by $S_k \to U$ provides the necessary notion for pairing the eigenvalues, which makes it a natural assumption for defining AKKT. In some sense, the definition says that the matrices must be approximately simultaneously diagonalizable. Note that when considering NLP as a particular case of an NLSDP with diagonal matrix, by considering a diagonal Lagrange multiplier matrix, both matrices are simultaneously diagonalizable by the identity, and we arrive at the usual AKKT concept for NLP.

When $\{x^k\}$ is a sequence as in the above definition, we say that $\{x^k\}$ is an AKKT sequence. Note that the sequence does not have to be formed by feasible points. The sequence $\{\Omega^k\}$ will be called the corresponding dual sequence.

Let us see that AKKT is closely related to a natural stopping criterion for numerical algorithms:

Lemma 4 *A point $\bar{x} \in \mathbb{R}^n$ satisfies AKKT if, and only if, there are sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\Omega^k\} \subset \mathbb{S}^m_+$, $\{\varepsilon_k\} \subset \mathbb{R}$ with $x^k \to \bar{x}$, $\varepsilon_k \to 0^+$ such that for all k,*

$$\|\nabla f(x^k) + DG(x^k)^* \Omega^k\| \leq \varepsilon_k, \tag{2.7}$$

$$\|[G(x^k)]_+\| \leq \varepsilon_k, \tag{2.8}$$

$$\text{for all } i = 1, \ldots, m \text{ and sufficiently large } k, \quad \lambda_i^{U_k}(G(x^k)) < -\varepsilon_k \Rightarrow \lambda_i^{S_k}(\Omega^k) = 0, \tag{2.9}$$

$$\|U_k - S_k\| \leq \varepsilon_k, \tag{2.10}$$

where $G(x^k) = U_k diag(\lambda_1^{U_k}(G(x^k)), \ldots, \lambda_m^{U_k}(G(x^k)))U_k^T$, $\Omega^k = S_k diag(\lambda_1^{S_k}(\Omega^k), \ldots, \lambda_m^{S_k}(\Omega^k))S_k^T$ are orthogonal diagonalizations of $G(x^k)$ and Ω^k for all k.

Proof 1 *Suppose that $\bar{x} \in \mathbb{R}^n$ satisfies AKKT. Let us take diagonalizations U_k of $G(x^k)$, S_k of Ω^k and U of $G(\bar{x})$ in such a way that, for an appropriate subsequence, $U_k \to U$, $S_k \to U$ and (2.5) and (2.6) hold. Let us define the sequence $\{\varepsilon_k\} \subset \mathbb{R}$ in order to satisfy (2.7-2.10). Define*

$$\varepsilon_k := \max\{\|\nabla f(x^k) + DG(x^k)^*\Omega^k\|, \|[G(x^k)]_+\|, \|U_k - S_k\|, -\lambda_i^{U_k}(G(x^k)) : i \in I(\bar{x})\},$$

where $I(\bar{x})$ is the set of all $i \in \{1, \ldots, m\}$ such that $\lambda_i^U(G(\bar{x})) = 0$. Hence, (2.7), (2.8) and (2.10) hold. To prove (2.9), note that if $j_0 \in \{1, \ldots, m\}$ is such that $\lambda_{j_0}^{U_k}(G(x^k)) < -\varepsilon_k$, then $-\lambda_{j_0}^{U_k}(G(x^k)) > \varepsilon_k \geq -\lambda_j^{U_k}(G(x^k))$ for all $j \in I(\bar{x})$. In particular, $j_0 \notin I(\bar{x})$. Therefore, by the definition of AKKT, $\lambda_{j_0}^{S_k}(\Omega^k) = 0$. The definition of AKKT ensures also that $\varepsilon_k \to 0^+$.

Suppose now that conditions (2.7-2.10) are valid. Since $\{U_k\}$ and $\{S_k\}$ are bounded, by (2.10) we may take a subsequence and U, that diagonalizes $G(\bar{x})$, such that $U_k \to U$ and $S_k \to U$. The limit $\nabla f(x^k) + DG(x^k)^*\Omega^k \to 0$ follows trivially, while the continuity of the functions involved ensures that $\bar{x} \in \mathscr{F}$. Now, suppose that $\lambda_i^U(G(\bar{x})) < 0$. Then, since $\lambda_i^{U_k}(G(x^k)) \to \lambda_i^U(G(\bar{x}))$, for k large enough $\lambda_i^{U_k}(G(x^k)) < -\varepsilon_k$. Therefore, $\lambda_i^{S_k}(\Omega^k) = 0$ and AKKT is satisfied.

Lemma 4 provides a natural stopping criterion associated with AKKT. Given small tolerances $\varepsilon_{\text{opt}} > 0$, $\varepsilon_{\text{feas}} > 0$, $\varepsilon_{\text{diag}} > 0$ and $\varepsilon_{\text{compl}} > 0$ associated with optimality, feasibility, simultaneous diagonalization and complementarity, respectively, an algorithm that aims at solving NLSDP and generates an AKKT sequence $\{x^k\} \subset \mathbb{R}^n$ together with a dual sequence $\{\Omega^k\} \subset \mathbb{S}^m_+$ should be stopped at iteration k when

$$\|\nabla f(x^k) + DG(x^k)^*\Omega^k\| \leq \varepsilon_{\text{opt}}, \tag{2.11}$$

$$\|[G(x^k)]_+\| \leq \varepsilon_{\text{feas}}, \tag{2.12}$$

$$\|U_k - S_k\| \leq \varepsilon_{\text{diag}}, \tag{2.13}$$

$$\lambda_i^{U_k}(G(x^k)) < -\varepsilon_{\text{compl}} \Rightarrow \lambda_i^{S_k}(\Omega^k) = 0. \tag{2.14}$$

As it is usual in the literature of sequential optimality condition, three properties must be satisfied: i) It must be a genuine necessary optimality condition, independently of the fulfillment of constraint qualifications. This gives a meaning with respect to optimality to the stopping criterion (2.11-2.14), independently of constraint qualifications. ii) The condition must be satisfied by limit points of sequences generated by relevant algorithms. Note that this simplifies the usual global convergence proofs to KKT points, also providing a guide to how new algorithms could be proposed. iii) It must be a strong optimality condition. This is shown by defining a weak constraint qualification that makes every point that satisfies the condition a true KKT point.

We are going to show that AKKT satisfies all these requirement. In the remainder of this section we show that it satisfies the first requirement, that is, it is indeed a necessary optimality condition. For this, let us first consider the external penalty algorithm for the problem:

$$\text{Minimize } f(x), \text{ subject to } G(x) \preceq 0, \; x \in \Sigma, \tag{2.15}$$

where $\Sigma \subseteq \mathbb{R}^n$ is a non-empty closed set.

Theorem 1 *Choose a sequence $\{\rho_k\} \subset \mathbb{R}$ with $\rho_k \to +\infty$ and for each k, let x^k be the global solution, if it exists, of the problem*

$$\textit{Minimize } f(x) + \frac{\rho_k}{2} P(x), \quad \textit{subject to } x \in \Sigma,$$

where $P(x) := \operatorname{tr}\left([G(x)]_+^2\right) = \| [G(x)]_+ \|^2$ is the penalty function. Then, any limit point of this sequence, if any exist, is a solution of problem (2.15) provided that its feasible region is non-empty.

Proof 2 *See [FM68].*

The penalization function $x \mapsto P(x)$ is a measure of infeasibility of x, as $P(x) \geq 0$ is continuous, and x is feasible if, and only if, $P(x) = 0$. We will also need the following result about the derivative of $P(x)$:

Lemma 5 *Let $G : \mathbb{R}^n \to \mathbb{S}^m$ be a differentiable function and $P(x) := \operatorname{tr}\left([G(x)]_+^2\right)$. Then, the gradient of P at $x \in \mathbb{R}^n$ is given by*

$$\nabla P(x) = 2DG(x)^*[G(x)]_+.$$

Proof 3 *It follows from [Lew93, Corollary 3.2].*

We are now ready to prove the main result of this section.

Theorem 2 *Let $\bar{x}$ be a local minimizer of (NLSDP). Then, $\bar{x}$ satisfies AKKT.*

Proof 4 *To show that AKKT is an optimality condition we will apply the external penalty algorithm to the following problem*

$$\textit{Minimize } f(x) + \frac{1}{2}\|x - \bar{x}\|^2 \textit{ subject to } G(x) \preceq 0, \; x \in B(\bar{x}, \delta), \tag{2.16}$$

where $\delta > 0$ is small enough and $B(\bar{x}, \delta)$ denotes the closed ball of center $\bar{x}$ and radius δ. Clearly, δ can be chosen such that $\bar{x}$ is the unique solution of (2.16). Let $\rho_k \to +\infty$ and for each k, let x^k be a solution of

$$\textit{Minimize } f(x) + \frac{1}{2}\|x - \bar{x}\|^2 + \frac{\rho_k}{2} \operatorname{tr}\left([G(x)]_+^2\right) \textit{ subject to } x \in B(\bar{x}, \delta). \tag{2.17}$$

By the non-emptiness and compactness of $B(\bar{x}, \delta)$, and continuity of the objective function, x^k is well defined for all $k \in \mathbb{N}$. By the boundedness of $\{x^k\}$, the uniqueness of the solution of problem (2.16), and Theorem 1, $x^k \to \bar{x}$. Furthermore, note that x^k is in the interior of $B(\bar{x}, \delta)$ for k large enough; hence, the gradient of $f(x) + \frac{1}{2}\|x - \bar{x}\|^2 + \frac{\rho_k}{2} P(x)$ vanishes at x^k, that is, from Lemma 5, we have

$$\nabla f(x^k) + x^k - \bar{x} + \rho_k DG(x^k)^* \left[G(x^k)\right]_+ = 0.$$

Defining $\Omega^k := \rho_k \left[G(x^k)\right]_+$, *and taking the limit when k tends to infinity in the above equality we obtain*

$$\lim_{k\to\infty} \nabla f(x^k) + DG(x^k)^*\Omega^k = 0.$$

Since $G(x^k)$, Ω^k, *and* $G(\bar{x})$ *are symmetric matrices we can diagonalize*

$$G(x^k) = U_k diag(\lambda_1^{U_k}(G(x^k)),\dots,\lambda_m^{U_k}(G(x^k)))U_k^T,$$

$$\Omega^k = \rho_k \left[G(x^k)\right]_+ = U_k diag(\rho_k[\lambda_1^{U_k}(G(x^k))]_+,\dots,\rho_k[\lambda_m^{U_k}(G(x^k))]_+)U_k^T,$$

and

$$G(\bar{x}) = U diag(\lambda_1^{U}(G(\bar{x})),\dots,\lambda_m^{U}(G(\bar{x})))U^T,$$

where $U_kU_k^T = I$, $UU^T = I$ *and* $U_k \to U$. *By the definition of* Ω^k, *the matrices* $G(x^k)$ *and* Ω^k *are simultaneously diagonalizable. Thus, the matrix* S_k *in the definition of AKKT is taken coinciding with* U_k. *Now, if* $\lambda_i^U(G(\bar{x})) < 0$ *then* $\lambda_i^{U_k}(G(x^k)) < 0$ *for all sufficiently large k; which implies that* $\lambda_i^{U_k}(\Omega^k) = \rho_k[\lambda_i^{U_k}(G(x^k))]_+ = 0$.

In the next example, we consider a local minimizer that does not satisfy the KKT conditions. Let us then build an AKKT sequence guaranteed to exist by Theorem 2.

Example 1 (AKKT sequence at a non-KKT solution) *Consider the following problem*

$$\begin{aligned} &\underset{x\in\mathbb{R}}{Minimize} && 2x,\\ &Subject\ to && G(x) := \begin{bmatrix} 0 & x \\ x & -1 \end{bmatrix} \preceq 0. \end{aligned}$$

Since the unique feasible point is $\bar{x} := 0$, *this is the unique global minimizer.*

(i) $\bar{x} = 0$ *is not a KKT point. Let* $\Omega := \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{12} & \mu_{22} \end{bmatrix}$. *We have*

$$\nabla f(\bar{x}) + DG(\bar{x})^*\Omega = 2 + 2\mu_{12} = 0 \Rightarrow \mu_{12} = -1,$$

and

$$\langle G(\bar{x}), \Omega\rangle = 0 \Rightarrow \mu_{22} = 0.$$

Thus, $\Omega = \begin{bmatrix} \mu_{11} & -1 \\ -1 & 0 \end{bmatrix}$, *which is an indefinite matrix, regardless of* μ_{11}. *Therefore,* $\bar{x}$ *can not be a KKT point.*

(ii) $\bar{x} = 0$ *is an AKKT point. Define* $x^k := -\dfrac{k}{k^2-1}$ *and* $\Omega^k := \begin{bmatrix} k & -1 \\ -1 & 1/k \end{bmatrix}$. *Then,*

$$\nabla f(x^k) + DG(x^k)^*\Omega^k = 2 - 2 = 0.$$

Note that, in the way that we have defined x^k *and* Ω^k, *we have that* $\Omega^kG(x^k) = G(x^k)\Omega^k$, *which is a well known necessary and sufficient condition for simultaneous diagonalization. Computing the eigenvectors,* $G(x^k)$ *and* Ω^k *share a common basis of eigenvectors given by* $u_1^k := \left(\dfrac{1}{\sqrt{1+k^2}}, \dfrac{k}{\sqrt{1+k^2}}\right)$ *and* $u_2^k := \left(\dfrac{-k}{\sqrt{1+k^2}}, \dfrac{1}{\sqrt{1+k^2}}\right)$, *with corresponding eigenvalues*

$$\lambda_1^{U_k}(G(x^k)) = \frac{-k^2}{k^2-1},\ \lambda_2^{U_k}(G(x^k)) = \frac{1}{k^2-1},\ \lambda_1^{S_k}(\Omega^k) = 0,\ \lambda_2^{S_k}(\Omega^k) = \frac{1}{k} + k,$$

where $U_k = S_k$ is the matrix with first and second columns given by u_1^k and u_2^k, respectively. Let $U := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ be the limit of $\{S_k\}$. Now, since $\lambda_1^U(G(\bar{x})) = -1 < 0$ and $\lambda_2^U(G(\bar{x})) = 0$, AKKT holds since the eigenvalue of Ω^k corresponding to the negative eigenvalue is $\lambda_1^{S_k}(\Omega^k)$, which is zero for all k.

2.5 An Augmented Lagrangian algorithm that generates AKKT sequences

In this section we will present an Augmented Lagrangian method for NLSDP based on the quadratic penalty function, which we will prove to generate an AKKT sequence. This is one of the most popular methods for solving NLSDPs [SSZ08]. For more details, see [ABMS07, cS03, cS10, HTY06, SXA15, SZW06, YY15]. Our algorithm is inspired by the Augmented Lagrangian method with safeguards for nonlinear programming, introduced in [ABMS07, ABMS08] (see also [BM14]), which turns out to be the same as the one from [SXA15] for NLSDP. Let us discuss in some details the contributions of [SXA15], which is most related to our approach. In [SXA15], the authors studied global convergence properties of four Augmented Lagrangian methods using different algorithmic strategies. For their study, the following NLSDP is considered:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & G(x) \preceq 0,\, h(x) = 0,\, x \in \mathscr{V}, \end{aligned} \tag{2.18}$$

where $\mathscr{V} := \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$ and $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^p$, $g : \mathbb{R}^n \to \mathbb{R}^q$ and $G : \mathbb{R}^n \to \mathbb{S}^m$ are continuously differentiable functions. Note that this problem differs from ours by the presence of additional standard nonlinear constraints. For (2.18), an Augmented Lagrangian method is presented using the safeguarded technique from [BM14]. The constraints $G(x) \preceq 0$ and $h(x) = 0$ are penalized (upper-level constraints) with the use of the Powell-Hestenes-Rockafellar (PHR) Augmented Lagrangian function for NLSDP, while the constraints in $\mathscr{V}$ are kept in the subproblems (lower-level constraints). The proposed algorithm then finds an approximate minimizer of the Augmented Lagrangian function over the set $\mathscr{V}$, followed by standard updates of the penalty parameter and safeguarded multipliers approximations. They then proceed to investigate global convergence results in a similar fashion as in [ABMS08, AHSS12c]. In one of their main results, the authors show that a limit point is either feasible or, if the constant positive linear dependence (CPLD) constraint qualification with respect to $\mathscr{V}$ holds, then it is a KKT point for an infeasibility problem with respect to the upper-level constraints. They then resort to MFCQ in order to prove that a feasible limit point is a KKT point of (2.18). That is, they only consider a weak constraint qualification (CPLD) for the standard nonlinear programming constraints in $\mathscr{V}$, but not for the conic constraints $G(x) \preceq 0$. We will provide global convergence results for this Augmented Lagrangian method for (NLSDP), which can be formulated with a new constraint qualification weaker than MFCQ. That is, for simplicity, we will consider problems only with conic constraints and no additional nonlinear constraints. Any additional nonlinear constraints could also be penalized or incorporated in the conic constraints in a standard way. Although we consider only unconstrained Augmented Lagrangian subproblems, the theory we develop is general enough in order to consider even subproblems with conic constraints, with an approach similar to what is done in [ABMS07, BHR18]. We leave the details of this more general approach for a later study, while we now focus on an Augmented Lagrangian method that penalizes all constraints and solves unconstrained subproblems.

Given a penalty parameter $\rho > 0$, the Augmented Lagrangian function $L_\rho : \mathbb{R}^n \times \mathbb{S}^m \to \mathbb{R}$, associated with (NLSDP), is defined by:

$$L_\rho(x, \Omega) := f(x) + \frac{1}{2\rho} \left\{ \|[\Omega + \rho G(x)]_+\|^2 - \|\Omega\|^2 \right\}. \tag{2.19}$$

This function is a natural extension of the Augmented Lagrangian function for nonlinear programming.

Similarly to the external penalty method, our goal is to solve (NLSDP) by solving a sequence of unconstrained minimization problems with respect to x, obtaining at each iteration k a primal iterate x^k, where the objective function is given by (2.19), and $\Omega := \bar{\Omega}^k$ and $\rho := \rho_k$ are iteratively updated according to some performance criterion. We expect that, for suitable choices of the parameters, the sequence of solutions generated will converge to the solution of problem (NLSDP). From the definition of the Augmented Lagrangian function above, we have that its gradient, with respect to x, is given by:

$$\nabla_x L_\rho(x,\Omega) = \nabla f(x) + DG(x)^* [\Omega + \rho G(x)]_+ .$$

Therefore, a natural update rule for the dual sequence is:

$$\Omega \leftarrow [\Omega + \rho G(x)]_+ ,$$

however, to avoid solving the next subproblem with a Lagrange multiplier approximation too large (given that we do not assume MFCQ, and Lagrange multipliers may be unbounded), we take the Lagrange multiplier approximation for the next subproblem as the projection of the natural update above onto a safeguarded box as in [BM14]. This means that when the penalty parameter is large, the method reduces to the external penalty method. To update the penalty parameter, we will rely on a joint measure of feasibility and complementarity. The formal statement of the algorithm is as follows:

Algorithm 1 Augmented Lagrangian Algorithm

STEP 0 (Initialization): Let $\tau \in (0,1)$, $\gamma > 1$, $\rho_1 > 0$ and $\Omega^{\max} \in \mathbb{S}^m_+$. Take $\{\varepsilon_k\} \subset \mathbb{R}$ a sequence of positive scalars such that $\lim \varepsilon_k = 0$. Define $0 \preceq \bar{\Omega}^1 \preceq \Omega^{\max}$. Choose $x^0 \in \mathbb{R}^n$ an arbitrary starting point. Initialize $k := 1$.

STEP 1 (Solve subproblem): Use x^{k-1} to find an approximate minimizer x^k of $L_{\rho_k}(x, \bar{\Omega}^k)$, that is, a point x^k such that

$$\|\nabla_x L_{\rho_k}(x^k, \bar{\Omega}^k)\| \leq \varepsilon_k .$$

STEP 2 (Penalty parameter update): Define

$$V^k := \left[\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right]_+ - \frac{\bar{\Omega}^k}{\rho_k}.$$

If

$$\|V^k\| \leq \tau \|V^{k-1}\|,$$

define

$$\rho_{k+1} := \rho_k,$$

otherwise, define

$$\rho_{k+1} := \gamma \rho_k.$$

STEP 3 (Multiplier update): Compute

$$\Omega^k := \left[\bar{\Omega}^k + \rho_k G(x^k)\right]_+ ,$$

and define $\bar{\Omega}^{k+1} := proj_S(\Omega^k)$, the orthogonal projection of Ω^k onto S, where $S := \{X \in \mathbb{S}^m | 0 \preceq X \preceq \Omega^{\max}\}$. Set $k := k+1$, and go to Step 1.

Note that $V^k = 0$ if, and only if x^k is feasible and complementarity holds. Indeed,

$$V^k = 0 \Rightarrow \left[\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right]_+ = \frac{\bar{\Omega}^k}{\rho_k} \Rightarrow \bar{\Omega}^k = \left[\bar{\Omega}^k + \rho_k G(x^k)\right]_+ ,$$

and then, computing the orthogonal decomposition, we have

$$\bar{\Omega}^k + \rho_k G(x^k) = U_k \mathrm{diag}(\lambda_1^{U_k}, \ldots, \lambda_m^{U_k}) U_k^T \text{ and}$$

$$\bar{\Omega}^k = [\bar{\Omega}^k + \rho_k G(x^k)]_+ = U_k \mathrm{diag}([\lambda_1^{U_k}]_+, \ldots, [\lambda_m^{U_k}]_+) U_k^T .$$

Hence,

$$G(x^k) = (1/\rho_k)U_k\text{diag}(\lambda_1^{U_k} - [\lambda_1^{U_k}]_+, \ldots, \lambda_m^{U_k} - [\lambda_m^{U_k}]_+)U_k^T.$$

As $\lambda_i^{U_k} - [\lambda_i^{U_k}]_+ \leq 0$ for all $i = 1, \ldots, m$, we have $G(x^k) \preceq 0$. Also, the multiplier $\bar{\Omega}^k$ is such that, if $\lambda_i^{U_k}(G(x^k)) < 0$ then $\lambda_i^{U_k} < [\lambda_i^{U_k}]_+$ and thus, $\lambda_i^{U_k}(\bar{\Omega}^k) = [\lambda_i^{U_k}]_+ = \max\{0, \lambda_i^{U_k}\} = 0$. The reciprocal implication follows similarly. Thus, the algorithm keeps the previous penalty parameter unchanged if $\|V^k\|$ is sufficiently reduced, otherwise, the penalty parameter is increased to force feasibility and complementarity.

Note also that we consider $\{\Omega^k\}$ as the dual sequence generated by the algorithm in order to check a stopping criterion, since this sequence together with $\{x^k\}$ fulfills the stopping criterion related to optimality (2.11) when $\varepsilon_k \leq \varepsilon_{\text{opt}}$.

The following result shows that Algorithm 1 finds stationary points of an infeasibility measure. This shows that the algorithm tends to find feasible points, which are global minimizers of the infeasibility measure, whenever the feasible region is non-empty.

Theorem 3 *Let $\bar{x} \in \mathbb{R}^n$ be a limit point of a sequence $\{x^k\}$ generate by Algorithm 1. Then, $\bar{x}$ is a stationary point for the optimization problem*

$$\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad P(x) := \text{tr}\left([G(x)]_+^2\right). \tag{2.20}$$

Proof 5 *If $\{\rho_k\}$ is bounded, that is, for $k \geq k_0$ the penalty parameter remains unchanged, we have that $V^k \to 0$. The sequence $\{\bar{\Omega}^k\}$ is bounded, then, there is an infinite subset $K_1 \underset{\infty}{\subset} \mathbb{N}$ such that $\lim_{k \in K_1} \bar{\Omega}^k = \bar{\Omega}$. Now, from $V^k \to 0$ we get*

$$\bar{\Omega} = \lim_{k \in K_1} \bar{\Omega}^k = \lim_{k \in K_1} \left[\bar{\Omega}^k + \rho_{k_0} G(x^k)\right]_+ = \lim_{k \in K_1} \Omega^k = \left[\bar{\Omega} + \rho_{k_0} G(\bar{x})\right]_+.$$

The computation is now similar to our previous discussion where V^k was assumed to be zero. Writing the orthogonal decomposition of the matrix $\bar{\Omega}$, we have

$$\bar{\Omega} = [\bar{\Omega} + \rho_{k_0} G(\bar{x})]_+ = U diag([\lambda_1^U]_+, \ldots, [\lambda_m^U]_+)U^T,$$

with $UU^T = I$. Moreover,

$$\bar{\Omega} + \rho_{k_0} G(\bar{x}) = U diag(\lambda_1^U, \ldots, \lambda_m^U)U^T.$$

In this way,

$$G(\bar{x}) = (1/\rho_{k_0})U diag((\lambda_1^U - [\lambda_1^U]_+), \ldots, (\lambda_m^U - [\lambda_m^U]_+))U^T,$$

thus, $G(\bar{x}) \preceq 0$. Therefore, $\bar{x}$ is a global minimizer of the optimization problem (2.20). If $\{\rho_k\}$ is unbounded, let us define

$$\delta^k := \nabla f(x^k) + DG(x^k)^* \Omega^k,$$

where $\Omega^k := [\bar{\Omega}^k + \rho_k G(x^k)]_+$. Clearly, from Step 1 of the algorithm, we have that $\|\delta^k\| \leq \varepsilon_k$. Dividing δ^k by ρ_k,

$$\frac{\delta^k}{\rho_k} = \frac{\nabla f(x^k)}{\rho_k} + DG(x^k)^* \left[\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right]_+,$$

since $\bar{\Omega}^k$ is bounded and $\nabla f(x^k) \to \nabla f(\bar{x})$, we have that $DG(\bar{x})^*[G(\bar{x})]_+ = 0$. Thus, the result follows from Lemma 5.

Next, we will show that a feasible limit point of a sequence $\{x^k\}$ generated by the algorithm is an AKKT point. In fact, $\{x^k\}$ is an associated AKKT sequence and $\{\Omega^k\}$ is the corresponding dual sequence.

Theorem 4 *Assume that $\bar{x} \in \mathbb{R}^n$ is a feasible limit point of a sequence $\{x^k\}$ generated by Algorithm 1. Then, $\bar{x}$ is an AKKT point.*

Proof 6 *Let $\bar{x} \in \mathscr{F}$ be a limit point of a sequence $\{x^k\}$ generated by Algorithm 1. Let us assume without loss of generality that $x^k \to \bar{x}$. From Step 1 of the algorithm we have that*

$$\|\nabla f(x^k) + DG(x^k)^* \Omega^k\| \leq \varepsilon_k \Rightarrow \lim_{k\to\infty} \nabla f(x^k) + DG(x^k)^* \Omega^k = 0,$$

where

$$\Omega^k = \left[\bar{\Omega}^k + \rho_k G(x^k)\right]_+ .$$

Now, let us prove that for appropriate matrices U and $S_k \to U$, we have that for all $i = 1, \dots, m$, if $\lambda_i^U(G(\bar{x})) < 0$ then $\lambda_i^{S_k}(\Omega^k) = 0$ for all sufficiently large k. We have two cases to analyze:

(i) If $\rho_k \to +\infty$, since the sequence $\{\bar{\Omega}^k\}$ is bounded we have that $\dfrac{\bar{\Omega}^k}{\rho_k} + G(x^k) \to G(\bar{x})$. Let us take a diagonalization

$$\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) = S_k diag(\lambda_1^k, \dots, \lambda_m^k) S_k^T,$$

where $S_k S_k^T = I$. Taking a subsequence if necessary, let us take a diagonalization

$$G(\bar{x}) = U diag(\lambda_1, \dots, \lambda_m) U^T,$$

where $UU^T = I$ with $S_k \to U$, $\lambda_i^k \to \lambda_i$ for all i. Then,

$$\Omega^k = [\bar{\Omega}^k + \rho_k G(x^k)]_+ = S_k diag(\rho_k [\lambda_1^k]_+, \dots, \rho_k [\lambda_m^k]_+) S_k^T.$$

Now, assume that $\lambda_i^U(G(\bar{x})) := \lambda_i < 0$. Then, $\lambda_i^k < 0$ for all sufficiently large k; which implies that $\lambda_i^{S_k}(\Omega^k) := \rho_k[\lambda_i^k]_+ = 0$.

(ii) If $\{\rho_k\}$ is bounded, similarly to what was done in Theorem 3, for $k \geq k_0$ we have $\rho_k = \rho_{k_0}$; hence, $V^k \to 0$. Thus, taking a subsequence if necessary, we have

$$\bar{\Omega} = \lim \bar{\Omega}^k = \lim \left[\bar{\Omega}^k + \rho_{k_0} G(x^k)\right]_+ = \lim \Omega^k = \left[\bar{\Omega} + \rho_{k_0} G(\bar{x})\right]_+ .$$

Let us take an orthogonal decomposition of the matrix $\bar{\Omega}^k + \rho_{k_0} G(x^k)$, that is,

$$\bar{\Omega}^k + \rho_{k_0} G(x^k) = S_k diag(\lambda_1^{S_k}, \dots, \lambda_m^{S_k}) S_k^T,$$

and let us take a subsequence such that $\{S_k\}$ converges to some orthogonal matrix U. Then,

$$\bar{\Omega} + \rho_{k_0} G(\bar{x}) = \lim \bar{\Omega}^k + \rho_{k_0} G(x^k) = U diag(\lambda_1^U, \dots, \lambda_m^U) U^T,$$

where $\lambda_i^{S_k} \to \lambda_i^U$ for all i. In this way, since

$$\bar{\Omega} = U diag([\lambda_1^U]_+, \dots, [\lambda_m^U]_+) U^T,$$

we have that

$$G(\bar{x}) = (1/\rho_{k_0}) U diag((\lambda_1^U - [\lambda_1^U]_+), \dots, (\lambda_m^U - [\lambda_m^U]_+)) U^T,$$

and then, $\lambda_i^U(G(\bar{x})) = \dfrac{\lambda_i^U - [\lambda_i^U]_+}{\rho_{k_0}}$. Assuming that it is negative, we have that $\lambda_i^U < [\lambda_i^U]_+$; hence, $\lambda_i^U < 0$. Then, $\lambda_i^{S_k} < 0$ for all sufficiently large k; which implies that $\lambda_i^{S_k}(\Omega^k) = [\lambda_i^{S_k}]_+ = 0$.

2.6 A new sequential optimality condition for NLSDP

In Section 2.4, we presented an extension of the classical AKKT sequential optimality condition known for NLP. However, for NLSDP, it turns out that we can define a much more natural and simpler sequential optimality condition, which is new even in the context of NLP. The new condition does not rely on eigenvalue computations for treating the complementarity, which are replaced by a simpler inner product of the constraint matrix and the dual matrix. The new condition also does not require the convergence $S_k \to U$, which relates the eigenvectors of the dual sequence with the ones of the limit of the constraint matrix. This new condition is called Trace-Approximate-KKT (TAKKT) and is defined below.

Definition 5 *We say that a point $\bar{x} \in \mathbb{R}^n$ satisfies the Trace-Approximate-KKT (TAKKT) condition if $G(\bar{x}) \preceq 0$ and there exist sequences $x^k \to \bar{x}$ and $\{\Omega^k\} \subset \mathbb{S}^m_+$ such that*

$$\lim_{k\to\infty} \nabla f(x^k) + DG(x^k)^*\Omega^k = 0, \tag{2.21}$$

$$\lim_{k\to\infty} \langle \Omega^k, G(x^k)\rangle = 0. \tag{2.22}$$

Note that, $\lim_{k\to\infty} \langle \Omega^k, G(x^k)\rangle = 0$ does not imply that Ω^k and $G(x^k)$ are (approximately) simultaneously diagonalizable. Indeed, consider the matrices

$$G(x^k) := \begin{bmatrix} 1/k & 0 \\ 0 & -1/k \end{bmatrix} \in \mathbb{S}^2 \text{ and } \Omega^k := \begin{bmatrix} k & k \\ k & k \end{bmatrix} \in \mathbb{S}^2_+.$$

Then, $\lim_{k\to\infty} \langle \Omega^k, G(x^k)\rangle = 0$ but these matrices are not simultaneously diagonalizable since $\Omega^k G(x^k) \neq G(x^k)\Omega^k$. More precisely, let us consider the orthogonal diagonalizations

$$\Omega^k = S_k \mathrm{diag}(\lambda_1^{S_k}(\Omega^k), \lambda_2^{S_k}(\Omega^k)) S_k^T = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 2k \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

and

$$G(x^k) = U_k \mathrm{diag}(\lambda_1^{U_k}(G(x^k)), \lambda_2^{U_k}(G(x^k))) U_k^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/k & 0 \\ 0 & -1/k \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Thus, no matter how the eigenvalues are ordered, we always have that $\lambda_i^{U_k}(G(x^k))\lambda_i^{S_k}(\Omega^k) \not\to 0$ for some i.

In this section, we will show that TAKKT is indeed an optimality condition and that the Augmented Lagrangian algorithm generates TAKKT sequences under the generalized Lojasiewicz inequality. In some sense, TAKKT plays the role of the Complementarity-AKKT (CAKKT) condition known for NLP [AMS10] in the global convergence analysis of NLSDP algorithms, under a very natural stopping criterion for NLSDP, which is based on the simple lemma below. More remarks about the relationship with CAKKT for NLP will follow at the end of this section.

Lemma 6 *The point $\bar{x} \in \mathbb{R}^n$ satisfies TAKKT if, and only if, there exist sequences $x^k \to \bar{x}$, $\{\Omega^k\} \subset \mathbb{S}^m_+$, and $\{\varepsilon_k\} \subset \mathbb{R}_+$ such that $x^k \to \bar{x}$, $\varepsilon_k \to 0^+$ and for all $k \in \mathbb{N}$,*

$$\|\nabla f(x^k) + DG(x^k)^*\Omega^k\| \leq \varepsilon_k, \tag{2.23}$$

$$\|[G(x^k)]_+\| \leq \varepsilon_k, \tag{2.24}$$

$$|\langle \Omega^k, G(x^k)\rangle| \leq \varepsilon_k. \tag{2.25}$$

To prove that TAKKT is an optimality condition for NLSDP, it is sufficient to note that in the proof of Theorem 2, where we proved that AKKT is an optimality condition, the dual sequence is defined as $\Omega^k := \rho_k[G(x^k)]_+$. We also note that from the definition of $\{x^k\}$ in the proof of Theorem 2 as the global solution of problem (2.17), the additional property holds:

$$f(x^k) - f(\bar{x}) + \frac{1}{2}\|x^k - \bar{x}\|^2 + \frac{\rho_k}{2} \mathrm{tr}\left(\left[G(x^k)\right]_+^2\right) \leq 0. \tag{2.26}$$

Let us see that these observations are sufficient to prove that TAKKT is an optimality condition.

Theorem 5 *Let $\bar{x}$ be a local minimizer of* (NLSDP). *Then $\bar{x}$ satisfies TAKKT.*

Proof 7 *Let $\{x^k\}$ be the AKKT sequence defined in the proof of Theorem 2. In particular, the dual sequence is defined for all k as $\Omega^k := \rho_k[G(x^k)]_+$ and (2.26) holds with some $\rho_k > 0$. It remains to prove that $\langle G(x^k), \Omega^k \rangle \to 0$. Taking the limit in (2.26) we have that*

$$\langle [G(x^k)]_+, \Omega^k \rangle = \sum_{i=1}^{m} \rho_k [\lambda_i(G(x^k))]_+^2 = \rho_k \operatorname{tr}\left(\left[G(x^k)\right]_+^2 \right) \to 0.$$

Since

$$\lambda_i(G(x^k))[\lambda_i(G(x^k))]_+ = [\lambda_i(G(x^k))]_+^2 \text{ for all } i = 1, \ldots, m,$$

it follows that $\langle G(x^k), \Omega^k \rangle = \sum_{i=1}^{m} \rho_k \lambda_i(G(x^k))[\lambda_i(G(x^k))]_+ \to 0$. Therefore, $\bar{x}$ is a TAKKT point.

Note that in Example 1, where a local minimizer that is not a KKT point is presented, the same sequence used to verify that AKKT holds can be used to check that TAKKT holds. Let us see an additional example from [Lov03, Example 6.3.6] that illustrates the same thing, but where there is a non-zero duality gap.

Example 2 (TAKKT sequence at a non-KKT solution) *Consider the following problem*

$$\begin{aligned} &\underset{x \in \mathbb{R}^2}{\text{Minimize}} && x_1, \\ &\text{Subject to} && G(x) := \begin{bmatrix} 0 & -x_1 & 0 \\ -x_1 & -x_2 & 0 \\ 0 & 0 & -1-x_1 \end{bmatrix} \preceq 0. \end{aligned}$$

The point $\bar{x} := (0,0)$ is a global minimizer, with an optimal value of zero. The dual problem can be stated as

$$\begin{aligned} &\underset{\Omega \in \mathbb{S}^3}{\text{Maximize}} && -\mu_{33}, \\ &\text{Subject to} && \Omega := \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{12} & \mu_{22} & \mu_{23} \\ \mu_{13} & \mu_{23} & \mu_{33} \end{bmatrix} \succeq 0, \\ & && \mu_{33} + 2\mu_{12} = 1, \\ & && \mu_{22} = 0, \end{aligned}$$

where the conic constraint implies $\mu_{12} = 0$ and hence the optimal value is -1.

Now, it is easy to check that the KKT condition does not hold at $\bar{x}$, but TAKKT holds with the sequence $x^k := (\frac{1}{k}, \frac{1}{k})$ and $\Omega^k := \begin{bmatrix} (k-1)^2 & \frac{k-1}{2k} & 0 \\ \frac{k-1}{2k} & \frac{1}{4k^2} & 0 \\ 0 & 0 & \frac{1}{k} \end{bmatrix} \in \mathbb{S}^3_+$.

In the following example we can see that AKKT does not imply TAKKT.

Example 3 (AKKT does not imply TAKKT) *Consider the following problem*

$$\begin{aligned} &\underset{x \in \mathbb{R}^2}{\text{Minimize}} && x_2, \\ &\text{Subject to} && G(x) := \begin{bmatrix} x_1 & 0 & 0 \\ 0 & -x_1 & 0 \\ 0 & 0 & x_1 - x_1^2 x_2 \end{bmatrix} \preceq 0. \end{aligned}$$

(i) *The feasible point* $\bar{x} := (0,1)$ *is an AKKT point. Define* $x_1^k := 1/k$, $x_2^k := 1$ *and* $\Omega^k := \begin{bmatrix} \mu_{11}^k & 0 & 0 \\ 0 & \mu_{22}^k & 0 \\ 0 & 0 & \mu_{33}^k \end{bmatrix} \in \mathbb{S}_+^3$ *with* $\mu_{11}^k := 2k, \mu_{22}^k := k^2$, *and* $\mu_{33}^k := k^2$. *Also, we have*

$$\nabla f(x^k) + DG(x^k)^* \Omega^k = (\mu_{11}^k - \mu_{22}^k + \mu_{33}^k - 2x_1^k x_2^k \mu_{33}^k, 1 - (x_1^k)^2 \mu_{33}^k) = (0,0).$$

Note that $G(x^k)$ *and* Ω^k *are simultaneously diagonalizable by the identity matrix. Since all eigenvalues of* $G(\bar{x})$ *are zero, there is nothing else to check.*

(ii) *TAKKT does not hold at* $\bar{x}$. *First, note that since* $G(x^k)$ *is diagonal we can take* Ω^k *a diagonal matrix without loss of generality. Now, suppose that there exist sequences* $x^k \to \bar{x}$ *and* $\Omega^k := \begin{bmatrix} \mu_{11}^k & 0 & 0 \\ 0 & \mu_{22}^k & 0 \\ 0 & 0 & \mu_{33}^k \end{bmatrix} \in \mathbb{S}_+^3$ *such that*

$$\nabla f(x^k) + DG(x^k)^* \Omega^k = (\mu_{11}^k - \mu_{22}^k + \mu_{33}^k - 2x_1^k x_2^k \mu_{33}^k, 1 - (x_1^k)^2 \mu_{33}^k) \to 0,$$

and

$$\langle G(x^k), \Omega^k \rangle = x_1^k(\mu_{11}^k - \mu_{22}^k + \mu_{33}^k) - (x_1^k)^2 x_2^k \mu_{33}^k = x_1^k(\mu_{11}^k - \mu_{22}^k + \mu_{33}^k - 2x_1^k x_2^k \mu_{33}^k) + (x_1^k)^2 x_2^k \mu_{33}^k \to 0.$$

Note that, since $\mu_{11}^k - \mu_{22}^k + \mu_{33}^k - 2x_1^k x_2^k \mu_{33}^k \to 0$, *we conclude that* $(x_1^k)^2 x_2^k \mu_{33}^k \to 0$. *Since* $x_2^k \to 1$, *this yields* $(x_1^k)^2 \mu_{33}^k \to 0$, *which constradicts* $1 - (x_1^k)^2 \mu_{33}^k \to 0$. *Thus, TAKKT does not hold at* $\bar{x} = (0,1)$.

Note that in the previous example, the point $\bar{x} = (0,1)$ is an AKKT point that is not a solution. Hence, following our global convergence result given by Theorem 4, we can not rule out the possibility of the augmented Lagrangian algorithm converging to this undesirable point. However, let us show that indeed the algorithm avoids the point $\bar{x}$ in this example. For this, let us prove the new global convergence result of the algorithm, that is, that it generates TAKKT sequences. We will need the following smoothness assumption, which is known as the generalized Lojasiewicz inequality [BDL07]:

Assumption 1 *Every feasible limit point* $\bar{x}$ *of a sequence* $\{x^k\}$ *generated by Algorithm 1 satisfies the generalized Lojasiewicz inequality below:*

There is $\delta > 0$ *and a function* $\varphi : B(\bar{x}, \delta) \to \mathbb{R}$, *where* $B(\bar{x}, \delta)$ *is the closed ball centered at* $\bar{x}$ *and radius* δ, *with* $\varphi(x) \to 0$ *when* $x \to \bar{x}$ *such that for all* $x \in B(\bar{x}, \delta)$,

$$|P(x) - P(\bar{x})| \leq \varphi(x) \|\nabla P(x)\|,$$

where $P(x) := \operatorname{tr}\left([G(x)]_+^2\right)$.

Note that, by Lemma 5, for all $x \in \mathbb{R}^n$ we have $\nabla P(x) = 2DG(x)^*[G(x)]_+$.

Assumption 1 is a natural extension of the smoothness assumption that is required in nonlinear programming to prove convergence of the Augmented Lagrangian method to CAKKT points [AMS10]. Note that this is an assumption about the smoothness of the function $G(\cdot)$, which holds, for instance, when $G(\cdot)$ is an analytic function (see [AMS10]). More importantly, this is not an assumption about how the function $G(\cdot)$ behaves in the conic constraint $G(x) \preceq 0$, such as in constraint qualification assumptions. We are now ready to prove the result.

Theorem 6 *Let Assumption 1 hold. Assume that* $\bar{x} \in \mathbb{R}^n$ *is a feasible limit point of a sequence* $\{x^k\}$ *generated by Algorithm 1. Then* $\bar{x}$ *is a TAKKT point.*

Proof 8 *By the proof of Theorem 4, it is sufficient to verify that* $\lim_{k\to\infty} \langle \Omega^k, G(x^k) \rangle = 0$. *Let us consider the following decomposition:*

$$\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) = S_k diag(\lambda_1^k, \ldots, \lambda_m^k) S_k^T \Rightarrow \Omega^k = S_k diag(\rho_k[\lambda_1^k]_+, \ldots, \rho_k[\lambda_m^k]_+) S_k^T, \tag{2.27}$$

and

$$G(x^k) = U_k diag(\lambda_1^{U_k}(G(x^k)), \ldots, \lambda_m^{U_k}(G(x^k)))U_k^T,$$

where, for all k, S_k and U_k are orthogonal matrices such that

$$\lambda_1^k \leq \lambda_2^k \leq \cdots \leq \lambda_m^k \text{ and } \lambda_1^{U_k}(G(x^k)) \leq \lambda_2^{U_k}(G(x^k)) \leq \cdots \leq \lambda_m^{U_k}(G(x^k)). \tag{2.28}$$

In addition, taking a subsequence if necessary, we have

$$G(\bar{x}) = U diag(\lambda_1^U(G(\bar{x})), \ldots, \lambda_m^U(G(\bar{x})))U^T,$$

where U is orthogonal and $\lambda_1^U(G(x^k)) \leq \lambda_2^U(G(\bar{x})) \leq \cdots \leq \lambda_m^U(G(\bar{x}))$ with $U_k \to U$.

(i) Assume that $\rho_k \to +\infty$. By Step 1 of the algorithm, the sequence $\{\nabla_x L_{\rho_k}(x^k, \Omega^k)\}$ is bounded. Also, as

$$\begin{aligned} \nabla_x L_{\rho_k}(x^k, \Omega^k) &= \nabla f(x^k) + DG(x^k)^*[\bar{\Omega}^k + \rho_k G(x^k)]_+ \\ &= \nabla f(x^k) + \rho_k DG(x^k)^* \left[\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right]_+, \end{aligned}$$

and since $\dfrac{\bar{\Omega}^k}{\rho_k} \to 0$, there is some $M > 0$ such that

$$\rho_k \|\nabla P(x^k)\| = \left\|\rho_k DG(x^k)^*[G(x^k)]_+\right\| \leq M$$

for all k. Taking the function φ given by Assumption 1 and using the fact that $P(\bar{x}) = 0$, we have for all k:

$$\left|\rho_k P(x^k)\right| \leq \varphi(x^k) \left\|\rho_k \nabla P(x^k)\right\| \leq M\varphi(x^k).$$

Taking the limit in k we have:

$$\lim_{k\to\infty} \rho_k P(x^k) = 0 \Rightarrow \lim_{k\to\infty} \rho_k \operatorname{tr}([G(x^k)]_+^2) = 0. \tag{2.29}$$

From (3.20), we get

$$G(x^k) = S_k diag(\lambda_1^k, \ldots, \lambda_m^k)S_k^T - \frac{\bar{\Omega}^k}{\rho_k}.$$

Now,

$$\begin{aligned} \langle \Omega^k, G(x^k)\rangle &= \left\langle \Omega^k, S_k diag(\lambda_1^k, \ldots, \lambda_m^k)S_k^T - \frac{\bar{\Omega}^k}{\rho_k}\right\rangle \\ &= \left\langle \Omega^k, S_k diag(\lambda_1^k, \ldots, \lambda_m^k)S_k^T\right\rangle - \left\langle \Omega^k, \frac{\bar{\Omega}^k}{\rho_k}\right\rangle. \end{aligned} \tag{2.30}$$

Let us show that $\langle \Omega^k, G(x^k)\rangle \to 0$ in an infinite subsequence $k \in K_1 \subset \mathbb{N}$. For this, let us show that

$$\left\langle \Omega^k, S_k diag(\lambda_1^k, \ldots, \lambda_m^k)S_k^T\right\rangle \to 0 \text{ and } \left\langle \Omega^k, \frac{\bar{\Omega}^k}{\rho_k}\right\rangle \to 0.$$

Firstly,

$$\left\langle \Omega^k, \frac{\bar{\Omega}^k}{\rho_k}\right\rangle = \left\langle \left[\bar{\Omega}^k + \rho_k G(x^k)\right]_+, \frac{\bar{\Omega}^k}{\rho_k}\right\rangle = \left\langle \left[\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right]_+, \bar{\Omega}^k\right\rangle. \tag{2.31}$$

Thus, by the boundedness of $\{\bar{\Omega}^k\}$ and since $\rho_k \to +\infty$ and $\left[G(x^k)\right]_+ \to 0$, we have that $\left[\dfrac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right]_+ \to$

0*; therefore,* $\left\langle \Omega^k, \frac{\bar{\Omega}^k}{\rho_k} \right\rangle \to 0$. *Furthermore,*

$$\begin{aligned} \left\langle \Omega^k, S_k diag(\lambda_1^k, \dots, \lambda_m^k) S_k^T \right\rangle &= \sum_{i=1}^m \rho_k [\lambda_i^k]_+ \lambda_i^k \\ &= \sum_{i=1}^m \left[\lambda_i^{S_k} \left(\bar{\Omega}^k + \rho_k G(x^k) \right) \right]_+ \lambda_i^{S_k} \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right). \end{aligned}$$

Let us now show that $\left[\lambda_i^{S_k} \left(\bar{\Omega}^k + \rho_k G(x^k) \right) \right]_+ \lambda_i^{S_k} \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right) \to 0$ *for all* $i = 1, \dots, m$. *From the order given in* (2.28)*, we can apply Lemma 3 to get*

$$\lambda_1(\bar{\Omega}^k) + \rho_k \lambda_i^{U_k}(G(x^k)) \le \lambda_i^{S_k}(\bar{\Omega}^k + \rho_k G(x^k)) \le \lambda_m(\bar{\Omega}^k) + \rho_k \lambda_i^{U_k}(G(x^k)). \tag{2.32}$$

From now on, since the diagonalizations were taken in such a way that the eigenvalues of all the matrices considered are ordered, we can omit the super-indexes in $\lambda_i^{U_k}(\cdot)$, $\lambda_i^{S_k}(\cdot)$, *and* $\lambda_i^{U}(\cdot)$. *Now, if* $\lambda_i(G(\bar{x})) < 0$, *we have from (2.32), since* $\lambda_i(G(x^k)) \to \lambda_i(G(\bar{x}))$, $\{\bar{\Omega}^k\}$ *is bounded, and* $\rho_k \to +\infty$, *that* $\lambda_i(\bar{\Omega}^k + \rho_k G(x^k)) < 0$ *for all sufficiently large k. Hence, for sufficiently large k,*

$$\left[\lambda_i \left(\bar{\Omega}^k + \rho_k G(x^k) \right) \right]_+ \lambda_i \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right) = 0.$$

Now, if $\lambda_i(G(\bar{x})) = 0$ *but* $\lambda_i(G(x^k)) \le 0$ *for an infinite set of indices* $K_1 \subset \mathbb{N}$, *we have that*

$$\begin{aligned} 0 \le \lambda_i(\Omega^k) &= \max\{0, \lambda_i(\bar{\Omega}^k + \rho_k G(x^k))\} \\ &\le \max\{0, \lambda_m(\bar{\Omega}^k) + \rho_k \lambda_i(G(x^k))\} \\ &\le \max\{0, \lambda_m(\bar{\Omega}^k)\} + \rho_k \max\{0, \lambda_i(G(x^k))\} \\ &= \lambda_m(\bar{\Omega}^k). \end{aligned} \tag{2.33}$$

Thus, for $k \in K_1$, $\{\lambda_i(\Omega^k)\}$ *is bounded. Since* $\lambda_i \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right) \to 0$, *we have*

$$\lim_{k \in K_1} \left[\lambda_i \left(\bar{\Omega}^k + \rho_k G(x^k) \right) \right]_+ \lambda_i \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right) = 0.$$

Finally, if $\lambda_i(G(\bar{x})) = 0$ *and* $\lambda_i(G(x^k)) > 0$ *for all sufficiently large k, let us multiply (2.32) by* $\lambda_i \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right) > 0$ *to arrive at*

$$\lambda_1(\bar{\Omega}^k) \lambda_i \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right) + \rho_k \lambda_i(G(x^k)) \lambda_i \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right) \le \lambda_i(\bar{\Omega}^k + \rho_k G(x^k)) \lambda_i \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right) \tag{2.34}$$

and

$$\lambda_i(\bar{\Omega}^k + \rho_k G(x^k)) \lambda_i \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right) \le \lambda_m(\bar{\Omega}^k) \lambda_i \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right) + \rho_k \lambda_i(G(x^k)) \lambda_i \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right). \tag{2.35}$$

Note that

$$\lambda_1(\bar{\Omega}^k) \lambda_i \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right) \to 0 \text{ and } \lambda_m(\bar{\Omega}^k) \lambda_i \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right) \to 0,$$

then, it remains to show that

$$\rho_k \lambda_i(G(x^k)) \lambda_i \left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \right) = \lambda_i(G(x^k)) \lambda_i \left(\bar{\Omega}^k + \rho_k G(x^k) \right) \to 0.$$

Since $\lambda_i(G(x^k)) \to 0$, $\{\bar{\Omega}^k\}$ is bounded, and $\rho_k \lambda_i(G(x^k))^2 \to 0$ (by (2.29)), this follows by multiplying (2.32) by $\lambda_i(G(x^k))$:

$$\lambda_1(\bar{\Omega}^k)\lambda_i(G(x^k)) + \rho_k \lambda_i(G(x^k))^2 \leq \lambda_i(\bar{\Omega}^k + \rho_k G(x^k))\lambda_i(G(x^k)) \leq \lambda_m(\bar{\Omega}^k)\lambda_i(G(x^k)) + \rho_k \lambda_i(G(x^k))^2. \tag{2.36}$$

Thus, we conclude that $\lim_{k\to\infty} \langle \Omega^k, G(x^k)\rangle = 0$.

(ii) If $\{\rho_k\}$ is bounded, the proof follows as in the proof of Theorem 4, item (ii), by noting that

$$\begin{aligned}\lim_{k\to\infty} \langle \Omega^k, G(x^k)\rangle &= \sum_{i=1}^{m} \lambda_i^U(\bar{\Omega})\lambda_i^U(G(\bar{x})) \\ &= \sum_{i=1}^{m} \frac{1}{\rho_{k_0}} [\lambda_i^U]_+ (\lambda_i^U - [\lambda_i^U]_+) = 0.\end{aligned}$$

To end this section, we discuss about the relevance of TAKKT in the context of sequential optimality conditions for NLP. For this, let us consider $G(x) := \operatorname{diag}(g_1(x), \ldots, g_m(x))$ and $\Omega^k := \operatorname{diag}(\mu_1^k, \ldots, \mu_m^k) \succeq 0$. Thus, a feasible point $\bar{x}$ satisfies TAKKT if, and only if,

$$\lim_{k\to\infty} \nabla f(x^k) + DG(x^k)^* \Omega^k = \lim_{k\to\infty} \nabla f(x^k) + \sum_{i=1}^{m} \mu_i^k \nabla g_i(x^k) = 0, \tag{2.37}$$

$$\lim_{k\to\infty} \langle \Omega^k, G(x^k)\rangle = \lim_{k\to\infty} \sum_{i=1}^{m} \mu_i^k g_i(x^k) = 0, \tag{2.38}$$

while for verifying AKKT, condition (2.38) is replaced by

$$g_i(\bar{x}) < 0 \Rightarrow \mu_i^k = 0, \text{ for all sufficiently large } k \text{ and } i = 1, \ldots, m,$$

which, for the sake of this discussion, we may consider its equivalent form

$$g_i(\bar{x}) < 0 \Rightarrow \mu_i^k \to 0^+, \text{ for all } i = 1, \ldots, m. \tag{2.39}$$

That is, the AKKT condition (2.37) + (2.39) says that the dual sequence $\{\mu^k\}$ may be unbounded, but the corresponding gradient of the Lagrangian function should vanish in the limit. Also, the gradient of an inactive constraint at the limit should not be considered, eventually, in the computation of the gradient of the Lagrangian. AKKT is typically considered as the most natural (and weakest) sequential optimality condition, as these properties are the least one would expect to hold at KKT-related conditions.

The complementarity condition imposed by TAKKT, namely, condition (2.38), has not been considered before in the context of NLP, as it is not natural in this context. The issue is that it does not immediately imply that $\mu_i^k \to 0^+$ when $g_i(\bar{x}) < 0$. Namely, (2.38) may not be adequate to assert complementarity in the limit, as one could satisfy it, for instance, with an inactive constraint at the limit but with a corresponding Lagrange multiplier sequence bounded away from zero. For instance, with constraints $(g_1(x^k), g_2(x^k)) := (-1 + \frac{1}{k}, \frac{1}{k}) \to (-1, 0)$ and $(\mu_1^k, \mu_2^k) := (1, k)$. That is, TAKKT may not correctly identify the active constraints in the verification of (2.37). Surprisingly, as we will show later, under MFCQ this issue is not present, as a TAKKT point is in fact a KKT point. The notion of a TAKKT point for NLP is closely related to the so-called Sum Converging to Zero (SCZ) concept defined in [GJN16]; however, a thorough investigation of the relevance of TAKKT in NLP, together with its relations with other sequential optimality conditions, would be in order, but out of the scope of this paper. For example, even for NLP, although we conjecture that TAKKT would not imply AKKT, we were not able to find an example. That is, it is not clear if the concern raised above can actually happen.

Another known way to measure complementarity, in the case of NLP, is the so-called Complementarity-Approximate-KKT (CAKKT) condition defined in [AMS10], where (2.38) is replaced by the complementarity measurement:

$$\lim_{k\to\infty} g_i(x^k)\mu_i^k = 0, \text{ for all } i = 1,\dots,m. \tag{2.40}$$

Clearly, (2.40) implies (2.38) and (2.39), hence CAKKT is stronger than TAKKT and AKKT for NLP. Note that since these are necessary optimality conditions, a stronger condition is more desirable than a weaker one. For inactive constraints, (2.40) and (2.39) impose the same condition on the multiplier, namely, that it should converge to zero. The difference is with respect to an active constraint $g_i(\bar{x}) = 0$: while AKKT allows the corresponding Lagrange multiplier μ_i^k to diverge to infinity arbitrarily fast, CAKKT imposes a bound on its rate of growth, namely, it should go to infinity slower than $\frac{1}{|g_i(x^k)|}$ (see [Hae18] for further discussions under this point of view). The CAKKT condition is natural in the context of interior point methods [Hae18] and it is also generated by augmented Lagrangian methods under the generalized Lojasiewicz inequality [AMS10]. In [AMS10] it was also shown that CAKKT is able to detect non-optimality in some cases where AKKT does not detect it.

Hence, another interesting endeavour would be to define an extension of CAKKT to NLSDP. Similarly to what was done for AKKT, such an extension would need to take into account the "correspondence" among the eigenvalues of Ω^k and $G(x^k)$, in such a way that the condition would reduce to CAKKT for NLP in the diagonal case. A first idea would be to define CAKKT for NLSDP by replacing the complementarity measurement in AKKT by the stronger one

$$\lim_{k\to\infty} \lambda_i^{U_k}(G(x^k))\lambda_i^{S_k}(\Omega^k) = 0, \text{ for all } i = 1,\dots,m,$$

where U_k and S_k are orthogonal matrices that diagonalize $G(x^k)$ and Ω^k, respectively, such that $S_k \to U$ and $U_k \to U$. Note that in the definition of AKKT we asked that $S_k \to U$ to obtain a correspondence between an inactive eigenvalue of the constraint matrix with an eigenvalue of the Lagrange multiplier matrix to vanish. In addition, for CAKKT we would ask that $U_k \to U$ to ensure that $\lambda_i^{U_k}(G(x^k)) \to \lambda_i^{U}(G(\bar{x}))$ for all $i = 1,\dots,m$.

Even though CAKKT defined in this way would be an optimality condition, we were not able to prove that the augmented Lagrangian algorithm generates this type of sequences under the generalized Lojasiewicz inequality. The crucial difficulty is that, as in the proof that the algorithm generates TAKKT sequences, it is important to take orthogonal diagonalizations of $G(x^k)$ and Ω^k in such a way that the eigenvalues are ordered. This is not an issue in TAKKT since it does not require any correspondence among the eigenvalues; however, the correspondence imposed by the definition of CAKKT does not allow us to arbitrarily order the eigenvalues of $G(x^k)$ and Ω^k.

In the next section we will show that our global convergence results to AKKT or TAKKT points are strictly stronger than standard global convergence results to a KKT point under MFCQ. In particular, our results do not require boundedness of Lagrange multipliers in order to assert that a feasible limit point of the algorithm is a KKT point.

2.7 Strength of the sequential optimality conditions

At this point, we have not yet mentioned that the optimality conditions that we have defined are strong in any sense. Namely, they could be satisfied at any feasible point, which would make our results up to now meaningless. Let us show that this is not the case. We start by discussing the strength of AKKT.

First, let us show that the global convergence results that we have proved are at least as good as standard global convergence results to a KKT point under MFCQ.

Theorem 7 *Let $\bar{x} \in \mathscr{F}$ be a feasible point that satisfies MFCQ. Then, for any objective function f in (NLSDP) such that $\bar{x}$ satisfies AKKT, $\bar{x}$ satisfies in addition the KKT conditions for this problem.*

Proof 9 *From the definition of AKKT, there exist sequences $\{x^k\} \subset \mathbb{R}^n, x^k \to \bar{x}$ and $\{\Omega^k\} \subset \mathbb{S}^m_+$, together with orthogonal matrices S_k and U that diagonalize Ω^k for each k and $G(\bar{x})$, respectively, with $S_k \to U$, such that*

$$\lim_{k\to\infty} \nabla f(x^k) + DG(x^k)^*\Omega^k = 0, \tag{2.41}$$

and

$$\lambda_i^U(G(\bar{x})) < 0 \Rightarrow \lambda_i^{S_k}(\Omega^k) = 0. \tag{2.42}$$

If $\{\Omega^k\}$ *is contained in a compact set, there is* $K \subset \mathbb{N}$, $\Omega \in \mathbb{S}^m_+$, *such that* $\lim_{k\in K} \Omega^k = \Omega$. *Let us consider the partition of* $\{1,\dots,m\}$ *into the index sets:*

$$I_1 := \{i \mid \lambda_i^U(G(\bar{x})) < 0\} \text{ and } I_2 := \{i \mid \lambda_i^U(G(\bar{x})) = 0\}.$$

From (2.41), we have that

$$\nabla f(\bar{x}) + DG(\bar{x})^*\Omega = 0,$$

and by noting that $G(\bar{x})$ *and* Ω *are simultaneously diagonalizable, we have that*

$$\lim_{k\to\infty} \langle \Omega^k, G(x^k)\rangle = \langle \Omega, G(\bar{x})\rangle = \sum_{i\in I_1} \lambda_i^U(\Omega)\lambda_i^U(G(\bar{x})) + \sum_{i\in I_2} \lambda_i^U(\Omega)\lambda_i^U(G(\bar{x})) = 0,$$

where the previous equality is due to the fact that $\lambda_i^U(\Omega) = 0$ *for* $i \in I_1$ *and* $\lambda_i^U(G(\bar{x})) = 0$ *for* $i \in I_2$. *Therefore,* $\bar{x}$ *satisfies the KKT conditions.*

Now, let us consider a subsequence such that $t_k := \|\Omega^k\| \to \infty$. *Thus, let us take* $K_1 \subset \mathbb{N}$ *such that* $\lim_{k\in K_1} \frac{\Omega^k}{t_k} = \bar{\Omega} \neq 0$ *for some* $\bar{\Omega} \in \mathbb{S}^m_+$ *simultaneously diagonalizable with* $G(\bar{x})$. *Then, from (2.41) and (2.42) we have that*

$$\lim_{k\in K_1} \frac{\nabla f(x^k)}{t_k} + DG(x^k)^* \frac{\Omega^k}{t_k} = DG(\bar{x})^*\bar{\Omega} = 0, \text{ and} \tag{2.43}$$

$$\lim_{k\in K_1} \left\langle \frac{\Omega^k}{t_k}, G(x^k)\right\rangle = \langle \bar{\Omega}, G(\bar{x})\rangle = \sum_{i\in I_1} \lambda_i^U(\bar{\Omega})\lambda_i^U(G(\bar{x})) + \sum_{i\in I_2} \lambda_i^U(\bar{\Omega})\lambda_i^U(G(\bar{x})) = 0. \tag{2.44}$$

To see that (2.43)-(2.44) *contradicts MFCQ, let* $d \in \mathbb{R}^n$ *be such that* $G(\bar{x}) + DG(\bar{x})d$ *is negative definite. Thus,*

$$\begin{aligned} 0 &= \langle \bar{\Omega}, G(\bar{x})\rangle + \langle DG(\bar{x})^*\bar{\Omega}, d\rangle \\ &= \langle \bar{\Omega}, G(\bar{x})\rangle + \langle \bar{\Omega}, DG(\bar{x})d\rangle \\ &= \langle \bar{\Omega}, G(\bar{x}) + DG(\bar{x})d\rangle, \end{aligned}$$

which implies by Lemma 2 that $\bar{\Omega} = 0$, *contradicting the definition of* $\bar{\Omega}$.

A result similar to Theorem 7 holds for TAKKT. The proof is very similar to the proof of Theorem 7 and is omitted.

Theorem 8 *Let* $\bar{x} \in \mathscr{F}$ *be a feasible point that satisfies MFCQ. Then, for any objective function* f *in* (NLSDP) *such that* $\bar{x}$ *satisfies TAKKT,* $\bar{x}$ *satisfies in addition the KKT conditions for this problem.*

One may view Theorem 7 as strenghtening the simple fact that MFCQ is a constraint qualification, that is, MFCQ implies that the KKT condition holds not only at local minimizers, but also at AKKT points (which includes local minimizers, as shown in Theorem 2). This has been called a *strict* constraint qualification in [AMRS18]. We use Theorem 7 to inspire the definition of a new constraint qualification: note that the property satisfied by points fulfilling MFCQ, given by Theorem 7, is a property of the feasible set of (NLSDP), independently of the objective function. Also, since AKKT is an optimality condition, it is clear that this property is a constraint qualification (weaker than MFCQ). The same holds true with respect to the property stated in Theorem 8. Let us make these definitions precise.

Definition 6 (AKKT/TAKKT-regularity) *We say that a feasible point* $\bar{x} \in \mathscr{F}$ *satisfies AKKT-regularity (TAKKT-regularity) if for any objective function* f *in* (NLSDP) *such that* $\bar{x}$ *satisfies AKKT (TAKKT, respectively),* $\bar{x}$ *satisfies in addition the KKT conditions for this problem.*

Thus, the global convergence result presented in Theorem 4 implies that under AKKT-regularity, feasible limit points of the algorithm are KKT points. Analogously, Theorem 6 shows the global convergence of the algorithm to a KKT point under TAKKT-regularity. This gives a more standard constraint qualification formulation of our global convergence results, which can be compared with other ones in the literature.

Let us now show that our global convergence results are strictly stronger than the more standard one based on MFCQ. We do this by showing that neither AKKT-regularity nor TAKKT-regularity imply MFCQ (note that Theorems 7 and 8 show that AKKT/TAKKT-regularity are weaker than MFCQ). The following example serve this purpose:

Example 4 (AKKT-regularity and TAKKT-regularity do not imply MFCQ) *Consider $\bar{x} := 0$ and the feasible set defined by*

$$G(x) := \begin{bmatrix} x & 0 \\ 0 & -x \end{bmatrix} \preceq 0.$$

(i) ($\bar{x}$ does not satisfies MFCQ.) Indeed,

$$G(\bar{x}) + DG(\bar{x})h = \begin{bmatrix} h & 0 \\ 0 & -h \end{bmatrix}.$$

Therefore, there is no $h \in \mathbb{R}$ such that $G(\bar{x}) + DG(\bar{x})h$ is negative definite.

(ii) ($\bar{x}$ satisfies AKKT-regularity and TAKKT-regularity.) This is trivially true as, independently of the objective function, $\bar{x} = 0$ is a KKT point.

The previous example is very simple, as it reduces trivially to a nonlinear programming problem. However one could arrive at exactly the same conclusions by considering a feasible set defined by $G(x_1, x_2) := \begin{bmatrix} x_1 & x_2 \\ x_2 & -x_1 \end{bmatrix} \preceq 0$.

Let us note that by revisiting the constraints in Example 3 in more details, that is, without a fixed objective function, we can see that it actually shows that TAKKT-regularity does not imply AKKT-regularity. Note also that in Example 1, AKKT-regularity does not hold, given that AKKT holds but KKT does not hold. This shows that AKKT-regularity may fail for linear constraints (which does not occur for AKKT-regularity in NLP). Similarly, Example 2 shows that TAKKT-regularity may fail for linear constraints.

Similarly to [AMRS18], we can provide a geometric interpretation of AKKT/TAKKT-regularity by means of the outer semicontinuity of a point-to-set mapping. We present this interpretation for TAKKT-regularity, but a similar definition can be made for AKKT-regularity.

Theorem 9 *A feasible point $\bar{x}$ satisfies TAKKT-regularity if, and only if, the following point-to-set mapping is outer semicontinuous at $(\bar{x}, 0)$:*

$$K^{\text{TAKKT}}(x, r) := \{DG(x)^*\Omega \;:\; |\langle \Omega, G(x)\rangle| \leq r,\ \Omega \in \mathbb{S}^m_+\},$$

that is, $\limsup_{(x,r)\to(\bar{x},0)} K^{\text{TAKKT}}(x, r) \subset K^{\text{TAKKT}}(\bar{x}, 0)$.

Proof 10 *Let f be a smooth objective function such that TAKKT holds at $\bar{x}$. Then, by the definition, there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\Omega^k\} \in \mathbb{S}^m_+$, $\{\varepsilon_k\} \subset \mathbb{R}^n$, and $\{r^k\} \subset \mathbb{R}_+$ such that*

$$\lim_{k\to\infty} x^k = \bar{x},\ \varepsilon_k := \nabla f(x^k) + DG(x^k)^*\Omega^k \to 0,\ \text{and } r^k := |\langle \Omega^k, G(x^k)\rangle| \to 0.$$

Define $w^k := DG(x^k)^*\Omega^k$. Clearly, the sequence w^k satisfies

$$w^k \in K^{\text{TAKKT}}(x^k, r^k) \text{ and } w^k \to -\nabla f(\bar{x}).$$

From the definition of lim sup *for the point-to-set mapping K^{TAKKT}, we have*

$$-\nabla f(\bar{x}) \in \limsup_{(x,r)\to(\bar{x},0)} K^{\text{TAKKT}}(x, r) \subset K^{\text{TAKKT}}(\bar{x}, 0),$$

which, by the definition of $K^{\text{TAKKT}}(\bar{x},0)$, implies that the KKT conditions hold. This proves that $\bar{x}$ satisfies TAKKT-regularity. Now, let us assume TAKKT-regularity at $\bar{x}$ and let us prove that

$$\limsup_{(x,r)\to(\bar{x},0)} K^{\text{TAKKT}}(x,r) \subset K^{\text{TAKKT}}(\bar{x},0).$$

Take $\overline{w} \in \limsup_{(x,r)\to(\bar{x},0)} K^{\text{TAKKT}}(x,r)$. So, there are sequences $\{x^k\}$, $\{w^k\}$ and $\{r^k\}$ such that $x^k \to \bar{x}$, $w^k \to \overline{w}$, $r^k \to 0$ and $w^k \in K^{\text{TAKKT}}(x^k,r^k)$ for all k. Now, let us define the linear function $f(x) := -\langle \overline{w}, x\rangle$ for all $x \in \mathbb{R}^n$. Let us see that TAKKT holds at $\bar{x}$ with this choice of f. Since $w^k \in K^{\text{TAKKT}}(x^k,r^k)$, there is a multiplier $\Omega^k \in \mathbb{S}^m_+$ such that

$$w^k = DG(x^k)^*\Omega^k \text{ and } |\langle \Omega^k, G(x^k)\rangle| \leq r^k.$$

Since $r^k \to 0$, we have $\lim_{k\to\infty} \langle \Omega^k, G(x^k)\rangle = 0$. Also, since $w^k \to \overline{w}$, we have $\varepsilon_k := \nabla f(x^k) + DG(x^k)^*\Omega^k = -\overline{w} + w^k \to 0$. Thus, TAKKT holds at $\bar{x}$, which implies by TAKKT-regularity that the objective function defined is such that the KKT conditions hold at $\bar{x}$. This can be written as $-\nabla f(\bar{x}) = \overline{w} \in K^{\text{TAKKT}}(\bar{x},0)$, which concludes the proof.

For NLP, several constraint qualifications strictly weaker than MFCQ that still imply AKKT-regularity have been proposed (such as CPLD [QW00], RCPLD [AHSS12c], CRSC [AHSS12b] and CPG [AHSS12b] – or CRCQ [Jan84] and RCRCQ [AHSS12a], which are independent of MFCQ). These weak constraint qualifications help in ensuring AKKT-regularity in a more tractable way by means of properties of the derivatives of the constraints in a neighborhood of the point of interest. We expect that extensions of these concepts to NLSDP would be relevant in characterizing global convergence of algorithms for NLSDP, while also showing that AKKT-regularity is much weaker than MFCQ, way beyond what our simple Example 4 suggests.

2.8 Conclusions

In the past ten years, sequential optimality conditions have played an increasing role in global convergence analysis of algorithms for nonlinear programming problems. Without a constraint qualification, the fact that the KKT conditions hold approximately at any local minimizer justifies the numerical practice of not verifying constraint qualifications at all when deciding when to stop the execution of an algorithm.

It can be conjectured that the stronger the theoretical properties of limit points of a sequence generated by an algorithm, the better the algorithm will behave in practice. This indicates the practical relevance of proving global convergence of algorithms under weak assumptions.

In this paper, we extended the NLP concept of Approximate-KKT point to nonlinear semidefinite programming. This extension is not straightforward as the KKT conditions for NLSDP require that the constraint matrix and the Lagrange multiplier matrix to be simultaneously diagonalizable; hence, it is not straightforward to define the notion of “satisfying the KKT conditions approximately”. Surprisingly, we were able to define a sequential optimality condition, new even in the context of NLP, that is naturally presented in NLSDP, as it does not rely on eigenvalue computations, which we call Trace-Approximate-KKT.

These tools allowed us to define an Augmented Lagrangian algorithm whose global convergence analysis can be done without relying on the Mangasarian-Fromovitz constraint qualification, that is, a limit point of a sequence generated by an algorithm may be proven to be a KKT point even though this point may have an unbounded Lagrange multiplier set.

Further research is needed to access the practical performance of the algorithm. We also envision that the theory presented can be generalized by considering an algorithm that solves constrained augmented Lagrangian subproblems, or by solving the subproblems up to second-order. Also, we expect other sequential optimality conditions for NLP to be generalized to NLSDP, such as the Approximate-Gradient-Projection [MS03], which can be relevant for algorithms that do not produce a Lagrange multiplier approximation.

Capítulo 3

Optimality conditions for nonlinear symmetric cone programming

Neste capítulo vamos apresentar o artigo: *Optimality conditions for nonlinear symmetric cone programming*. O objetivo deste artigo é estender o conceito de Aproximadamente KKT para cones simétricos quaisquer. No capítulo anterior apresentamos a definição de AKKT para NSDP onde a condição de complementariedade é exigida em termos dos autovalores das matrizes envolvidas. Além disso, uma relação entre as matrizes ortogonais que efetuam as diagonalizações foi exigida. Em programação não linear sob cones simétricos (NSCP)[1] a ideia é similar a adotada em NSDP. Faraut e Konráyi provaram em [FK94] a existência de uma decomposição espectral para cones simétricos através da teoria de álgebras de Jordan Euclidianas. Esses resultados nos permitirão realizar a extensão de AKKT para NSCP.

Embora NSCP tenha esse papel unificador, vamos apresentar em mais detalhes como definir AKKT para o problema de otimização sob cones de segunda-ordem (NSOCP)[2], uma vez que para NSOCP várias condições de qualificação foram estendidas da programação não linear como CRCQ, RCRCQ e CRSC.

3.1 Article: Optimality conditions for nonlinear symmetric cone programming

Nonlinear symmetric cone programming (NSCP) generalizes important optimization problems such as nonlinear programming (NLP), nonlinear semidefinite programming (NSDP) and nonlinear second-order programming (NSOCP). In this work, we present a new optimality condition for NSCP based on an extension of the so-called Approximate KKT condition (AKKT), which provides a KKT point under a condition weaker than Robinson's constraint qualification. In addition, we prove that an augmented Lagrangian method proposed for NSOCPs generates AKKT sequences and, under a weak constraint qualification, we prove the global convergence of the algorithm to a KKT point. In particular, due to the particular symmetric cone structure of NSOCP, we were able to exploit extensions of constant rank constraint qualifications in our global convergence analysis.

3.2 Introduction

The *nonlinear symmetric cone programming* (NSCP) is an optimization problem where the constraints are given by a relation between a function and a general symmetric cone. In recent years, the interest in NSCP has grown considerably. The reason is that many well-known optimization problems are particular cases of NSCP, including the *nonlinear programming* (NLP), the *nonlinear second-order cone programming* (NSOCP) and *nonlinear semidefinite programming* (NSDP) problems. The theoretical tool that allows us to unify the study of all these problems is the notion of an Euclidean Jordan algebra. Faraut and Korányi studied in [FK94] the concept of cone of squares and proved that every symmetric cone can be represented

[1]Do inglês: nonlinear symmetric cone programming

[2]Do inglês: nonlinear second-order cone programming

as a cone of squares of some Euclidean Jordan algebra. Although the interest in optimization problems with conic constraints is increasing, the development is still very preliminary.

For the NLP case, sequential optimality conditions have been considered very useful in the past years due to the possibility of unifying and extending results of global convergence for several algorithms [AHM11]. These necessary optimality conditions appeared in NLP as an alternative for conditions of the form "KKT or not-CQ", which means that every local optimizer is either a Karush-Kuhn-Tucker point or it does not satisfy a *constraint qualification* (CQ). In fact, sequential optimality conditions are necessary for optimality without the need of requiring a constraint qualification, besides, they imply a condition of the form "KKT or not-CQ" under a so-called *strict* constraint qualification [AMRS18] in place of "CQ", which are stronger than the ones appearing in more theoretical venues but, more interestingly, are weaker than CQs usually employed in global convergence analysis (such as Mangasarian-Fromovitz CQ/Robinson's CQ). Due to their relevance, sequential optimality conditions were extended to different classes of optimization problems such as multiobjective optimization [GJN16], NSDP [AHV18], NLP with complementarity constraints [AHSS18, Ram16], generalized Nash equilibrium problems [BHR19], optimization of discontinuous functions [BKM17], optimization in Banach Spaces [KSW18], variational and quasi-variational inequalities [HS11, KS18], quasi-equilibrium problems [BHFR18] and others.

One of the most relevant sequential optimality conditions is the so-called Approximate-Karush-Kuhn-Tucker (AKKT) condition, which is related to several first and second-order algorithms. For more details see [AHRS17, AMRS16, AMRS18, AMS16, BM14]. In this paper, we consider an extension of the AKKT condition for NSCP problems. The particular case of NSOCP is studied in details, where an augmented Lagrangian algorithm is proposed and constant rank constraint qualifications are exploited. In particular, we deal with the constraint qualifications introduced in [ZZ18] where several constant rank constraint qualifications have been extended to NSOCP, namely, the constant rank constraint qualification (CRCQ), the relaxed constant rank constraint qualification (RCRCQ) and the constant rank of the subspace component (CRSC).

This paper is organized as follows. In Section 3.3, we recall some theoretical results on Euclidean Jordan algebras. The sequential optimality conditions for NSCP is introduced in Section 3. In Section 3.4, we analyse in details the case of NSOCP. In Section 3.5, we introduce the augmented Lagrangian algorithm for NSOCP and show that this method generates sequences that satisfy our sequential optimality condition. Finaly, in Section 3.6, some conclusions are presented.

The following notations will be adopted in this paper. For any matrix $A \in \mathbb{R}^{n\times\ell}$, its transpose is denoted by $A^T \in \mathbb{R}^{\ell\times n}$. A vector $z \in \mathbb{R}^\ell$ can be written as $(z_0, \bar{z}) \in \mathbb{R}\times\mathbb{R}^{\ell-1}$. Similarly, a function $h\colon \mathbb{R}^n \to \mathbb{R}^\ell$ with components $h_i\colon \mathbb{R}^n \to \mathbb{R}$ can be written as $h(x) = \left([h(x)]_0, \overline{h(x)}\right) \in \mathbb{R}\times\mathbb{R}^{\ell-1}$. The Jacobian of h at $x \in \mathbb{R}^n$ is denoted by $Jh(x) \in \mathbb{R}^{\ell\times n}$ and the gradient of h_i at x by $\nabla h_i(x)$. The Euclidean norm and the Euclidean inner product will be denoted by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$, respectively. Consider $\mathscr{J}$ a finite dimensional inner-product space and an operator $P\colon \mathbb{R}^n \to \mathscr{J}$. The adjoint of P with repect to the Euclidean inner product in $\mathbb{R}^n$ is given by $P^*\colon \mathscr{J} \to \mathbb{R}^n$. Let C be a cone, the notation $y \succeq_C 0$ means that y belongs to the cone C, while $\operatorname{int}(C)$ denotes the interior of C. The matrix I_ℓ is the $\ell\times\ell$ identity matrix. The ℓ-dimensional second-order cone is given by

$$K_\ell := \left\{ z = (z_0, \bar{z}) \in \mathbb{R}\times\mathbb{R}^{\ell-1} \;\middle|\; \begin{array}{ll} \|\bar{z}\| \le z_0 & \text{for } \ell \ge 2 \\ z \ge 0 & \text{for } \ell = 1 \end{array} \right\}.$$

3.3 Symmetric cones and Euclidean Jordan algebras

In this section, we review some results on symmetric cones and Euclidean Jordan algebras. A more complete approach showing details of how Euclidean Jordan algebras and symmetric cones relate can be seen in the book by Faraut and Korányi [FK94] and [Bae07]. Let $\mathscr{E}$ be an n-dimensional real vector space with an inner product $\langle\cdot,\cdot\rangle$. In order to define a symmetric cone, we need the concept of dual cone and homogeneous cone. The dual of a cone $\mathscr{K} \subseteq \mathscr{E}$ is defined by $\mathscr{K}^* := \{u \in \mathscr{E} \mid \langle u, v\rangle \ge 0 \text{ for all } v \in \mathscr{K}\}$, and $\mathscr{K}$ is self-dual if $\mathscr{K} = \mathscr{K}^*$. Furthermore, the cone $\mathscr{K}$ is *homogeneous* if for each points u and v in the interior of $\mathscr{K}$ there exists a linear bijection T such that $T(u) = v$ and $T(\mathscr{K}) = \mathscr{K}$.

Definition 7 *The cone $\mathcal{K}$ is symmetric if it has nonempty interior, is self-dual and homogeneous.*

Some well-known optimization problems are defined on symmetric cones, as in the case of: the nonlinear programming on the nonnegative orthant, the nonlinear second-order cone programming on the second-order cone (or Lorentz cone), the nonlinear semidefinite programming on the positive semidefinite cone and others. The concept of an Euclidean Jordan algebra allows us to unify the study of NLP, NSOCP, NSDP, and other problems obtained through a mix of different conic constraints by means of a symmetric conic constraint. The definition is as follows

Definition 8 *Let $\mathcal{E}$ be an n-dimensional real vector space with an inner product $\langle \cdot, \cdot \rangle$ and a bilinear operator $\circ : \mathcal{E} \times \mathcal{E} \to \mathcal{E}$. Then, we say that $(\mathcal{E}, \circ)$ is an* Euclidean Jordan algebra *if for all $u, v, w \in \mathcal{E}$:*

(i) $u \circ v = v \circ u$,

(ii) $u \circ (u^2 \circ v) = u^2 \circ (u \circ v)$ *where* $u^2 = u \circ u$,

(iii) $\langle v \circ w, u \rangle = \langle v, w \circ u \rangle$.

It is well-known that an Euclidean Jordan algebra $\mathcal{E}$ can be decomposed as a direct sum of *simple* Euclidean Jordan algebras

$$\mathcal{E} = \mathcal{E}_1 \oplus \mathcal{E}_2 \oplus \cdots \oplus \mathcal{E}_r,$$

which induces a decomposition of a symmetric cone $\mathcal{K}$ as a direct sum of symmetric cones, i.e.,

$$\mathcal{K} = \mathcal{K}_1 \oplus \mathcal{K}_2 \oplus \cdots \oplus \mathcal{K}_r,$$

where each algebra $(\mathcal{E}_i, \circ)$ is such that it is not possible to obtain nonzero sub-algebras U_i, V_i of $\mathcal{E}_i$ such that $\mathcal{E}_i = U_i \oplus V_i$. Let us present a characterization of symmetric cones in terms of a cone of squares.

Theorem 10 *([FK94, Theorem III.2.1]). Let $\mathcal{E}$ be a finite dimensional inner-product space. A cone $\mathcal{K} \subseteq \mathcal{E}$ is symmetric if and only if $\mathcal{K}$ is a cone of squares of some Euclidean Jordan algebra $(\mathcal{E}, \circ)$, i.e., $\mathcal{K} := \{u \circ u \mid u \in \mathcal{E}\}$.*

With the above theorem, we can use Euclidean Jordan algebra theory to study properties of symmetric cones. In particular, given a symmetric cone $\mathcal{K}$ we will consider its natural Euclidean Jordan algebra such that $\mathcal{K}$ is its cone of squares. For example, we can characterize the second-order cone K_ℓ as a cone of squares defining the operator

$$(z_0, \bar{z}) \circ (y_0, \bar{y}) = (\langle z, y \rangle, z_0 \bar{y} + y_0 \bar{z}),$$

for all $z = (z_0, \bar{z}) \in \mathbb{R} \times \mathbb{R}^{\ell-1}$ and $y = (y_0, \bar{y}) \in \mathbb{R} \times \mathbb{R}^{\ell-1}$. Moreover, to characterize the positive semidefinite cone $\mathbb{S}^m_+$ as a cone of square, we define $X \circ Y = \frac{YX+XY}{2}$ for $X, Y \in \mathbb{S}^m$. An important concept in Euclidean Jordan algebras is the notion of a spectral decomposition, which is a natural extension from the usual decomposition for symmetric matrices.

Theorem 11 *([FK94, Theorem III.1.2]). Let $(\mathcal{E}, \circ)$ be an Euclidean Jordan algebra and $u \in \mathcal{E}$. Then there exist so-called idempotents $c_i(u)$, $i = 1, \ldots, r$ satisfying*

$$c_i(u) \circ c_j(u) = 0, \qquad i \neq j, \tag{3.1}$$

$$c_i(u) \circ c_i(u) = c_i(u), \qquad i = 1, \ldots, r, \tag{3.2}$$

$$c_1(u) + \cdots + c_r(u) = \boldsymbol{e}, \qquad i = 1, \ldots, r, \tag{3.3}$$

and so-called eigenvalues $\lambda_i(u) \in \mathbb{R}$ with $i = 1, \ldots, r$ such that

$$u = \lambda_1(u) c_1(u) + \lambda_2(u) c_2(u) + \cdots + \lambda_r(u) c_r(u), \tag{3.4}$$

where $\boldsymbol{e}$ satisfies $u \circ \boldsymbol{e} = \boldsymbol{e} \circ u = u$ for all u.

The element $\mathbf{e}$ in the above theorem is unique and called identity. We also say that $c_i(u)$, $i = 1,\dots,r$, in the previous theorem form a Jordan frame for u. We can see in Proposition II.2.1 of [FK94] that the functions λ_i are all continuous over $\mathscr{E}$ and uniquely determined by u.

The elements $u \in \mathscr{E}$ and $v \in \mathscr{E}$ are said to *operator commute* if they share a Jordan frame, that is, if there exists a common Jordan frame $\{c_1,\dots,c_r\}$ such that

$$u = \lambda_1(u)c_1 + \cdots + \lambda_r(u)c_r \quad \text{and} \quad v = \lambda_1(v)c_1 + \cdots + \lambda_r(v)c_r.$$

The number of nonzero eigenvalues of $u \in \mathscr{E}$ is an invariant called the *rank* of u. For a symmetric matrix, its spectral decomposition coincide with the classic eigenvalue decomposition. For the particular case of the Euclidean Jordan algebra associated with the second order-cone K_ℓ, the spectral decomposition of $z \in \mathbb{R}^\ell$ is given by $z = \lambda_1(z)c_1(z) + \lambda_2(z)c_2(z)$ where $\lambda_1(z) = z_0 - \|\bar{z}\|$, $\lambda_2(z) = z_0 + \|\bar{z}\|$ and

$$c_1(z) = \begin{cases} \frac{1}{2}\left(1, -\frac{\bar{z}}{\|\bar{z}\|}\right) & \text{if } \bar{z} \neq 0, \\ \frac{1}{2}(1, -\overline{w}) & \text{if } \bar{z} = 0, \end{cases} \qquad c_2(z) = \begin{cases} \frac{1}{2}\left(1, \frac{\bar{z}}{\|\bar{z}\|}\right) & \text{if } \bar{z} \neq 0, \\ \frac{1}{2}(1, \overline{w}) & \text{if } \bar{z} = 0, \end{cases}$$

where $\overline{w}$ is any vector in $\mathbb{R}^{\ell-1}$ such that $\|\overline{w}\| = 1$.

The following result summarizes some interesting properties on eigenvalues of Euclidean Jordan algebras.

Proposition 1 *[LFF18, Proposition 2.4] Let $(\mathscr{E}, \circ)$ be an Euclidean Jordan algebra, $\mathscr{K}$ its cone of squares and $u, v \in \mathscr{E}$.*

(i) $u \in \mathscr{K}$ ($u \in \operatorname{int}(\mathscr{K})$) if and only if the eigenvalues of u are nonnegative (positive);

(ii) if $u, v \in \mathscr{K}$, then, $u \circ v = 0$ if and only if $\langle u, v\rangle = 0$;

(iii) if $u, v \in \mathscr{K}$ is such that $\langle u, v\rangle = 0$ then u and v operator commute.

Theorem 12 *[LFF18, Proposition 2.5]. Let $(\mathscr{E}, \circ)$ be an Euclidean Jordan algebra, $\mathscr{K}$ its cone of squares and $u \in \mathscr{E}$ with spectral decomposition $u = \sum_{i=1}^{r} \lambda_i(u)c_i(u)$. Then, the projection of u onto $\mathscr{K}$ is given by*

$$[u]_+ = \sum_{i=1}^{r} \max\{0, \lambda_i(u)\}c_i(u).$$

In the case of the second-order cone, we have $[z]_+ = \max\{0, \lambda_1(z)\}c_1(z) + \max\{0, \lambda_2(z)\}c_2(z)$ which was explicit in [FLT01, Proposition 3.3] as follows

$$[z]_+ = \begin{cases} z & \text{if } \|\bar{z}\| \leq z_0, \\ 0 & \text{if } \|\bar{z}\| \leq -z_0, \\ \frac{\|\bar{z}\| + z_0}{2}\left(1, \frac{\bar{z}}{\|\bar{z}\|}\right) & \text{otherwise.} \end{cases} \tag{3.5}$$

By (3.5) it is easy to see that if $z \in K_\ell$ then $[z]_+ = z$. Moreover, if $z \in -K_\ell$ then $[z]_+ = 0$. In [CSS03], it is proved that the function $z \mapsto [z]_+$ is strongly semismooth and continuously differentiable at $z \in \mathbb{R}^\ell$ if the eigenvalues of z are nonzero.

3.4 Nonlinear Symmetric Cone Programming

In this section, we are interested in studying the problem that generalizes all the previously mentioned ones. The *nonlinear symmetric cone programming* (NSCP) problem that we consider here is defined below:

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

where $F\colon \mathbb{R}^n \to \mathbb{R}$ and $G\colon \mathbb{R}^n \to \mathcal{E}$ are continuously differentiable functions, and $\mathcal{K} \subseteq \mathcal{E}$ is a symmetric cone and $\mathcal{E}$ is a finite dimensional inner-product space. The Lagrangian function $\mathcal{L} : \mathbb{R}^n \times \mathcal{E} \to \mathbb{R}$ of (NSCP) is defined by

$$\mathcal{L}(x,\sigma) := F(x) - \langle G(x), \sigma\rangle.$$

We say that $(x^*,\sigma) \in \mathbb{R}^n \times \mathcal{E}$ is a KKT pair for (NSCP) if the following conditions are satisfied:

$$\nabla\mathcal{L}(x^*,\sigma) = \nabla F(x^*) - JG(x^*)^*\sigma = 0, \tag{3.6}$$
$$\langle G(x^*), \sigma\rangle = 0, \tag{3.7}$$
$$G(x^*) \in \mathcal{K}, \tag{3.8}$$
$$\sigma \in \mathcal{K}. \tag{3.9}$$

By considering the Euclidean Jordan algebra $(\mathcal{E},\circ)$ such that $\mathcal{K}$ is its cone of squares, by Proposition 1 we can replace condition (3.7) by $G(x^*)\circ\sigma = 0$. Since this implies that $G(x^*)$ and σ operator commute it is easy to see that condition (3.7) can also be replaced by

$$\lambda_i(G(x^*))\lambda_i(\sigma) = 0, \quad i = 1,\dots,r, \tag{3.10}$$

where $G(x^*)$ and σ operator commute and the ordering of the eigenvalues is the order given by their common Jordan frame. In addition, we have the following definition.

Definition 9 *We say that a KKT pair $(x^*,\sigma) \in \mathbb{R}^n \times \mathcal{E}$ satisfies the* strict complementarity *condition if*

$$\mathit{rank}\ G(x^*) + \mathit{rank}\ \sigma = r.$$

A constraint qualification is required in order for the KKT conditions to hold at a local solution. As well-known constraint qualifications for (NSCP), we can cite nondegeneracy and Robinson's CQ.

Definition 10 *We say that $x^* \in \mathbb{R}^n$ satisfies the nondegeneracy condition if*

$$\mathit{Im}\ JG(x^*) + \mathcal{T}^{lin}_{\mathcal{K}}(G(x^*)) = \mathcal{K},$$

where $\mathcal{T}^{lin}_{\mathcal{K}}(G(x^*))$ is the lineality space of the tangent cone of $\mathcal{K}$ at $G(x^*)$ and $\mathit{Im}\ JG(x^*)$ denotes the image of the linear mapping $JG(x^*)$.

Nondegeneracy was studied by Bonnans and Shapiro in [BS00], which is also related to the so-called transversality condition. When nondegeneracy is satisfied at a point, the associated Lagrange multiplier is unique. Robinson's CQ, which is weaker than nondegeneracy, is defined below

Definition 11 *We say that $x^* \in \mathbb{R}^n$ satisfies Robinson's CQ if there exists $d \in \mathbb{R}^n$ such that*

$$G(x^*) + JG(x^*)d \in \operatorname{int}(\mathcal{K}),$$

where $\operatorname{int}(\mathcal{K})$ denotes the interior of $\mathcal{K}$.

When Robinson's CQ is satisfied, the set of Lagrange multipliers is nonempty and bounded.

3.4.1 Sequential Optimality Condition for NSCP

As previously stated, in general, most optimality conditions are of the form "KKT or not-CQ". An alternative to this type of optimality conditions, that do not require any CQ, is a so-called *sequential optimality condition*. The sequence needed for checking this condition is usually the one generated by standard algorithms. This provides global convergence results stronger than the usual ones. In this section, we intend to present an extension for the condition called Approximate- Karush-Kuhn-Tucker (AKKT), that was studied in [AHM11, QW00] for NLP and [AHV18] for NSDP. In NLP, the AKKT condition is a strong optimality condition satisfied by limit points of many first and second-order methods. For more details see [BHR18, BKM17, DDTA13, Hae18, HM15, HS11, AHSS12a, TYW17]. Let us start by presenting our definition of AKKT for NSCP problems.

Definition 12 *Let $x^* \in \mathbb{R}^n$ be a feasible point. We say that x^* is an* Approximate Karush-Kuhn-Tucker *(AKKT) point for* (NSCP) *if there exist sequences $\{x^k\} \subset \mathbb{R}^n$ and $\{\sigma^k\} \subset \mathscr{K}$ with $x^k \to x^*$ such that*

$$\lim_{k\to\infty} \nabla \mathscr{L}(x^k, \sigma^k) = 0, \tag{3.11}$$

$$\text{If } \lambda_i(G(x^*)) > 0 \text{ then } \lambda_i(\sigma^k) = 0 \text{ for all } i = 1, \dots, r, \text{ and sufficiently large } k, \tag{3.12}$$

$$\lim_{k\to\infty} c_i(\sigma^k) = c_i(G(x^*)) \text{ for all } i = 1, \dots, r, \tag{3.13}$$

where

$$G(x^*) = \lambda_1(G(x^*))c_1(G(x^*)) + \cdots + \lambda_r(G(x^*))c_r(G(x^*)), \tag{3.14}$$

$$\sigma^k = \lambda_1(\sigma^k)c_1(\sigma^k) + \cdots + \lambda_r(\sigma^k)c_r(\sigma^k), \tag{3.15}$$

are spectral decompositions of $G(x^*)$ and σ^k.

Notice that the definition of the AKKT condition is independent of the choices of $c_i(G(x^*))$ and $c_i(\sigma^k)$ for $i = 1, \dots, r$. We recall that in the definition of AKKT for NLP, the Lagrange multipliers associated with inactive constraints at x^* are taken equal to zero. In NSCP we need a notion that gives us a way to associate the "inactive" eigenvalues of $G(x^*)$ with the zero eigenvalues of σ^k. In the above definition, the relation (3.13) provides the necessary condition for pairing the eigenvalues correctly. The equivalence below provides a way to detect an AKKT sequence in terms of a sequence of tolerances $\varepsilon_k \to 0$.

Lemma 7 *A feasible point $x^* \in \mathbb{R}^n$ satisfies the AKKT condition if, and only if, there are sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\sigma^k\} \subset \mathscr{K}$, $\{\varepsilon_k\} \subset \mathbb{R}_+$ with $x^k \to x^*$, $\varepsilon_k \to 0$, such that*

$$\|\nabla \mathscr{L}(x^k, \sigma^k)\| \le \varepsilon_k, \tag{3.16}$$

$$\|[-G(x^k)]_+\| \le \varepsilon_k, \tag{3.17}$$

$$\lambda_i(G(x^k)) > \varepsilon_k \Rightarrow \lambda_i(\sigma^k) = 0 \text{ for all } i = 1, \dots, r, \text{ and sufficiently large } k, \tag{3.18}$$

$$\|c_i(\sigma^k) - c_i(G(x^k))\| \le \varepsilon_k \text{ for all } i = 1, \dots, r, \tag{3.19}$$

where

$$G(x^k) = \lambda_1(G(x^k))c_1(G(x^k)) + \cdots + \lambda_r(G(x^k))c_r(G(x^k)) \text{ and } \sigma^k = \lambda_1(\sigma^k)c_1(\sigma^k) + \cdots + \lambda_r(\sigma^k)c_r(\sigma^k)$$

are spectral decompositions of $G(x^k)$ and σ^k.

Proof 11 *Let us assume that $x^* \in \mathbb{R}^n$ satisfies the AKKT condition. By the definition of AKKT, we can take the spectral decompositions for $G(x^*)$ and σ^k given by (3.14) and (3.15) with $c_i(G(x^k)) \to c_i(G(x^*))$ and $c_i(\sigma^k) \to c_i(G(x^*))$ for $i = 1, \dots, r$, such that (3.11)-(3.13) hold. Now, note that*

$$\|[-G(x^k)]_+\| = \left\| \sum_{i=1}^r \max\{0, -\lambda_i(G(x^k))\} c_i(G(x^k)) \right\| \to \left\| \sum_{i=1}^r \max\{0, -\lambda_i(G(x^*))\} c_i(G(x^*)) \right\| = 0, \tag{3.20}$$

because x^* is feasible. Define the sequence $\{\varepsilon_k\} \subset \mathbb{R}_+$ as follows

$$\varepsilon_k := \max\left\{\|\nabla\mathscr{L}(x^k,\, \sigma^k)\|,\ \|[-G(x^k)]_+\|,\ \lambda_i(G(x^k)) : i \in I(x^*),\ \|c_i(\sigma^k) - c_i(G(x^k))\| : i = 1,\dots,r\right\},$$

where, $I(x^*) := \{i \mid \lambda_i(G(x^*)) = 0\}$. Observe that the limit for $k \to \infty$ of each term inside the above maximum is zero from (3.11), (3.13) and (3.20). By the continuity of the involved functions, we have that $\varepsilon_k \to 0$. Hence (3.16), (3.17) and (3.19) hold. Now, let k be sufficiently large. To prove (3.18), note that for $j \in \{1,\dots,r\}$ such that $\lambda_j(G(x^k)) > \varepsilon_k$ we have that

$$\lambda_j(G(x^k)) > \varepsilon_k \geq \lambda_i(G(x^k)) \quad \textit{for all} \quad i \in I(x^*).$$

In particular, $j \notin I(x^*)$, that is, $\lambda_i(G(x^*)) > 0$. Hence, from (3.12), $\lambda_j(\sigma^k) = 0$, and so (3.18) holds.

Let us now assume that there are $\{x^k\}$, $\{\sigma^k\}$, $\{\varepsilon_k\}$ satisfying $x^k \to x^*$, $\varepsilon_k \to 0$ and (3.16)-(3.19). The continuity of functions involved and (3.17) ensures that x^* is feasible. The limit (3.11) follows trivially by (3.16). Since $\{c_i(\sigma^k)\}$ and $\{c_i(G(x^k))\}$ are bounded for all $i = 1,\dots,r$, we may take a subsequence if necessary such that $c_i(G(x^k)) \to c_i(G(x^*))$ for all $i = 1,\dots,r$. Hence, (3.13) follows from (3.19). Now, if we suppose that $\lambda_i(G(x^*)) > 0$ then, $\lambda_i(G(x^k)) > \varepsilon_k$ for k large enough . Thus, by (3.18) we have that $\lambda_i(\sigma^k) = 0$, which means that (3.13) holds. Therefore, $x^* \in \mathbb{R}^n$ is an AKKT point.

We can use the previous lemma to define a simple stopping criterion for algorithms for NSCP. Let ε_{opt}, $\varepsilon_{\text{feas}}$, $\varepsilon_{\text{compl}}$ and $\varepsilon_{\text{spec}}$ be small tolerances associated with optimality, feasibility, complementarity and spectral decomposition, respectively. Then, an algorithm for solving NSCP that generates an AKKT sequence $\{x^k\} \subset \mathbb{R}^n$ and a dual sequence $\{\sigma^k\} \subset \mathscr{K}$ can be safely stopped when:

$$\|\nabla\mathscr{L}(x^k,\sigma^k)\| \leq \varepsilon_{\text{opt}}, \tag{3.21}$$
$$\|[-G(x^k)]_+\| \leq \varepsilon_{\text{feas}}, \tag{3.22}$$
$$\lambda_i(G(x^k)) > \varepsilon_{\text{compl}} \quad \Rightarrow \quad \lambda_i(\sigma^k) = 0, \tag{3.23}$$
$$\|c_i(\sigma^k) - c_i(G(x^k))\| \leq \varepsilon_{\text{spec}}. \tag{3.24}$$
$$\tag{3.25}$$

In order to prove that the AKKT conditions are genuine optimality conditions for NSCP, we will use the penalty technique, which is quite common in nonlinear programming. More precisely, consider the problem

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & F(x), \\ \text{subject to} & G(x) \in \mathscr{K},\, x \in \Omega, \end{array} \tag{3.26}$$

where, $\Omega \subseteq \mathbb{R}^n$ is a nonempty closed set. For this problem, the function $P(x) := \|\,[-G(x)]_+\,\|^2$ is a measure of infeasibility, in the sense that $P(x) \geq 0$ for all $x \in \Omega$, and it is feasible if, and only if, $P(x) = 0$. This function can be used for constructing a penalized problem that has the property described below.

Lemma 8 ([FM68]) *Choose a sequence $\{\rho_k\} \subset \mathbb{R}$ with $\rho_k \to +\infty$. For each k, let x^k be a solution, if it exists, for the following penalized problem:*

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\textit{Minimize}} & F(x) + \rho_k P(x), \\ \textit{subject to} & x \in \Omega. \end{array}$$

Then, all limit points of $\{x^k\}$ are global minimizers of (3.26).

Theorem 13 *Let $x^* \in \mathbb{R}^n$ be a local minimizer of (NSCP). Then, x^* satisfies the AKKT condition.*

Proof 12 *Consider the problem*

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & F(x)+\frac{1}{2}\|x-x^*\|^2, \\ \text{subject to} \quad & G(x)\in\mathscr{K}, \\ & x\in\mathscr{B}(x^*,\delta), \end{aligned} \tag{3.27}$$

where $\mathscr{B}(x^*,\delta):=\{x\in\mathbb{R}^n \mid \|x-x^*\|\leq\delta\}$ *for* $\delta>0$ *such that* x^* *is the unique solution of (3.27). Let* $x^k\in\mathbb{R}^n$ *be a solution of*

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & F(x)+\frac{1}{2}\|x-x^*\|^2+\frac{\rho_k}{2}\|[-G(x)]_+\|^2, \\ \text{subject to} \quad & x\in\mathscr{B}(x^*,\delta), \end{aligned} \tag{3.28}$$

where $\{\rho_k\}\subset\mathbb{R}$ *with* $\rho_k\to+\infty$*. By the continuity of the involved functions and the compactness of* $\mathscr{B}(x^*,\delta)$*, the sequence* x^k *is well-defined for all k. Note that,*

$$F(x^k)+\frac{1}{2}\|x^k-x^*\|^2+\frac{\rho_k}{2}\|[-G(x^k)]_+\|^2\leq F(x^*), \tag{3.29}$$

since $x^*\in\mathscr{B}(x^*,\delta)$ *is a feasible point. In addition, the set* $\mathscr{B}(x^*,\delta)$ *is nonempty and compact, hence, by Lemma 8 we have that all limit points of* x^k *are global solutions of (3.28) and* $x^k\to x^*$*. Now, for large enough k, we have that* x^k *is a local minimizer of* $F(x)+\frac{1}{2}\|x-x^*\|^2+\frac{\rho_k}{2}\|[-G(x)]_+\|^2$*. Then*

$$\nabla F(x^k)+(x^k-x^*)-\rho_k JG(x^k)^*[-G(x^k)]_+=0. \tag{3.30}$$

Let us define for all k, $\sigma^k:=\rho_k[-G(x^k)]_+\in\mathscr{K}$*. Let us consider the spectral decompositions*

$$G(x^k)=\lambda_1(G(x^k))c_1(G(x^k))+\cdots+\lambda_r(G(x^k))c_r(G(x^k)),$$

$$\sigma^k=\rho_k\max\{0,\lambda_1(-G(x^k))\}c_1(G(x^k))+\cdots+\rho_k\max\{0,\lambda_r(-G(x^k))\}c_r(G(x^k)).$$

Note that $-\lambda_i(G(x^k))=\lambda_i(-G(x^k))$ *for all* $i=1,\ldots,r$*. If* $\lambda_i(G(x^*))>0$ *then* $\lambda_i(G(x^k))>0$ *for k large enough. Thus,* $\lambda_i(\sigma^k)=\rho_k\max\{0,\lambda_i(-G(x^k))\}=0$*. Moreover,* σ^k *operator commutes with* $G(x^k)$*. Then, the assumption* $c_i(\sigma^k)\to c_i(G(x^*))$ *is trivially satisfied. Taking the limit when* $k\to\infty$ *in (3.30) we have*

$$\lim_{k\to\infty}\nabla\mathscr{L}(x^k,\sigma^k)=\lim_{k\to\infty}\nabla F(x^k)-JG(x^k)^*\sigma^k=0. \tag{3.31}$$

Hence, x^* *satisfies the AKKT condition.*

3.4.2 The strength of the AKKT condition for NSCP

In this section, we will measure the strength of the AKKT condition for NSCP in comparison with an optimality condition of the type "KKT or not-CQ".

Let us show that our necessary optimality condition is at least as good as "KKT or not-Robinson's CQ".

Theorem 14 *Let* $x^*\in\mathbb{R}^n$ *be an AKKT point that satisfies Robinson's CQ. Then,* x^* *satisfies the KKT condition.*

Proof 13 *By the AKKT definition for NSCP, there exist sequences* $\{x^k\}\subset\mathbb{R}^n$ *and* $\{\sigma^k\}\subset\mathscr{K}$ *with* $x^k\to x^*$ *such that*

$$\lim_{k\to\infty}\nabla\mathscr{L}(x^k,\sigma^k)=0, \tag{3.32}$$

$$\text{If } \lambda_i(G(x^*))>0 \text{ then } \lambda_i(\sigma^k)=0 \text{ for all } i=1,\ldots,r, \text{ and sufficiently large } k, \tag{3.33}$$

$$\lim_{k\to\infty}c_i(\sigma^k)=c_i(G(x^*)) \text{ for all } i=1,\ldots,r, \tag{3.34}$$

where

$$G(x^*) = \lambda_1(G(x^*))c_1(G(x^*)) + \cdots + \lambda_r(G(x^*))c_r(G(x^*)), \tag{3.35}$$
$$\sigma^k = \lambda_1(\sigma^k)c_1(\sigma^k) + \cdots + \lambda_r(\sigma^k)c_r(\sigma^k), \tag{3.36}$$

are spectral decompositions of $G(x^*)$ *and* σ^k*, respectively.*

If $\{\sigma^k\}$ *is contained in a compact set, there exists* $\sigma \in \mathscr{K}$ *such that, taking a subsequence if necessary,* $\sigma^k \to \sigma$*. Then, by (3.32) we have that*

$$\nabla F(x^*) - JG(x^*)^*\sigma = 0.$$

Now, consider the index sets

$$I_1 := \{i \mid \lambda_i(G(x^*)) > 0\} \quad \textit{and} \quad I_2 := \{i \mid \lambda_i(G(x^*)) = 0\}.$$

Note that, by (3.34)*,* $G(x^*)$ *and* σ *operator commute. Then,*

$$G(x^*) \circ \sigma = \sum_{i \in I_1} \lambda_i(G(x^*))\lambda_i(\sigma)c_i(G(x^*)) + \sum_{i \in I_2} \lambda_i(G(x^*))\lambda_i(\sigma)c_i(G(x^*)) = 0, \tag{3.37}$$

since $\lambda_i(G(x^*)) = 0$ *for* $i \in I_2$ *and* $\lambda_i(\sigma) = 0$ *for* $i \in I_1$*, which implies KKT. Now, suppose that* $\{\sigma^k\}$ *is not contained in a compact set. Let us consider a subsequence such that* $t_k := \|\sigma^k\| \to \infty$*. Then,* $\dfrac{\sigma^k}{t_k} \to \sigma^* \neq 0$ *for some* $\sigma^* \in \mathscr{K}$*. Then, by (3.32) and* (3.33) *we have that*

$$\lim_{k \to \infty} \frac{\nabla F(x^k)}{t_k} - JG(x^k)^* \frac{\sigma^k}{t_k} = -JG(x^*)^*\sigma^* = 0, \tag{3.38}$$

and by taking the limit in (3.36)*,*

$$G(x^*) \circ \sigma^* = \sum_{i \in I_1} \lambda_i(G(x^*))\lambda_i(\sigma^*)c_i(G(x^*)) + \sum_{i \in I_2} \lambda_i(G(x^*))\lambda_i(\sigma^*)c_i(G(x^*)) = 0, \tag{3.39}$$

To see that (3.38)-(3.39) contradicts Robinson's condition, let $d \in \mathbb{R}^n$ *be such that*

$$G(x^*) + JG(x^*)d \in \operatorname{int}(\mathscr{K}).$$

Thus,

$$0 = \langle G(x^*), \sigma^* \rangle + \langle JG(x^*)^*\sigma^*, d \rangle \tag{3.40}$$
$$= \langle G(x^*) + JG(x^*)d, \sigma^* \rangle, \tag{3.41}$$

which is a contradiction, since $0 \neq \sigma^* \in \mathscr{K}$ *and* $G(x^*) + JG(x^*)d \in \operatorname{int}(\mathscr{K})$*.*

Similarly to [AHV18], one can define a constraint qualification (AKKT-regularity), strictly weaker than Robinson's CQ, emposing that the constraint set is such that for every objective function such that AKKT holds at x^*, the KKT conditions also holds. A geometric interpretation of AKKT-regularity is also available. See [AHV18].

3.5 Nonlinear Second-Order Cone Programming

Besides the non-negative orthant and the semidefinite cone, an important example of a symmetric cone is the second-order cone. In this section, we will study the nonlinear second-order cone programming (NSOCP) problem, which is a particular case of NSCP when the symmetric cone is taken as a second-order cone. Since the second-order cone has a very particular structure, we present the AKKT condition in a more adequate form. Also, there are constraint qualifications for NSOCP that have not yet been extended to NSCPs as it is the case of the constant rank constraint qualification (CRCQ) and others, which allows us to compare our optimality conditions with a stronger condition “KKT or not-CQ”, namely, for CQs strictly weaker than Robinson’s CQ.

The nonlinear second-order cone programming that we are interested consists in:

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

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $g\colon \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable functions and $K := K_{m_1} \times \cdots \times K_{m_r}$ is the second-order cone given by the Cartesian product of second-order cones K_{m_i}, $i = 1, \ldots, r$, such that $m_1 + \cdots + m_r = m$. Let us define the feasible set of (NSOCP) by

$$F := \{x \in \mathbb{R}^n \mid g_i(x) \in K_{m_i} \ \text{ for all } \ i = 1, \ldots, r\}.$$

The topological interior and boundary of the cone K_{m_i} are characterized, respectively, by

$$\text{int}(K_{m_i}) := \{(x_0, \bar{x}) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid \|\bar{x}\| < x_0\} \ \text{ and } \ \text{bd}(K_{m_i}) := \{(x_0, \bar{x}) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid \|\bar{x}\| = x_0\}.$$

The boundary of K_{m_i} excluding the null vector will be denoted by $\text{bd}^+(K_{m_i})$. For $x^* \in \mathbb{R}^n$ we define the following index sets:

$$\begin{aligned} I_I(x^*) &:= \{i = 1, \ldots, r \mid g_i(x^*) \in \text{int}(K_{m_i})\}, \\ I_B(x^*) &:= \{i = 1, \ldots, r \mid g_i(x^*) \in \text{bd}^+(K_{m_i})\}, \\ I_0(x^*) &:= \{i = 1, \ldots, r \mid g_i(x^*) = 0\}. \end{aligned}$$

The KKT conditions for (NSOCP) are given as follows. Consider the Lagrangian function $L\colon \mathbb{R}^n \times \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_r} \to \mathbb{R}$ given by

$$L(x, \mu_1, \ldots, \mu_r) := f(x) - \sum_{i=1}^{r} \langle g_i(x), \mu_i \rangle.$$

We say that $x^* \in \mathbb{R}^n$ is a KKT point for (NSOCP) if there exist multipliers $\mu_i \in \mathbb{R}^{m_i}$, $i = 1, \ldots, r$, such that

$$\nabla L(x^*, \mu_1, \ldots, \mu_r) = 0, \tag{3.42}$$
$$g_i(x^*) \circ \mu_i = 0, \qquad i = 1, \ldots, r, \tag{3.43}$$
$$g_i(x^*) \in K_{m_i}, \qquad i = 1, \ldots, r, \tag{3.44}$$
$$\mu_i \in K_{m_i}, \qquad i = 1, \ldots, r. \tag{3.45}$$

Condition (3.43) was studied by Alizadeh and Goldfarb in [AG03] where the authors concluded that (3.43) is equivalent to the following conditions:

$$\begin{aligned} &(i) \quad i \in I_I(x^*) \Rightarrow \mu_i = 0, \\ &(ii) \quad i \in I_B(x^*) \Rightarrow \mu_i = 0 \ \text{ or } \ \mu_i = \Big([\mu_i]_0, -\alpha_i(x^*)\overline{g_i(x^*)}\Big) \ \text{ with } [\mu_i]_0 \neq 0, \end{aligned}$$

where $\alpha_i(x^*) = \frac{[\mu_i]_0}{[g_i(x^*)]_0}$ with $i = 1, \ldots, r$. The relations $i \in I_I(x^*)$ and $i \in I_B(x^*)$ can be written in terms of the eigenvalues of $g_i(x^*)$ as follows:

$$g_i(x^*) \in \text{int}(K_{m_i}) \quad \Leftrightarrow \quad \lambda_1(g_i(x^*)) > 0,\ \lambda_2(g_i(x^*)) > 0, \tag{3.46}$$
$$g_i(x^*) \in \text{bd}^+(K_{m_i}) \quad \Leftrightarrow \quad \lambda_1(g_i(x^*)) = 0,\ \lambda_2(g_i(x^*)) > 0. \tag{3.47}$$

Note that, in the case of (ii), the formula for the multiplier being $\mu_i = \Big([\mu_i]_0, -\alpha_i(x^*)\overline{g_i(x^*)}\Big)$ with $[\mu_i]_0 \neq 0$ can be equivalently written as $\mu_i \in \text{bd}^+(K_{m_i})$. In this case, due to the expression of μ_i and the definition of

$c_1(\cdot)$ and $c_2(\cdot)$, we can conclude that μ_i and $g_i(x^*)$ operator commute but $c_1(\mu_i) = c_2(g_i(x^*))$ and $c_2(\mu_i) = c_1(g_i(x^*))$.

Similarly to NSCP, we say that a KKT pair (x^*, μ) of (NSOCP) satisfies *strict complementarity* if

$$g_i(x^*) + \mu_i \in \text{int } (K_{m_i}) \quad \text{for all} \quad i = 1, \dots, r.$$

Note that under strict complementarity, it is not the case that $\mu_i = g_i(x^*) = 0$, nor that $g_i(x^*) \in \text{int}(K_{m_i})$ with $\mu_i = 0$, nor that $\mu_i \in \text{int}(K_{m_i})$ with $g_i(x^*) = 0$.

3.5.1 Sequential Optimality Condition for NSOCP

In Section 3.4, we introduced our definition of sequential optimality condition for NSCP. This definition was made in terms of the eigenvalues that happened when we take the spectral decomposition. In (NSOCP), we do not need to use the eigenvalues of the constraints and the Lagrange multipliers. The next result gives us the definition of the AKKT condition for NSOCP, which is obtained by the definition of the AKKT condition for (NSCP) in the particular case where the symmetric cone is a second-order cone. Let us consider $\mathscr{K}$ a simple symmetric cone. Note that, the generalization for several cones is straightforward since every symmetric cone can be decomposed as a direct sum of simple symmetric cones.

Theorem 15 *A feasible point $x^* \in \mathbb{R}^n$ satisfies the AKKT condition for (NSCP) when $\mathscr{K} = K_{m_i}$ if, and only if, there exist sequences $\{x^k\} \subset \mathbb{R}^n$ and $\{\mu_i^k\} \subset K_{m_i}$ with $x^k \to x^*$ such that*

$$\lim_{k\to\infty} \nabla f(x^k) - \sum_{i=1}^{r} Jg_i(x^k)^T \mu_i^k = 0, \tag{3.48}$$

$$i \in I_I(x^*) \Rightarrow \mu_i^k = 0 \textit{ for sufficiently large } k, \tag{3.49}$$

$$i \in I_B(x^*) \Rightarrow \mu_i^k = 0 \textit{ or } \mu_i^k \in bd^+(K_{m_i}) \textit{ with } -\frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|} \to \frac{\overline{g_i(x^*)}}{\|\overline{g_i(x^*)}\|}. \tag{3.50}$$

Proof 14 *By the definition of the AKKT for (NSCP) we need only verify the relation between the complementarity condition requested in AKKT for NSCP (3.12) and (3.13) and the complementarity conditions given by (3.49) and (3.50).*

Assume that the feasible point $x^* \in \mathbb{R}^n$ satisfies the AKKT condition for NSCP. Since $\mathscr{K} = K_{m_i}$, we have the following spectral decomposition for $g_i(x^*) \in K_{m_i}$

$$g_i(x^*) = \lambda_1(g_i(x^*))e_1(g_i(x^*)) + \lambda_2(g_i(x^*))e_2(g_i(x^*)). \tag{3.51}$$

We have two cases to consider:

(i) If $i \in I_I(x^*)$ then, $\lambda_1(g_i(x^*)) > 0$ and $\lambda_2(g_i(x^*)) > 0$. By AKKT for NSCP, we have that $\lambda_1(\mu_i^k) = \lambda_2(\mu_i^k) = 0$ and therefore, $\mu_i^k = 0$.

(ii) If $i \in I_B(x^*)$ then, $\lambda_1(g_i(x^*)) = 0$ and $\lambda_2(g_i(x^*)) > 0$. By AKKT for NSCP, we have that $\lambda_2(\mu_i^k) = 0$. Note that, we have two options for $\lambda_2(\mu_i^k)$.

(a_1) If $\lambda_2(\mu_i^k) = \|\overline{\mu_i^k}\| + [\mu_i^k]_0$ then, $\lambda_2(\mu_i^k) = 0$ implies $\mu_i^k = 0$.

(a_2) If $\lambda_2(\mu_i^k) = \|\overline{\mu_i^k}\| - [\mu_i^k]_0$ with $\lambda_1(\mu_i^k) = \|\overline{\mu_i^k}\| + [\mu_i^k]_0 > 0$, we have that $\lambda_2(\mu_i^k) = 0$ implies $\mu_i^k \in bd^+(K_{m_i})$. Moreover, by the format of $\lambda_1(\mu_i^k)$ and $\lambda_2(\mu_i^k)$ we have that

$$e_1(\mu_i^k) = \left(\frac{1}{2}, \frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|}\right) \quad \textit{and} \quad e_2(\mu_i^k) = \left(\frac{1}{2}, -\frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|}\right).$$

By assumption of the AKKT for NSCP, we have that $c_2(\mu_i^k) \to c_2(g_i(x^*))$. Thus,we get $-\frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|} \to$

$$\frac{\overline{g_i(x^*)}}{\|\overline{g_i(x^*)}\|}.$$

Now, let us prove the converse. Note that, if $\lambda_2(g_i(x^*)) = 0$ then $\lambda_1(g_i(x^*)) = 0$ and we do not need to check $\lambda_1(\mu_i^k)$ and $\lambda_2(\mu_i^k)$. Consider the following cases:

(i) $\lambda_1(g_i(x^*)) > 0$ and $\lambda_2(g_i(x^*)) > 0$. Thus, $i \in I_I(x^*)$ and then, by AKKT for NSOCP, $\mu_i^k = 0$. Hence, $\lambda_1(\mu_i^k) = \lambda_2(\mu_i^k) = 0$ and $c_1(\mu_i^k) \to c_1(g_i(x^*))$ and $c_2(\mu_i^k) \to c_2(g_i(x^*))$ is trivially satisfied.

(ii) $\lambda_1(g_i(x^*)) = 0$ and $\lambda_2(g_i(x^*)) > 0$. Thus, $i \in I_B(x^*)$. By AKKT for NSOCP, $\mu_i^k = 0$ or $\mu_i^k \in bd^+(K_{m_i})$ with $-\frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|} \to \frac{\overline{g_i(x^*)}}{\|\overline{g_i(x^*)}\|}$.

(b_1) If $\mu_i^k = 0$ then $\lambda_1(\mu_i^k) = \lambda_2(\mu_i^k) = 0$ and $c_1(\mu_i^k) \to c_i(g_i(x^*))$ and $c_2(\mu_i^k) \to c_2(g_i(x^*))$ is trivially satisfied.

(b_2) If $\mu_i^k \in bd^+(K_{m_i})$ with $-\frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|} \to \frac{\overline{g_i(x^*)}}{\|\overline{g_i(x^*)}\|}$ we have that $\lambda_2(\mu_i^k) = [\mu_i^k]_0 - \|\overline{\mu_i^k}\| = 0$ with $\lambda_1(\mu_i^k) = \|\overline{\mu_i^k}\| + [\mu_i^k]_0 > 0$. Moreover, by the format of $\lambda_1(\mu_i^k)$ and $\lambda_2(\mu_i^k)$ we have that

$$e_1(\mu_i^k) = \left(\frac{1}{2}, \frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|}\right) \quad \textit{and} \quad e_2(\mu_i^k) = \left(\frac{1}{2}, -\frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|}\right).$$

Furthermore, $-\frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|} \to \frac{\overline{g_i(x^*)}}{\|\overline{g_i(x^*)}\|}$, implies that $c_1(\mu_i^k) \to c_1(g_i(x^*))$ and $c_2(\mu_i^k) \to c_2(g_i(x^*))$.

Note that, the complementarity condition (3.49) can be represented equivalently by

$$i \in I_B(x^*) \Rightarrow \overline{\mu_i^k}[g_i(x^*)]_0 + \overline{g_i(x^*)}[\mu_i^k]_0 \to 0 \text{ with } \mu_i^k \in \text{bd}^+(K_{m_i}).$$

The previous Theorem provides the AKKT definition for NSOCP so that our definition of AKKT for NSCP and for NSOCP are equivalent. Moreover, the definition does not need to use the eigenvalues what do AKKT for NSOCP simple to verify. Similar to the Definition of AKKT for NSCP we can define a stopping criterion for algorithms to solve NSOCP. The following is a Lemma that shows how to define the stopping criterion.

Lemma 9 *The point $x^* \in \mathbb{R}^n$ is an AKKT point if, and only if, there are sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\mu^k\} \subset K$, $\{\varepsilon_k\} \subset \mathbb{R}$ with $x^k \to x^*$, $\varepsilon_k \to 0^+$ such that*

$$\|\nabla L(x^k, \mu^k)\| \leq \varepsilon_k, \quad (3.52)$$

$$\|[-g_i(x^k)]_+\| \leq \varepsilon_k, \quad (3.53)$$

$$\left([g_i(x^k)]_0 - \|\overline{g_i(x^k)}\|\right) > \varepsilon_k \;\Rightarrow\; \mu_i^k = 0 \;\textit{ for } \; k \textit{ large enough}, \quad (3.54)$$

$$[g_i(x^k)]_0 > \varepsilon_k, \;\left|\|\overline{g_i(x^k)}\| - [g_i(x^k)]_0\right| \leq \varepsilon_k \;\Rightarrow\; \|\overline{\mu_i^k}[g_i(x^k)]_0 + \overline{g_i(x^k)}[\mu_i^k]_0\| \leq \varepsilon_k \;\textit{ for } \|\overline{\mu_i^k}\| = [\mu_i^k]_0. \quad (3.55)$$

The Lemma 9 suggests how to proceed to define a stopping criterion from the definition of AKKT. Consider the small tolerances ε_{opt}, $\varepsilon_{\text{feas}}$, $\varepsilon_{\text{compl}_1}$, $\varepsilon_{\text{compl}_2}$ and $\varepsilon_{\text{compl}_3}$ associated to optimality, feasibility and complementarity, respectively, the algorithm generates AKKT sequences $\{x^k\} \subset \mathbb{R}^n$ and $\{\mu^k\} \subset K$ such that

- $\|\nabla L(x^k, \mu^k)\| \leq \varepsilon_{\text{opt}}$,
- $\|[-g_i(x^k)]_+\| \leq \varepsilon_{\text{feas}}$,
- if $\left([g_i(x^k)]_0 - \|\overline{g_i(x^k)}\|\right) < -\varepsilon_{\text{compl}_1}$ then $\mu_i^k = 0$,

- if $[g_i(x^k)]_0 > \varepsilon_{\text{compl}_2}$ and $\left| \|\overline{g_i(x^k)}\| - [g_i(x^k)]_0 \right| \leq \varepsilon_{\text{compl}_3}$ then

$$\|\overline{\mu_i^k}[g_i(x^k)]_0 + \overline{g_i(x^k)}[\mu_i^k]_0\| \leq \varepsilon_{\text{compl}_3} \text{ for } \|\overline{\mu_i^k}\| = [\mu_i^k]_0$$

The next theorem follows directly from AKKT for NSOCP implies AKKT for NSCP.

Theorem 16 *Let x^* a local minimizer of NSOCP. Then, x^* satisfies AKKT.*

In the next example we present a case where the minimizer does not satisfy the KKT conditions but it is possible to build an AKKT sequence.

Example 5 *(AKKT sequence at a non-KKT solution). Consider the following problem*

$$\begin{aligned} &\underset{x\in\mathbb{R}}{\textit{Minimize}} && x, \\ &\textit{subject to} && g_1(x) = (-x,0) \in K_2,\ g_2(x) = (0,x^2) \in K_2, \\ &&& g_3(x) = (1,x) \in K_2,\ g_4(x) = (1+x,1+x) \in K_2. \end{aligned} \tag{3.56}$$

The point $x^* = 0$ is the optimal solution of (3.56).

(i) $x^* = 0$ is not a KKT point. Firstly, note that

$$g_1(x^*) = (0,0) \in K_2, \quad g_2(x^*) = (0,0) \in K_2, \quad g_3(x^*) = (1,0) \in \mathit{int}\,(K_2), \quad g_4(x^*) = (1,1) \in \mathit{bd}^+(K_2).$$

Let $\mu_1 = ([\mu_1]_0, \overline{\mu_1})$, $\mu_2 = ([\mu_2]_0, \overline{\mu_2})$, $\mu_3 = ([\mu_3]_0, \overline{\mu_3})$ and $\mu_4 = ([\mu_4]_0, \overline{\mu_4})$. We have

$$\nabla f(x^*) - \sum_{i=1}^{4} Jg_i(x^*)^T \mu_i = 0 \Rightarrow 1 + [\mu_1]_0 - \overline{\mu_3} - ([\mu_4]_0 + \overline{\mu_4}) = 0. \tag{3.57}$$

Moreover, $g_3(x^*) = (1,0) \in \mathit{int}\,(K_2)$ implies that $\mu_3 = 0$. In addition, $g_4(x^*) = (1,1) \in \mathit{bd}^+(K_2)$ implies that $\mu_4 = ([\mu_4]_0, -[\mu_4]_0)$. Thus, by (3.57) we have that $[\mu_1]_0 = -1$, hence $\mu_1 \notin K_2$ regardless of $\overline{\mu_1}$. Therefore, x^* can not be a KKT point.

(ii) $x^* = 0$ is an AKKT point. Define $x^k = \dfrac{1}{2k}$, $\mu_1^k = \left(\dfrac{1}{k}, \dfrac{1}{k}\right)$, $\mu_2^k = (k,k)$, $\mu_3^k = (0,0)$ and $\mu_4^k = (1,-1)$. Then,

$$\nabla f(x^k) - \sum_{i=1}^{4} Jg_i(x^k)^T \mu_i^k = 1 + \frac{1}{k} - 2\frac{1}{2k}k = \frac{1}{k} \to 0.$$

Note that, $g_1(x^*) = g_2(x^*) = 0$, $g_3(x^*) \in \mathit{int}\,(K_2)$ and $\mu_3^k = (0,0)$. Moreover, $g_4(x^*) \in \mathit{bd}^+(K_2)$ and $\mu_4^k = (1,-1) \in \mathit{bd}^+(K_2)$ such that $\overline{\mu_4^k}[g_4(x^*)]_0 + \overline{g_4(x^*)}[\mu_4^k]_0 = 0$. Therefore, $x^* = 0$ is an AKKT point.

3.6 An algorithm that satisfy the new optimality condition

The (linear) second-order cone programming has many interesting algorithms such as primal-dual interior-point method [ART03]. However, the recent interest in NSOCP compared to linear SOCP and mainly with NLP makes optimality conditions and algorithms for NSOCP a work in progress. In this section, we will present an Augmented Lagrangian algorithm, proposed by Yong-Jin Liu and Li-Wei Zhang in [LZ08]. This algorithm is an extension of the well-known augmented Lagrangian algorithm used in NLP.

3.6.1 Augmented Lagrangian Method

The augmented Lagrangian is a very popular algorithm in the field of optimization, we can see it in nonlinear programming [Buy72], in cone constrained problems [SS04] and others. In this subsection, we will prove that the augmented Lagrangian method proposed recently by Yong-Jin Liu and Li-Wei Zhang [LZ08] generates sequences that satisfy AKKT for NSOCP without any constraint qualification. In the augmented Lagrangian method, the authors use the additional assumption that the constraints satisfy the nondegeneracy condition. In the next section, we will present a constraint qualification associated with AKKT, which we will prove that our global convergence for augmented Lagrangian result generalizes the previous in the literature. The augmented Lagrangian function of NSOCP is the Powell-Hestenes-Rockafellar (PHR) defined as follows

$$L_\rho(x,\mu_1,\dots,\mu_r) := f(x) + \frac{1}{2\rho}\sum_{i=1}^{r}\left(\|[\mu_i - \rho g_i(x)]_+\|^2 - \|\mu_i\|^2\right), \tag{3.58}$$

where $\rho > 0$ is the penalty parameter. The gradient with respect to x of the augmented Lagrangian function 3.58 is given by:

$$\nabla_x L_\rho(x,\mu_1,\dots,\mu_r) = \nabla f(x) - \sum_{i=1}^{r} Jg_i(x)^T[\mu_i - \rho g_i(x)]_+.$$

The formal statement of the algorithm is as follows:

Algorithm 2 Augmented Lagrangian Algorithm for NSOCP

Let $\tau \in (0,1)$, $\gamma > 1$, $\rho_1 > 0$ and $\mu_i^0 \in \text{int}\,(K_{m_i})$. Take a sequence of tolerances $\{\varepsilon_k\} \subset \mathbb{R}_+$. Define $0 \preceq_{K_{m_i}} \hat{\mu}_i^1 \preceq_{K_{m_i}} \mu_i^{\max}$. Choose an arbitrary starting point $x^0 \in \mathbb{R}^n$. Initialize $k := 1$ and $\|V_i^0\| := +\infty$.

Step 1: Find an approximate minimizer x^k of $\nabla_x L_{\rho_k}(x,\mu_1^k,\dots,\mu_r^k)$. That is, find x^k satisfying

$$\text{Minimize } L_{\rho_k}(x,\mu_1^k,\dots,\mu_r^k). \tag{3.59}$$

Step 2: Define

$$V_i^k := \left[\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right]_+ - \frac{\hat{\mu}_i^k}{\rho_k}.$$

If $\|V^k\| \le \tau\|V^{k-1}\|$, define $\rho_{k+1} := \rho_k$, otherwise define $\rho_{k+1} := \gamma\rho_k$.
Step 3: Compute

$$\mu_i^k := [\hat{\mu}_i^k - \rho_k g_i(x^k)]_+,$$

and define $\hat{\mu}_i^{k+1} := \text{proj}(\mu_i^k)$ as the orthogonal projection of μ_i^k onto $\{y_i \in K_{m_i} | 0 \preceq_{K_{m_i}} y_i \preceq_{K_{m_i}} \mu_i^{\max}\}$.
Step 4: Set $k := k+1$, and go back to Step 1.

In above algorithm the function $V_i^k = \left[\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right]_+ - \frac{\hat{\mu}_i^k}{\rho_k}$ is a joint measure of feasibility and complementarity also used for update of penalty parameter. If the function $\|V^k\|$ is sufficiently reduced the penalty parameter ρ_k is keeps, otherwise he penalty parameter is increased. It is not difficulty to show that $V_i^k = 0$ if, and only if, x^k is a feasible sequence and the complementarity holds. The Lagrange multiplier sequence $\hat{\mu}^k$ is built bounded. However, the sequence of multipliers μ^k is not necessary bounded. The next result says that the limit points of sequence generated by algorithm previous are feasible points and global minimizer of the infeasibility measure, whenever of feasible set is nonempty.

Theorem 17 *Let $x^* \in \mathbb{R}^n$ be a feasible limit point of a sequence $\{x^k\}$ generated by Algorithm 2. Then, x^* is a stationary point for the following problem*

$$\textit{Minimize} \quad P(x) := \|[-g_i(x)]_+\|^2. \tag{3.60}$$

Proof 15 *We need consider two cases: the sequence $\{\rho_k\}$ is bounded or unbounded.*

(i) $\{\rho_k\}$ is bounded. Then, for $k \ge k_0$ we have $\rho_k = \rho_{k_0}$. Thus, $V_i^k \to 0$ and

$$\lim_{k\to\infty} \mu_i^k = \lim_{k\to\infty}\left[\hat{\mu}_i^k - \rho_{k_0} g_i(x^k)\right]_+ = [\hat{\mu}_i - \rho_{k_0} g_i(x^*)]_+ = \lim_{k\to\infty}\hat{\mu}_i^k = \hat{\mu}_i. \tag{3.61}$$

Take the following decomposition for $\hat{\mu}_i \in K_{m_i}$

$$\hat{\mu}_i = [\hat{\mu}_i - \rho_{k_0} g_i(x^*)]_+ = [\lambda_1]_+ c_1 + [\lambda_2]_+ c_2.$$

Then, we have

$$\hat{\mu}_i - \rho_{k_0} g_i(x^*) = \lambda_1 c_1 + \lambda_2 c_2 \Rightarrow g_i(x^*) = (1/\rho_{k_0})\left(([\lambda_1]_+ - \lambda_1)c_1 + ([\lambda_2]_+ - \lambda_2)c_2\right).$$

Note that, $[\lambda_1]_+ - \lambda_1 \geq 0$ *and* $[\lambda_2]_+ - \lambda_2 \geq 0$*. Hence,* $g_i(x^*) \in K_{m_i}$ *and* x^* *is a global minimizer of (3.60).*

(ii) $\{\rho_k\}$ *is unbounded. Let us define* $\delta^k := \nabla f(x^k) - \sum_{i=1}^{r} J g_i(x^k)^T \mu_i^k$ *where* $\mu_i^k := [\hat{\mu}_i^k - \rho_k g_i(x^k)]_+$*. By step 1 of algorithm 2 we have that* $\|\delta^k\| \leq \varepsilon_k$*. Thus,*

$$\frac{\delta^k}{\rho_k} = \frac{\nabla f(x^k)}{\rho_k} - \sum_{i=1}^{r} J g_i(x^k)^T \left[\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right]_+.$$

Since that, $\hat{\mu}_i^k$ *is bounded, the functions* ∇f *and* g *are continuous we have that* $\sum_{i=1}^{r} J g_i(x^*)^T [-g_i(x^*)]_+ = 0$*. Therefore,* x^* *is a stationary point of (3.60).*

Theorem 18 *Assume that* $x^* \in \mathbb{R}^n$ *is a feasible limit point of a sequence* $\{x^k\}$ *generated by Algorithm 2. Then,* x^* *is an AKKT point.*

Proof 16 *Assume that* $x^* \in \mathbb{R}^n$ *feasible is a limit point of* $\{x^k\}$ *generated by Algorithm 2 such that* $x^k \to x^*$*. From step 1 of algorithm we have that*

$$\left\|\nabla f(x^k) - \sum_{i=1}^{r} J g_i(x^k)^T [\hat{\mu}_i^k - \rho g_i(x^k)]_+\right\| \leq \varepsilon \quad \Rightarrow \quad \lim_{k\to\infty} \nabla f(x^k) - \sum_{i=1}^{r} J g_i(x^k)^T \mu_i^k = 0, \tag{3.62}$$

where, $\mu_i^k = [\hat{\mu}_i^k - \rho g_i(x^k)]_+$*. Now, we will prove that*

(a) $i \in I_I(x^*) \Rightarrow \mu_i^k = 0,$

(b) $i \in I_B(x^*) \Rightarrow \mu_i^k = 0$ *or* $\mu_i^k \in bd^+(K_{m_i})$ *with* $-\frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|} \to \frac{\overline{g_i(x^*)}}{\|\overline{g_i(x^*)}\|}$.

Similarly to Theorem 17 we have two cases to analyze: the sequence $\{\rho_k\}$ *is bounded or unbounded.*

(a_1) *Let us consider that* $i \in I_I(x^*)$ *and* $\{\rho_k\}$ *is bounded, this is, for* $k \geq k_0$ *we have* $\rho_k = \rho_{k_0}$*. Taking the spectral decomposition for* μ_i^k *as follows*

$$\mu_i^k = [\hat{\mu}_i^k - \rho_{k_0} g_i(x^k)]_+ = \max\{0, \lambda_1^k\} c_1^k + \max\{0, \lambda_2^k\} c_2^k$$

we have by item (i) *of Theorem 17 that*

$$g_i(x^*) = \frac{1}{\rho_{k_0}}\left((\max\{0, \lambda_1^k\} - \lambda_1)c_1 + (\max\{0, \lambda_2^k\} - \lambda_2)c_2\right), \tag{3.63}$$

with $c_1^k \to c_1$ *and* $c_2^k \to c_2$*. Since* $i \in I_I(x^*)$*, we have that the eigenvalues of* $g_i(x^*)$ *are such that*

$$\lambda_1(g_i(x^*)) = \frac{\max\{0, \lambda_1^k\} - \lambda_1}{\rho_k} > 0 \quad \text{and} \quad \lambda_2(g_i(x^*)) = \frac{\max\{0, \lambda_2^k\} - \lambda_2}{\rho_k} > 0.$$

Thus, $\max\{0, \lambda_1^k\} > \lambda_1$ *and* $\max\{0, \lambda_2^k\} > \lambda_2$*. Then,* $\lambda_1 < 0$, $\lambda_2 < 0$ *and* $\lambda_1^k < 0$, $\lambda_2^k < 0$ *for all k large enough. Moreover,* $\lambda_1(\mu_i^k) = \max\{0, \lambda_1^k\} = 0$ *and* $\lambda_2(\mu_i^k) = \max\{0, \lambda_2^k\} = 0$ *for all k large enough. Therefore, we have that* $\mu_i^k = 0$.

(b_1) *Let us consider that $i \in I_B(x^*)$ and $\{\rho_k\}$ is bounded. Take (3.63), since $i \in I_B(x^*)$, we have that the eigenvalues of $g_i(x^*)$ are such that*

$$\lambda_1(g_i(x^*)) = 0 \quad \text{and} \quad \lambda_2(g_i(x^*)) > 0.$$

With argument similar to item (a_1), $\lambda_2(g_i(x^*)) = \dfrac{\max\{0,\lambda_2^k\} - \lambda_2}{\rho_{k_0}} > 0$ implies $\lambda_2(\mu_i^k) = \max\{0,\lambda_1^k\} = 0$ for all k large enough. Thus, we have two possibilities. First,

$$\lambda_2(\mu_i^k) = [\mu_i^k]_0 + \|\overline{\mu_i^k}\| = 0 \quad \text{and} \quad \lambda_1(\mu_i^k) = [\mu_i^k]_0 - \|\overline{\mu_i^k}\| > 0. \tag{3.64}$$

Therefore, $\mu_i^k = 0$. Second,

$$\lambda_2(\mu_i^k) = [\mu_i^k]_0 - \|\overline{\mu_i^k}\| = 0 \quad \text{and} \quad \lambda_1(\mu_i^k) = [\mu_i^k]_0 + \|\overline{\mu_i^k}\| > 0. \tag{3.65}$$

This way, $\mu_i^k \in bd^+(K_{m_i})$. In addition, we have

$$g_i(x^*) = \lambda_1(g_i(x^*))c_1(g_i(x^*)) + \lambda_2(g_i(x^*))c_2(g_i(x^*)), \tag{3.66}$$

such as in (3.63) with $c_1^k \to c_1$ and $c_2^k \to c_2$ for

$$c_1(g_i(x^*)) = \left(\frac{1}{2}, -\frac{\overline{\mu_i^k}}{2\|\overline{\mu_i^k}\|}\right), \quad c_2(g_i(x^*)) = \left(\frac{1}{2}, \frac{\overline{\mu_i^k}}{2\|\overline{\mu_i^k}\|}\right). \tag{3.67}$$

Note that, since $\mu_i^k = [\hat{\mu}_i^k - \rho_{k_0} g_i(x^k)]_+ = \max\{0,\lambda_1^k\}c_1^k + \max\{0,\lambda_2^k\}c_2^k$, we get by (3.65) that our decomposition of μ_i^k is such that

$$c_1(\mu_i^k) = \left(\frac{1}{2}, \frac{\overline{\mu_i^k}}{2\|\overline{\mu_i^k}\|}\right), \quad c_2(\mu_i^k) = \left(\frac{1}{2}, -\frac{\overline{\mu_i^k}}{2\|\overline{\mu_i^k}\|}\right). \tag{3.68}$$

Thus, since that $c_1(\mu_i^k) \to c_1(g_i(x^*))$ and $c_2(\mu_i^k) \to c_2(g_i(x^*))$ we have $-\dfrac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|} \to \dfrac{\overline{g_i(x^*)}}{\|\overline{g_i(x^*)}\|}$. Therefore, $\mu_i^k = 0$ or $\mu_i^k \in bd^+(K_{m_i})$ with $-\dfrac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|} \to \dfrac{\overline{g_i(x^*)}}{\|\overline{g_i(x^*)}\|}$.

(a_2) *Let us consider that $i \in I_I(x^*)$ and $\{\rho_k\}$ is unbounded. Since the sequence $\hat{\mu}_i^k$ is bounded we have that $\dfrac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k) \to -g_i(x^*)$. Let us take a spectral decomposition*

$$\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k) = \lambda_1^k c_1^k + \lambda_2^k c_2^k \tag{3.69}$$

where c_1^k, c_2^k are such as in Theorem 11. Taking a subsequence if necessary, let us consider a spectral decomposition

$$\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k) \to -g_i(x^*) = \lambda_1 c_1 + \lambda_2 c_2, \tag{3.70}$$

where c_1, c_2 are such as in Theorem 11 with $c_1^k \to c_1$, $c_2^k \to c_2$, $\lambda_1^k \to \lambda_1$ and $\lambda_2^k \to \lambda_2$. In addition, we can get a spectral decomposition for μ_i^k using (3.69) given by

$$\mu_i^k = [\hat{\mu}_i^k - \rho_k g_i(x^k)]_+ = \rho_k \max\{0,\lambda_1^k\}c_1^k + \rho_k \max\{0,\lambda_2^k\}c_2^k. \tag{3.71}$$

Then, if $i \in I_I(x^*)$ we have that $\lambda_1(g_i(x^*)) > 0$ and $\lambda_2(g_i(x^*)) > 0$. In particular, $\lambda_1 < 0$ and $\lambda_2 < 0$ since that it is eigenvalues of $-g_i(x^*)$. Thus, $\lambda_1^k < 0$ and $\lambda_2^k < 0$ for all sufficiently large k. Therefore,

$\lambda_1(\mu_i^k) = \rho_k \max\{0, \lambda_1^k\} = 0$ *and* $\lambda_2(\mu_i^k) = \rho_k \max\{0, \lambda_2^k\} = 0$*; hence,* $\mu_i^k = 0$.

(b₂) Let us consider that $i \in I_B(x^*)$ *and* $\{\rho_k\}$ *is unbounded. Similarly to* a_2*, we may take spectral decompositions*

$$\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k) = \lambda_1^k c_1^k + \lambda_2^k c_2^k \tag{3.72}$$

and

$$-g_i(x^*) = \lambda_1 c_1 + \lambda_2 c_2, \tag{3.73}$$

where c_1, c_2 *are such as in Theorem 11 with*

$$c_1^k \to c_1, \quad c_2^k \to c_2, \quad \lambda_1^k \to \lambda_1 \quad \text{and} \quad \lambda_2^k \to \lambda_2.$$

In addition, we can get a spectral decomposition for μ_i^k *using (3.72) given by*

$$\mu_i^k = [\hat{\mu}_i^k - \rho_k g_i(x^k)]_+ = \rho_k \max\{0, \lambda_1^k\} c_1^k + \rho_k \max\{0, \lambda_2^k\} c_2^k. \tag{3.74}$$

If $i \in I_B(x^*)$ *we have that* $\lambda_1(g_i(x^*)) = 0$ *and* $\lambda_2(g_i(x^*)) > 0$*. Note that, by (3.73)*

$$0 = \lambda_1(g_i(x^*)) = -\lambda_1 \quad \text{and} \quad 0 < \lambda_2(g_i(x^*)) = -\lambda_2.$$

Thus, $\lambda_2 < 0$ *and* $\lambda_2^k < 0$ *for all sufficiently large* k*, implying that* $\lambda_2(\mu_i^k) = \rho_k \max\{0, \lambda_2^k\} = 0$*. Moreover, we have two options for* $\lambda_2(\mu_i^k)$*. If*

$$\lambda_2(\mu_i^k) = [\mu_i^k]_0 + \|\overline{\mu_i^k}\| = 0 \quad \text{and} \quad \lambda_1(\mu_i^k) = [\mu_i^k]_0 - \|\overline{\mu_i^k}\| > 0$$

then, $\mu_i^k = 0$*. If*

$$\lambda_2(\mu_i^k) = [\mu_i^k]_0 - \|\overline{\mu_i^k}\| = 0 \quad \text{and} \quad \lambda_1(\mu_i^k) = [\mu_i^k]_0 + \|\overline{\mu_i^k}\| > 0$$

then, $\mu_i^k \in bd^+(K_{m_i})$*. The format of* $\lambda_1(\mu_i^k)$ *and* $\lambda_2(\mu_i^k)$ *suggest that*

$$c_1(\mu_i^k) = \left(\frac{1}{2}, \frac{\overline{\mu_i^k}}{2\|\overline{\mu_i^k}\|}\right), \quad c_2(\mu_i^k) = \left(\frac{1}{2}, -\frac{\overline{\mu_i^k}}{2\|\overline{\mu_i^k}\|}\right). \tag{3.75}$$

However,

$$c_1(g_i(x^*)) = \left(\frac{1}{2}, -\frac{\overline{g_i(x^*)}}{2\|\overline{g_i(x^*)}\|}\right), \quad c_2(g_i(x^*)) = \left(\frac{1}{2}, \frac{\overline{g_i(x^*)}}{2\|\overline{g_i(x^*)}\|}\right). \tag{3.76}$$

Thus, since that $c_1(\mu_i^k) \to c_1(g_i(x^*))$ *and* $c_2(\mu_i^k) \to c_2(g_i(x^*))$ *we have that* $-\frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|} \to \frac{\overline{g_i(x^*)}}{\|\overline{g_i(x^*)}\|}$.

3.7 Conclusions

The well-known sequential optimality conditions introduced in nonlinear programming have proved to be a powerful tool in obtaining results of global convergence of several algorithms and in the unification of developed theory. In this paper, we presented an extension of an Approximate-KKT point to nonlinear symmetric cone programming, which is direct from the definition proposed in [AHV18]. However, the definition for nonlinear second-order cone programming was not straightforward.

A weaker constraint qualification was proposed for NSOCP. From the new constraint qualification, we can prove that our global convergence result for the augmented Lagrangian algorithm generalized the others.

Some interesting questions about the NSOCP that we intended to investigate in the future are: prove that other algorithms for NSOCP generate AKKT points, to extend other sequential optimality condition as Approximate-Gradient-Projection, which is interesting for algorithms where the Lagrange multiplier sequence are not available, study other qualification conditions for NSOCP.

Capítulo 4

On a Conjecture in Second-Order Optimality Conditions

Neste capítulo, vamos apresentar o artigo: *On a Conjecture in Second-Order Optimality Conditions*. O objetivo deste artigo é apresentar uma prova para um caso particular da conjectura feita por R. Andreani, J.M. Martínez, M.L. Schuverdt em [AMS07]. Vamos considerar o problema de programação não linear abaixo

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimizar}} \quad & f(x), \\ \text{sujeito a} \quad & h(x) = 0, \\ & g(x) \leq 0, \end{aligned} \tag{NLP}$$

onde $f:\mathbb{R}^n \to \mathbb{R}$, $h:\mathbb{R}^n \to \mathbb{R}^m$, e $g:\mathbb{R}^n \to \mathbb{R}^p$ são duas vezes continuamente diferenciáveis.

Em geral, análises de convergência de algoritmos em programação não linear utilizam como condição de otimalidade de segunda-ordem o seguinte:

minimizador local $+$ **CQ** $\Rightarrow$ ***WSOC***[1].

WSOC é uma condição de otimalidade onde um ponto KKT a cumpre se existem multiplicadores de Lagrange tais que a matriz Hessiana da função Lagrangiana é semidefinida positiva no subespaço crítico abaixo definido

$$S(x^*) := \{d \in \mathbb{R}^n \mid \nabla h_i(x^*)^T d = 0,\ i = 1,\ldots,,m, \nabla g_j(x^*)^T d = 0, j \in A(x^*)\ \},$$

onde $A(x^*)$ é o conjunto das restrições ativas em x^*. Nosso interesse em WSOC se dá devido seu apelo prático. É sabido que somente MFCQ não é suficiente para garantir WSOC. Podemos encontrar na literatura vários trabalhos cujo objetivo é verificar o que pode ser exigido, em conjunto com MFCQ, para garantir WSOC no ponto. Condições envolvendo informação sobre o posto da matriz Jacobiana se mostraram úteis nessa tarefa. Dentre as condições propostas, podemos destacar o trabalho de Baccari e Trad em [BT05] e Andreani, Martínez e Schuverdt em [AMS07] cujos detalhes serão apresentados na próxima seção.

Andreani, Martínez e Schuverdt conjecturaram que o posto da Jacobiana pode aumentar em no máximo uma unidade na vizinhança do ponto e dessa forma, juntamente a validade de MFCQ no ponto, WSOC é satisfeita. A seguir iremos apresentar uma prova para um caso particular dessa conjectura. Vale citar que, essa conjectura foi provada recentemente por Mascarenhas, detalhes da prova podem ser vistos em [Mas18].

4.1 Article: On a Conjecture in Second-Order Optimality Conditions

In this paper, we deal with a conjecture formulated in [*R. Andreani, J.M. Martínez, M.L. Schuverdt, "On second-order optimality conditions for nonlinear programming", Optimization, 56:529–542, 2007*], which states that whenever a local minimizer of a nonlinear optimization problem fulfills the Mangasarian-

[1] Do inglês: weak second-order condition (WSOC)

Fromovitz Constraint Qualification and the rank of the set of gradients of active constraints increases at most by one in a neighborhood of the minimizer, a second-order optimality condition that depends on one single Lagrange multiplier is satisfied. This conjecture generalizes previous results under a constant rank assumption or under a rank deficiency of at most one. We prove the conjecture under the additional assumption that the Jacobian matrix has a smooth singular value decomposition. Our proof also extends to the case of the strong second-order condition, defined in terms of the critical cone instead of the critical subspace.
Keywords: Nonlinear optimization, Constraint qualifications, Second-order optimality conditions, Singular value decomposition

4.1.1 Introduction

This paper considers a conjecture about second-order necessary optimality conditions for constrained optimization. Our interest in such conjecture comes from practical considerations; see the extended version of this article [BHRV18].

It is well known that under the Mangasarian Fromovitz Constraint Qualification (MFCQ), a second-order optimality condition that depends on the whole set of Lagrange multipliers can be formulated (see Theorem 19). We are interested in additional weak assumptions such that a second-order optimality condition that can be checked with a single Lagrange multiplier holds. This is known as the Weak Second-Order Condition (WSOC), which states that the Hessian of the Lagrangian at a Karush-Kuhn-Tucker (KKT) point, for some Lagrange multiplier, is positive semidefinite on the subspace orthogonal to the gradients of active constraints (critical subspace). We are also interested in the Strong Second-Order Condition (SSOC), which replaces the critical subspace by the larger, and more accurate, critical cone (see equation (4.3)).

Baccari and Trad [BT05] have shown that WSOC holds at a local minimizer, under MFCQ, when the rank of the Jacobian matrix is not smaller than the number of active constraints minus one. In [AMS07], Andreani, Martínez and Schuverdt, with the aim of stating a verifiable condition guaranteeing global convergence of a second-order augmented Lagrangian method to a second-order stationary point, proposed a new condition suitable for that purpose, where the Baccari-Trad condition was replaced by the local constant rank of the Jacobian matrix. There, they conjectured that WSOC would hold at a local minimizer satisfying MFCQ, under the unifying condition that the rank of the Jacobian matrix would locally increase at most by one.

We are aware of three papers dealing with this conjecture. A proof of it under an additional technical condition has appeared in [SXA15], and a counter-example has appeared in [ML16]. As it is shown in the extended technical report [BHRV18], these results are incorrect. Also, the recent paper [Hae17] proves the conjecture for a special form of quadratically constrained problems.

Our approach is based on an additional assumption of smoothness of the singular value decomposition of the Jacobian of the active constraints around the given point.

The paper is organized as follows. In Section 4.1.2, we present our notation and basic definitions. In Section 4.1.3, we prove our main result. In Section 4.1.4, we present some conclusions and future directions of research on this topic.

4.1.2 Basic Definitions and the Statement of the Conjecture

First, we start with the basic notation. The set $\mathbb{R}^n$ stands for the n-dimensional real Euclidean space, $n \in \mathbb{N}$. The set $\mathbb{R}^n_+ \subset \mathbb{R}^n$ is the set of vectors whose components are nonnegative. The canonical basis of $\mathbb{R}^n$ is denoted by $e_1, \dots, e_n$. A set $\mathscr{R} \subset \mathbb{R}^n$ is a ray if $\mathscr{R} = \{rd_0 : r \geq 0\}$ for some vector $d_0 \in \mathbb{R}^n$. We say that $\mathscr{K}$ is a first-order cone if $\mathscr{K}$ is the direct sum of a subspace and a ray.

Now, consider the nonlinear constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x), \\ \text{subject to} & h_i(x) = 0, \ \ \forall i \in \mathscr{E} := \{1, \dots, m\}, \\ & g_j(x) \leq 0, \ \ \forall j \in \mathscr{I} := \{1, \dots, p\}, \end{array} \tag{4.1}$$

where f, h_i, $g_j : \mathbb{R}^n \to \mathbb{R}$ are assumed to be, at least, twice continuously differentiable functions.

Denote by Ω the feasible set of (4.1). For a feasible point $x^* \in \Omega$, we use $A(x^*) := \{j \in \mathscr{I} : g_j(x^*) = 0\}$ to denote the set of indices of active inequalities. The point $x^* \in \Omega$ satisfies the Mangasarian-Fromovitz Constraint Qualification (MFCQ) if $\{\nabla h_i(x^*) : i \in \mathscr{E}\}$ is a linearly independent set, and there is a direction $d \in \mathbb{R}^n$ such that $\nabla h_i(x^*)^{\mathrm{T}} d = 0$, for $i \in \mathscr{E}$ and $\nabla g_j(x^*)^{\mathrm{T}} d < 0$, for $j \in A(x^*)$. Given $x^* \in \Omega$, we define $J(x)$ as the $(m+q) \times n$ matrix whose first m rows are formed by $\nabla h_i(x)^{\mathrm{T}}$, $i \in \mathscr{E}$ and the remaining rows by $\nabla g_j(x)^{\mathrm{T}}, j \in A(x^*)$, where the cardinality of $A(x^*)$ will be denoted by q. We denote the Lagrangian function by $L(x,\lambda,\mu) := f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x)$, where (x,λ,μ) is in $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+$. The symbols $\nabla_x L(x,\lambda,\mu)$ and $\nabla^2_{xx} L(x,\lambda,\mu)$ stand for the gradient and Hessian of $L(x,\lambda,\mu)$ with respect to x, respectively. We denote by $\Lambda(x^*)$ the set of all Lagrange multipliers at x^*. The standard second-order optimality condition under MFCQ alone is the following well known result [BS00, BT80]:

Theorem 19 *Let x^* be a local minimizer of (4.1), with MFCQ holding at x^*. Then,*

$$\forall d \in C(x^*), \exists (\lambda,\mu) \in \Lambda(x^*) \text{ such that } d^{\mathrm{T}} \nabla^2_{xx} L(x^*,\lambda,\mu) d \geq 0, \tag{4.2}$$

where

$$C(x^*) := \left\{ d \in \mathbb{R}^n : \begin{array}{l} \nabla f(x^*)^{\mathrm{T}} d = 0; \nabla h_i(x^*)^{\mathrm{T}} d = 0, i \in \mathscr{E} \\ \nabla g_j(x^*)^{\mathrm{T}} d \leq 0, j \in A(x^*) \end{array} \right\} \tag{4.3}$$

is the critical cone.

We will also make use of Yuan's Lemma [Yua90]:

Lemma 10 (Yuan [Yua90, BT05]) *Let $P, Q \in \mathbb{R}^{n \times n}$ be two symmetric matrices and $K \subset \mathbb{R}^n$ a first-order cone. Then, the following conditions are equivalent:*

- $\max\{d^{\mathrm{T}} P d, d^{\mathrm{T}} Q d\} \geq 0, \ \ \forall d \in K$;
- *There exist $\alpha, \beta \geq 0$ with $\alpha + \beta = 1$ such that $d^{\mathrm{T}}(\alpha P + \beta Q) d \geq 0$, $\forall d \in K$.*

Now, we proceed to state the conjecture from [AMS07, Section 5].

Let x^* be a local minimizer of (4.1). Assume that MFCQ holds at x^* and that the rank of $J(x)$ is at most $r+1$ in a neighborhood of x^*, where r is the rank of $J(x^*)$. Then, there exists a Lagrange multiplier $(\lambda,\mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$ such that

$$\nabla_x L(x^*,\lambda,\mu) = 0 \quad \text{with} \quad \mu_j g_j(x^*) = 0, \forall j, \tag{4.4}$$

and for every $d \in S(x^*)$, we have

$$d^{\mathrm{T}} \nabla^2_{xx} L(x^*,\lambda,\mu) d \geq 0, \tag{4.5}$$

where $S(x^*) := \{d \in \mathbb{R}^n : \nabla h_i(x^*)^{\mathrm{T}} d = 0, \forall i; \nabla g_j(x^*)^{\mathrm{T}} d = 0, \forall j \in A(x^*)\}$ is the critical subspace (or weak critical cone).

Equations (4.4)-(4.5) are known as the Weak Second-Order Condition (WSOC) and its main relevance appears in global convergence of second-order algorithms. See the detailed discussion in [BHRV18].

4.1.3 Main Result

In this section, we prove the conjecture under an additional assumption based on the smoothness of the singular value decomposition of the Jacobian matrix.

Let us first show that Baccari and Trad's result [BT05] can be generalized in order to consider column-rank deficiency. The proof is a simple application of the rank-nullity theorem, namely, that the rank of a matrix is equal to the co-dimension of its kernel.

Theorem 20 *Let x^* be a local minimizer of (4.1), such that MFCQ holds and the rank of the Jacobian matrix $J(x^*) \in \mathbb{R}^{(m+q) \times n}$ is $n-1$. Then, there exists a Lagrange multiplier $(\lambda,\mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$ such that WSOC holds.*

Proof 17 *Applying the rank-nullity theorem to $S(x^*) = Ker(J(x^*))$, we get that $dim(S(x^*)) = 1$. Hence, there is a $d_0 \in S(x^*)$ such that $S(x^*) = \{td_0 : t \in \mathbb{R}\}$. Since MFCQ holds, Theorem 19 yields (4.2). In particular, for $d := d_0$, there is a Lagrange multiplier (λ,μ) such that $d_0^{\mathrm{T}}\nabla^2_{xx}L(x^*,\lambda,\mu)d_0 \geq 0$. This same Lagrange multiplier can be used for all other directions $d := td_0 \in S(x^*)$, since $d^{\mathrm{T}}\nabla^2_{xx}L(x^*,\lambda,\mu)d = t^2 d_0^{\mathrm{T}}\nabla^2_{xx}L(x^*,\lambda,\mu)d_0 \geq 0$. Thus, WSOC holds at x^*.*

We now present our main result. We prove the conjecture under an additional technical assumption on the smoothness of the singular value decomposition (SVD) of the Jacobian matrix around x^*.

Assumption 2 *Let $J(x) \in \mathbb{R}^{(m+q)\times n}$ be the Jacobian matrix for x near x^*. We assume that there exist differentiable functions, around x^*, given by*

$$x \mapsto U(x) \in \mathbb{R}^{(m+q)\times(m+q)}, \;\; x \mapsto \Sigma(x) \in \mathbb{R}^{(m+q)\times n} \text{ and } x \mapsto V(x) \in \mathbb{R}^{n\times n},$$

such that $J(x) = U(x)\Sigma(x)V(x)^{\mathrm{T}}$, where $\Sigma(x)$ is diagonal with diagonal elements $\sigma_1(x), \sigma_2(x), \ldots, \sigma_k(x)$, where $k := \min\{m+q, n\}$ and $\sigma_i(x) = 0$ when i is greater than the rank of $J(x)$. We assume also that $U(x^*)$ and $V(x^*)$ are matrices with non-zero orthogonal columns.

Note that only at $x = x^*$ we assume that $U(x^*)$ and $V(x^*)$ are matrices with orthogonal columns. This implies that $U(x)$ and $V(x)$ are at least invertible matrices in a small enough neighborhood of x^*, but not necessarily with orthogonal columns.

Theorem 21 *Assume that x^* is a local minimizer of (4.1) that fulfills MFCQ. Let r be the rank of $J(x^*)$, and assume that for every x in some neighborhood of x^*, the rank of $J(x)$ is at most $r+1$. Suppose also that Assumption 2 holds. Then, there is a Lagrange multiplier (λ,μ) such that WSOC holds.*

Proof 18 *Let us consider the column functions $U(x) = [u_1(x) \ldots u_{m+q}(x)]$ and $V(x) = [v_1(x) \ldots v_n(x)]$. Clearly, $J(x) = \sum_{k=1}^{r+1} \sigma_k(x)u_k(x)v_k(x)^{\mathrm{T}}$. To simplify the notation, let us assume that $m = 0$ and $A(x^*) = \{1,\ldots,p\}$, that is, $q = p$, hence, it holds that*

$$\nabla g_i(x)^{\mathrm{T}} = \sum_{k=1}^{r+1} \sigma_k(x)[u_k(x)]_i v_k(x)^{\mathrm{T}}, \;\; i = 1,\ldots,p,$$

where $[u]_i$ is the i-th coordinate of the vector u of appropriate dimension.

Since the functions are smooth, we can compute derivatives of $\nabla g_i(x)$ to get, for $i = 1,\ldots,p$, that

$$\nabla^2 g_i(x) = \sum_{k=1}^{r+1} \sigma_k(x)[u_k(x)]_i J_{v_k}(x) + \sum_{k=1}^{r+1} ([u_k(x)]_i \nabla\sigma_k(x) + \sigma_k(x)\nabla[u_k(x)]_i) v_k(x)^{\mathrm{T}},$$

where $J_{v_k}(x) \in \mathbb{R}^{n\times n}$ is the Jacobian matrix of the function v_k at x.

Now, let us fix a direction $d \in S(x^*)$ and a Lagrange multiplier $\mu \in \Lambda(x^*)$ (we identify a Lagrange multiplier (λ,μ) with μ since we are assuming $m = 0$). Now, we proceed to evaluate $d^{\mathrm{T}}\nabla^2_{xx}L(x^*,\mu)d$. We omit the dependency on x^* for simplicity. Then, we have that $d^{\mathrm{T}}\nabla^2_{xx}L(x^*,\mu)d$ is equal to

$$\begin{array}{rl} & d^{\mathrm{T}}\nabla^2 f d + d^{\mathrm{T}}\left[\sum_{i=1}^{p}\mu_i\sum_{k=1}^{r+1}([u_k]_i\nabla\sigma_k + \sigma_k\nabla[u_k]_i)v_k^{\mathrm{T}} + \sigma_k[u_k]_i J_{v_k}\right]d \\ = & d^{\mathrm{T}}\nabla^2 f d + \sum_{k=1}^{r+1}[(d^{\mathrm{T}}\nabla\sigma_k)(\mu^{\mathrm{T}}u_k) + \sigma_k d^{\mathrm{T}}J_{u_k}^{\mathrm{T}}\mu]v_k^{\mathrm{T}}d + \sum_{k=1}^{r+1}\sigma_k(\mu^{\mathrm{T}}u_k)d^{\mathrm{T}}J_{v_k}d. \end{array} \tag{4.6}$$

From $S(x^*) = Ker(J(x^*))$, and from the SVD, $J(x) = U(x)\Sigma(x)V(x)^{\mathrm{T}}$, we can conclude that there are $s_{r+1},\ldots,s_n \in \mathbb{R}$ such that $d = \sum_{j=r+1}^{n} s_j v_j$. Hence, from the orthogonality of $\{v_1,\ldots,v_p\}$ (which holds at $x = x^*$ from our assumption), we get $v_k^{\mathrm{T}}d = 0$, $k < r+1$. Furthermore, since $\sigma_{r+1} = 0$, we obtain that

$$d^{\mathrm{T}}\nabla^2_{xx}L(x^*,\mu)d = d^{\mathrm{T}}\nabla^2 f d + (d^{\mathrm{T}}\nabla\sigma_{r+1})(\mu^{\mathrm{T}}u_{r+1})(v_{r+1}^{\mathrm{T}}d) + \sum_{k=1}^{r}\sigma_k(\mu^{\mathrm{T}}u_k)d^{\mathrm{T}}J_{v_k}d. \tag{4.7}$$

For a fixed Lagrange multiplier $\bar{\mu} \in \Lambda(x^*)$ (note that MFCQ ensures non-emptyness of $\Lambda(x^*)$), we can write $\Lambda(x^*) = (\bar{\mu} + Ker(J(x^*)^{\mathrm{T}})) \cap \mathbb{R}^p_+$, hence, there are $t_{r+1}, \dots, t_p \in \mathbb{R}$ such that $\mu = \bar{\mu} + \sum_{j=r+1}^{p} t_j u_j$, and we can write (4.7) as

$$d^{\mathrm{T}} \nabla^2_{xx} L(x^*, \mu) d = d^{\mathrm{T}} \nabla^2 f d + (d^{\mathrm{T}} \nabla \sigma_{r+1})(\bar{\mu}^{\mathrm{T}} u_{r+1} + t_{r+1})(v_{r+1}^{\mathrm{T}} d) + \sum_{k=1}^{r} \sigma_k (\bar{\mu}^{\mathrm{T}} u_k) d^{\mathrm{T}} J_{v_k} d. \tag{4.8}$$

Observe that for a fixed $d \in S(x^*)$, the value of $d^{\mathrm{T}} \nabla^2_{xx} L(x^*, \mu) d$, for $\mu \in \Lambda(x^*)$, depends on a single parameter t_{r+1}. Since MFCQ holds, condition (4.2) holds, and we may write it as

$$\max_{\mu \in \Lambda(x^*)} d^{\mathrm{T}} \nabla^2_{xx} L(x^*, \mu) d \geq 0, \ \ \forall d \in S(x^*). \tag{4.9}$$

In virtue of the fact that the value of $d^{\mathrm{T}} \nabla^2_{xx} L(x^*, \mu) d$ depends only on one parameter, in order to apply Yuan's Lemma, we will rewrite (4.9) as a maximization problem over a line segment. For that purpose, we define the set

$$M := \{(t_{r+1}, t_{r+2}, \dots, t_p) \in \mathbb{R}^{p-r} \text{ such that } \bar{\mu} + \sum_{j=r+1}^{p} t_j u_j \in \Lambda(x^*)\}, \tag{4.10}$$

and the following optimization problems

$$a_* := \ inf\{t_{r+1} : (t_{r+1}, t_{r+2}, \dots, t_p) \in M\}, \tag{4.11}$$

and

$$b_* := \ sup\{t_{r+1} : (t_{r+1}, t_{r+2}, \dots, t_p) \in M\}. \tag{4.12}$$

The set M is a compact convex set. Indeed, it is easy to see that M is convex and closed, since $\Lambda(x^*)$ is also convex and closed. To show that M is bounded, suppose by contradiction that there is a sequence $(t^k_{r+1}, \dots, t^k_p) \in M$ with $T_k := \max\{|t^k_{r+1}|, \dots, |t^k_p|\} \to \infty$. Since $\Lambda(x^*)$ is bounded, there is a scalar K such that $\|\bar{\mu} + \sum_{j=r+1}^{p} t^k_j u_j\| \leq K$, $\forall k \in \mathbb{N}$. Dividing this expression by T_k and taking an adequate subsequence, we conclude that there are $\bar{t}_{r+1}, \dots, \bar{t}_p$, not all zero, such that $\sum_{j=r+1}^{p} \bar{t}_j u_j = 0$, which is a contradiction with the linear independence of $\{u_{r+1}, \dots, u_p\}$. Therefore, both values a_* and b_* are finite and attained.

Denote by $\theta(t_{r+1}, d) := d^{\mathrm{T}} \nabla^2 f d + (\bar{\mu}^{\mathrm{T}} u_{r+1} + t_{r+1})(d^{\mathrm{T}} \nabla \sigma_{r+1})(v^{\mathrm{T}}_{r+1} d) + \sum_{k=1}^{r} \sigma_k (\bar{\mu}^{\mathrm{T}} u_k) d^{\mathrm{T}} J_{v_k} d$. From (4.8), $\theta(t_{r+1}, d)$ coincides with $d^{\mathrm{T}} \nabla^2_{xx} L(x^*, \mu) d$, whenever $\mu = \bar{\mu} + \sum_{j=r+1}^{p} t_j u_j$.

Now, consider the optimization problem:

$$\max\{\theta(t_{r+1}, d) : t_{r+1} \in [a_*, b_*]\}. \tag{4.13}$$

From (4.9), we get $\max\{\theta(t_{r+1}, d) : t_{r+1} \in [a_*, b_*]\} \geq 0$ for all $d \in S(x^*)$. Observe that $\theta(t, d)$ is linear in t, and for each t fixed, it defines a quadratic form as function of d. Then, since $\theta(t, d)$ is linear in t, the maximum of (4.13) is attained either at $t_{r+1} = a_*$ or $t_{r+1} = b_*$. For simplicity, let us call the quadratic forms $\theta(a_*, d)$ and $\theta(b_*, d)$ by $d^{\mathrm{T}} P d$ and $d^{\mathrm{T}} Q d$, respectively. Thus, we arrive at

$$\max\{d^{\mathrm{T}} P d, d^{\mathrm{T}} Q d\} \geq 0, \ \ \forall d \in S(x^*). \tag{4.14}$$

Applying Yuan's Lemma (Lemma 10), we get the existence of $\alpha \geq 0$ and $\beta \geq 0$, with $\alpha + \beta = 1$, such that

$$d^{\mathrm{T}} (\alpha P + \beta Q) d \geq 0, \ \ \forall d \in S(x^*). \tag{4.15}$$

Additionally, due to the linearity of $\theta(t, d)$, we see that $\eta := \alpha a_* + \beta b_*$ satisfies $\theta(\eta, d) = d^{\mathrm{T}} (\alpha P + \beta Q) d$. Denote $\pi_{r+1}(t_{r+1}, \dots, t_p) := t_{r+1}$ the projection onto the first coordinate. From the continuity of π_{r+1} and the compactness and convexity of M, we get $[a_*, b_*] = \pi_{r+1}(M)$. Since $\eta \in [a_*, b_*]$, we conclude that there are some scalars $\hat{t}_{r+1}, \dots, \hat{t}_p$ with $\hat{t}_{r+1} = \eta$ and $\hat{\mu} := \bar{\mu} + \sum_{j=r+1}^{p} \hat{t}_j u_j \in \Lambda(x^*)$, such that $\theta(\eta, d) = d^{\mathrm{T}} (\alpha P + \beta Q) d = d^{\mathrm{T}} \nabla^2_{xx} L(x^*, \hat{\mu}) d$. From (4.15), we get that $\hat{\mu}$ is a Lagrange multiplier for which WSOC holds at x^*.

The problem of minimizing x_3, subject to $\cos(x_1 + x_2) - x_3 - 1 \leq 0$, $-\cos(x_1 + x_2) - x_3 + 1 \leq 0$ and

$-2x_3 \leq 0$, at the local minimizer $x^* := (0,0,0)$, shows that Theorem 21 applies for obtaining WSOC, while previous results do not. See details in [BHRV18, Example 3.2].

Note that our proof suggests that when the rank is constant, the Hessian of the Lagrangian in the critical subspace does not depend on the Lagrange multiplier. In fact, this can be proved without additional assumptions. See the extended technical report [BHRV18]. This explains why constant rank assumptions imply the validity of second-order conditions at any Lagrange multiplier (see [AES10]).

Despite our focus on conditions implying WSOC, the above analysis allows us to obtain conclusions about SSOC, related with [BT05, Theorem 5.1]. Recall that the *generalized strict complementary slackness* (GSCS) condition holds at the feasible point x^* if there exists, at most, one index $i_0 \in A(x^*)$ such that $\mu_{i_0} = 0$ whenever $(\lambda, \mu) \in \Lambda(x^*)$.

Theorem 22 *Assume that x^* is a local minimizer of (4.1) that fulfills MFCQ. Let r be the rank of $J(x^*)$, and assume that for every x in some neighborhood of x^*, the rank of $J(x)$ is at most $r+1$. Suppose also that Assumption 2 and GSCS hold at x^*. Then, there is a Lagrange multiplier (λ, μ) such that SSOC holds.*

Proof 19 *From [BT05, Theorem 5.1] or [ABHS17, Definition 3.3 and Lemma 3.3], it follows that GSCS implies that the critical cone $C(x^*)$, from equation (4.3), is a first-order cone. Hence, we can still apply Yuan's Lemma and prove the result in the same lines of Theorem 21.*

Condition SSOC is relevant beyond the scope of this paper, since it is associated with a sufficient optimality condition (given by replacing positive semidefiniteness with positive definiteness in its definition). Checking the sufficient version of SSOC is easier than checking so-called "no-gap" conditions (see discussion in [BHRV18]), since it depends on a single Lagrange multiplier, rather than on the whole set of Lagrange multipliers. Weak conditions ensuring the necessity of SSOC, as in Theorem 22, reduces the gap with respect to its sufficient counterpart. Other conditions implying the validity of SSOC at a local minimizer are the ones defined in terms of variations of the constant rank constraint qualification. See [ML16, AES10, MS11]. This approach can be extended by considering Abadie-type assumptions, namely, assuming that the critical cone is a subset of the tangent cone of a modified feasible set, as in [ABHS17, Bom16]. Also, SSOC has been proved under GSCS in [BT05], when the set of Lagrange multipliers is a bounded line segment. This has been extended to a larger set in [AMS07], namely, when the set of Hessians of the Lagrangian at the vertices of the bounded set of Lagrange multipliers has rank at most two.

4.1.4 Conclusions

In order to analyse limit points of a sequence generated by a second-order algorithm, one usually relies on WSOC, the stationarity concept based on the critical subspace. Most conditions guaranteeing WSOC at local minimizers are based on a constant rank assumption on the Jacobian matrix. In this paper we developed new tools to deal with the non-constant rank case, by partially solving a conjecture formulated in [AMS07]. We also presented a condition ensuring the validity of SSOC. Possible future lines of research include investigating the full conjecture using generalized notions of derivative.

Acknowledgements

This research was partially conducted while the second author held a Visiting Scholar position at Department of Management Science and Engineering, Stanford University, Stanford CA, USA. The fourth author was a PhD student at Department of Applied Mathematics, University of São Paulo-SP, Brazil. We thank the referees and the handling editor for insightful remarks. This research was funded by FAPESP grants 2013/05475-7 and 2016/02092-8, by CNPq grants 454798/2015-6, 303264/2015-2 and 481992/2013-8, and CAPES.

Capítulo 5

Conclusões

O problema de programação não linear foi muito estudado nos últimos anos; os avanços são inúmeros e podem ser acompanhados na vasta bibliografia que trata do tema. Entretanto, problemas tais como NSDP, NSOCP e outros cujas restrições pertencem a cones simétricos têm sido estudados mais ativamente nas duas últimas décadas e ainda de forma preliminar. As condições sequenciais de otimalidade são bem conhecidas em NLP por permitirem provar resultados de convergência global sob hipóteses fracas para algoritmos de primeira e segunda-ordem.

Nesta tese, estendemos a conhecida condição sequencial AKKT para o problema NSDP. Também apresentamos uma nova condição de otimalidade chamada Trace-AKKT (TAKKT) que se mostrou mais prática e simples dentro do contexto de programação semidefinida não linear. Essa nova condição de otimalidade pode ser definida em NLP, porém sua relevância neste contexto ainda está para ser investigada.

O problema NSCP é conhecido por unificar a teoria de vários problemas relevantes em otimização tais como os problemas citados anteriormente. Neste trabalho, apresentamos uma extensão da condição AKKT para essa classe de problemas. A condição AKKT definida para NSCP possui as mesmas propriedades que AKKT em NSDP. Em relação a condições de otimalidade de segunda-ordem, apresentamos uma prova para um caso particular da conjectura feita em [AMS16], cuja prova completa apareceu posteriormente em [Mas18].

Como pesquisas futuras pretendemos analisar a teoria de condições sequenciais para cones gerais. Ainda como possibilidade de pesquisa futura, durante o desenvolvimento do artigo: *Optimality conditions for non-linear symmetric cone programming* pretendemos realizar um estudo mais aprofundado sobre as condições CRCQ, RCRCQ e CRSC propostas em [ZZ18] para NSOCP. Acreditamos que seja possível definir tais condições de modo a coincidirem com a definição usual no caso de NLP e de modo que garantam que um ponto AKKT seja KKT.

Referências Bibliográficas

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

[ABMS07] R. Andreani, E. G. Birgin, J. M. Martínez e M. L. Schuverdt. On augmented Lagrangian methods with general lower-level constraint. *SIAM Journal on Optimization*, 18(4):1286–1309, 2007.

[ABMS08] R. Andreani, E. G. Birgin, J. M. Martínez e M. L. Schuverdt. Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. *Mathematical Programming*, 111:5–32, 2008.

[AES10] R. Andreani, C. E. Echagüe e M. L. Schuverdt. Constant-rank condition and second-order constraint qualification. *Journal of Optimization theory and Applications*, 146:2:255–266, 2010.

[AFSS19] R. Andreani, N. S. Fazzio, M. L. Schuverdt e L. D. Secchin. A sequential optimality condition related to the quasinormality constraint qualification and its algorithmic consequences. *SIAM Journal on Optimization, to appear*, 2019.

[AG03] F. Alizadeh e D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.

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

[AHRS17] R. Andreani, G. Haeser, A. Ramos e 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.

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

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

[AHSS12c] R. Andreani, G. Haeser, M.L. Schuverdt e P.J.S. Silva. A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programming*, 135(1-2):255–273, 2012.

[AHSS18] R. Andreani, G. Haeser, L.D. Secchin e P.J.S. Silva. New sequential optimality conditions for mathematical problems with complementarity constraints and algorithmic consequences. *Optimization-Online*, 2018.

[AHV18] R. Andreani, G. Haeser e D.S. Viana. Optimality conditions and global convergence for nonlinear semidefinite programming. *Mathematical Programming*, DOI: 10.1007/s10107-018-1354-5, 2018.

[AMRS16] R. Andreani, J. M. Martínez, A. Ramos e P. J. S. Silva. A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization*, 26(1):96–110, 2016.

[AMRS18] R. Andreani, J. M. Martínez, A. Ramos e P. J. S. Silva. Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Mathematics of Operations Research*, 43(3):693–717, 2018.

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

[AMS10] R. Andreani, J. M. Martínez e B. F. Svaiter. A new sequencial optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization*, 20(6):3533–3554, 2010.

[AMS16] R. Andreani, J.M. Martínez e L.T. Santos. Newton's method may fail to recognize proximity to optimal points in constrained optimization. *Mathematical Programming*, 160:547–555, 2016.

[ART03] E.D. Andersen, C. Roos e T. Terlaky. On implementing a primal-dual interior-point method for conic quadratic programming. *Mathematical Programming*, 95:249–277, 2003.

[ASS18] R. Andreani, L.D. Secchin e P.J.S. Silva. Convergence properties of a second order augmented Lagrangian method for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 28:2574–2600, 2018.

[Bae07] M. Baes. Convexity and differentiability properties of spectral functions and spectral mappings on Euclidean Jordan algebras. *Linear Algebra and its Applications*, 422(2):664–700, 2007.

[BDL07] J. Bolte, A. Daniilidis e A. S. Lewis. The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, 17(4):1205–1223, 2007.

[BGM$^+$16] E. G. Birgin, J. L. Gardenghi, J. M. Martínez, S. A. Santos e Ph. L. Toint. Evaluation complexity for nonlinear constrained optimization using unscaled KKT conditions and high-order models. *SIAM Journal on Optimization*, 26:951–967, 2016.

[BHFR18] L. F. Bueno, G. Haeser, F.Lara e F. N. Rojas. Augmented Lagrangian method for the solution of quasi-equilibrium problems. *Optimization-Online*, 2018.

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

[BHR19] L.F. Bueno, G. Haeser e F.N. Rojas. Optimality conditions and constraint qualifications for Generalized Nash Equilibrium Problems and their practical implications. *SIAM Journal on Optimization*, 29(1):31–54, 2019.

[BHRV18] R. Behling, G. Haeser, A. Ramos e D. Viana. On a conjecture in second-order optimality conditions. *Journal of Optimization Theory and Applications*, 176(3):625–633, 2018. Extended Technical Report arXiv:1706.07833.

[BKM17] E. G. Birgin, N. Krejic e J. M. Martínez. On the minimization of possibly discontinuous functions by means of pointwise approximations. *Optimization Letters*, 11(8):1623–1637, 2017.

[BM14] E. Birgin e J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. SIAM Publications, 2014.

[Bom16] I. Bomze. Copositivity for second-order optimality conditions in general smooth optimization problems. *Optimization*, 65(4):779–795, 2016.

[BS00] J.F. Bonnans e A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer Verlag, New York, 2000.

[BSS06] M. S. Bazaraa, H. D. Sherali e C. M. Shetty. *Practical methods of Optimization: theory and algorithms*. John Wiley and Sons, New Jersey, USA, 2006.

[BT80] A. Ben-Tal. Second-order and related extremality conditions in nonlinear programming. *Journal of Optimization Theory and Applications*, 31:2, 1980.

[BT05] A. Baccari e A. Trad. On the classical necessary second-order optimality conditions in the presence of equality and inequality constraints. *SIAM Journal on Optimization*, 15(2):394–408, 2005.

[Buy72] J. D. Buys. *Dual algorithms for constrained optimization problems*. Tese de Doutorado, University of Leiden, The Netherlands, 1972.

[CR04] R. Correa e H. Ramírez. A global algorithm for nonlinear semidefinite programming. *SIAM Journal on Optimization*, 15(1):303–318, 2004.

[cS03] M. Kočvara e M. Stingl. PENNON - a generalized augmented lagrangian method for semidefinite programming. *High performance algorithms and software for nonlinear optimization (eds. G. Di Pillo and A. Murli),Kluwer Academic Publishers*, páginas 297–315, 2003.

[cS07] M. Kočvara e M. Stingl. On the solution of large-scale SDP problems by the modified barrier method using iterative solvers. *Mathematical Programming*, 109:413–444, 2007.

[cS10] M. Kočvara e M. Stingl. PENNON - a code for convex nonlinear and semidefinite programming. *Optimization Methods and Software*, 18(3):317–333, 2010.

[CSS03] X.D. Chen, D. Sun e J. Sun. Complementarity functions and numerical experiments on some smoothing Newton methods for second-order-cone complementarity problems. *Computational optimization and applications*, 25:39–56, 2003.

[DDTA13] J. Dutta, K. Deb, R. Tulshyan e R. Arora. Approximate KKT points and a proximity measure for termination. *Journal Global Optimization*, 56(4):1463–1499, 2013.

[FAN01] B. Fares, P. Apkarian e 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.

[FJV07] R.W. Freund, F. Jarre e C.H. Vogelbusch. Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling. *Mathematical Programming*, 109:581–611, 2007.

[FK94] J. Faraut e A. Korányi. *Analysis on Symmetric Cone*. Oxford mathematical monograph, Clarendon Press, Oxfor, 1994.

[FLT01] M. Fukushima, Z-Q. Luo e P. Tseng. Smoothing functions for second-order-cone complementarity problems. *SIAM Journal on Optimization*, 12(2):436–460, 2001.

[FM68] A. V. Fiacco e G. P. McCormick. *Nonlinear Programming Sequential Unconstrained Minimization Techniques*. John Wiley, New York, 1968.

[FNA02] B. Fares, D. Noll e P. Apkarian. Robust control via sequential semidefinite programming. *SIAM Journal on Control and Optimization*, 40(6):1791–1820, 2002.

[For00] A. Forsgren. Optimality conditions for nonconvex semidefinite programming. *Mathematical Programming*, 88(1):105–128, 2000.

[GJN16] G. Giorgi, B. Jiménez e V. Novo. Approximate Karush-Kuhn-Tucker condition in multiobjective optimization. *Journal of Optimization Theory and Applications*, 171(1):70–89, 2016.

[GR10] W. Gómez e H. Ramírez. A filter algorithm for nonlinear semidefinite programming. *Computational and Applied Mathematics*, 29(2):297–328, 2010.

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

[Hae18] 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.

[HJ91] R.A. Horn e C.R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, 1991.

[HM15] G. Haeser e V.V. Melo. Convergence detection for optimization algorithms: approximate-KKT stopping criterion when Lagrange multipliers are not available. *Operations Research Letters*, 43(5):484–488, 2015.

[HS11] G. Haeser e M.L. Schuverdt. On approximate KKT condition and its extension to continuous variational inequalities. *Journal of Optimization Theory and Applications*, 149(3):528–539, 2011.

[HTY06] X.X. Huang, K.L. Teo e X.Q. Yang. Approximate augmented lagrangian functions and nonlinear semidefinite programs. *Acta Mathematica Sinica*, 22(5):1283–1296, 2006.

[Jan84] R. Janin. Direction derivative of the marginal function in nonlinear programming. *Mathematical Programming Studies*, 21:127–138, 1984.

[Jar12] F. Jarre. Elementary optimality conditions for nonlinear SDPs. Em *Handbook on Semidefinite, Conic and Polynomial Optimization*, Spring, 2012. International Series in Operations Research & Management Science.

[KKW03] H. Konno, N. Kawadai e D. Wu. Estimation of failure probability using semi-definite Logit model. *Computational Management Science*, 1(1):59–73, 2003.

[KS18] D. Kanzow e D. Steck. Approximated Lagrangian and exact penalty methods for quasi-variational inequalities. *Computational Optimization and Applications*, 60(3):801–824, 2018.

[KSW18] C. Kanzow, D. Steck e D. Wachsmuth. An augmented Lagrangian methods for optimization problems in Banach spaces. *SIAM Journal Control Optimization*, 56(1):272–291, 2018.

[KT06] Y. Kanno e I. Takewaki. Sequential semidefinite program for maximum robustness design of structures under load uncertainty. *Journal of Optimization Theory and Applications*, 130:265–287, 2006.

[Lew93] A. S. Lewis. Convex analysis on the Hermitian matrices. *SIAM Journal on Optimization*, 6(1):164–177, 1993.

[LFF16] B. F. Lourenço, E. H. Fukuda e M. Fukushima. Optimality conditions for nonlinear semidefinite programming via squared slack variables. *Mathematical Programming*, 166:1–24, 2016.

[LFF18] B. F. Lourenço, E. H. Fukuda e M. Fukushima. Optimality conditions for problems over symmetric cones and a simple augmented Lagrangian method. *Mathematics of Operations Research*, 43(4):1233–1251, 2018.

[Lov03] L. Lovász. *Semidefinite Programs and Combinatorial Optimization*. Springer, New York, 2003.

[LWC12] H.Z. Luo, H.X. Wu e G.T. Chen. On the convergence of augmented lagrangian methods for nonlinear semidefinite programming. *Journal of Global Optimization*, 54(3):599–618, 2012.

[LZ08] Y.J. Liu e L.W. Zhang. Convergence of the augmented Lagrangian method for nonlinear optimization problems over second-order cone. *Journal of Optimization Theory and Applications*, 139(3):557–575, 2008.

[Mas18] W. Mascarenhas. A simple canonical form for nonlinear programming problems and its use. *Journal of Optimization Theory and Applications*, DOI: 10.1007/s10957-018-1381-7, 2018.

[ML16] L. Minchenko e A. Leschov. On strong and weak second-order necessary optimality conditions for nonlinear programming. *Optimization*, 65(9):1693–1702, 2016.

[MP00] J. M. Martínez e E.A. Pilotta. Inexact restoration algorithm for constrained optimization. *Journal of Optimization Theory and Applications*, 104(1):135–163, 2000.

[MS03] J. M. Martínez e B. F. Svaiter. A practical optimality condition without constraint qualifications for nonlinear programming. *Journal of Optimization Theory and Applications*, 118(1):117–133, 2003.

[M.S05] M.Stingl. *On the Solution of Nonlinear Semidefinite Programs by Augmented Lagrangian Methods*. Tese de Doutorado, University of Erlangen, 2005.

[MS11] L. Minchenko e S. Stakhovski. Parametric nonlinear programming problems under the relaxed constant rank condition. *SIAM Journal on Optimization*, 21:314–332, 2011.

[QS06] H. Qi e D. Sun. A quadratically convergent newton method for computing the nearest correlation matrix. *SIAM Journal on Matrix Analysis and Applications*, 28(2):360–385, 2006.

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

[Ram16] A. Ramos. Mathematical programs with equilibrium constraints: A sequential optimality condition, new constraint qualifications and algorithmic consequences. *Optimization-online*, 2016.

[ScL09] M. Stingl, M. Kočvara e G. Leugering. A sequential convex semidefinite programming algorithm with an application to multiple-load free material optimization. *SIAM Journal on Optimization*, 20(1):130–155, 2009.

[Sha97] A. Shapiro. First and second order analysis of nonlinear semidefinite programs. *SIAM Journal on Optimization*, 77(1):301–320, 1997.

[SS04] A. Shapiro e J. Sun. Some properties of the augmented Lagrangian in cone constrained optimization. *Mathematics of Operations Research*, 29:479–491, 2004.

[SSZ08] D. Sun, J. Sun e L. Zhang. The rate of convergence of the augmented lagrangian method for nonlinear semidefinite programming. *Mathematical Programming*, 114(2):349–391, 2008.

[SXA15] C. Shen, W. Xue e Y. An. A new result on second-order necessary conditions for nonlinear programming. *Operations Research Letters*, 43:117–122, 2015.

[SZW06] J. Sun, L.W. Zhang e Y. Wu. Properties of the augmented Lagrangian in nonlinear semidefinite optimization. *Journal of Optimization Theory and Applications*, 12(3):437–456, 2006.

[The75] C. M. Theobald. An inequality for the trace of the product of two symmetric matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 77(2):265–267, 1975.

[Tod03] M. J. Todd. Semidefinite optimization. *Acta Numerica*, 10:515–560, 2003.

[TYW17] N. V. Tuyen, J. Yao e C. Wen. A note on approximate Karush-Kuhn-Tucker conditions in locally Lipschitz multiobjective optimization. *ArXiv:1711.08551*, 2017.

[VB96] L. Vandenberghe e S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):549–95, 1996.

[VBW98] L. Vandenberghe, S. Boyd e S.P. Wu. Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, 19(2):499–533, 1998.

[Yua90] Y. Yuan. On a subproblem of trust region algorithms for constrained optimization. *Mathematical Programming*, 47:53–63, 1990.

[YY12] H. Yamashita e H. Yabe. Local and superlinear convergence of a primal-dual interior point method for nonlinear semidefinite programming. *Mathematical Programming*, 132(1-2):1–30, 2012.

[YY15] H. Yamashita e H. Yabe. A survey of numerical methods for nonlinear semidefinite programming. *Journal of the Operations Research Society of Japan*, 58(1):24–60, 2015.

[YYH12] H. Yamashita, H. Yabe e K. Harada. A primal-dual interior point method for nonlinear semidefinite programming. *Mathematical Programming*, 135(1-2):89–121, 2012.

[ZZ14] Z.B. Zhu e H.L. Zhu. A filter method for nonlinear semidefinite programming with global convergence. *Acta Mathematica Sinica*, 30(10):1810–1826, 2014.

[ZZ18] Y. Zhang e L. Zhang. New constraint qualifications and optimality conditions for second order cone programs. *Set-Valued and Variational Analysis*, DOI: 10.1007/s11228-018-0487-2, 2018.