UNIVERSIDADE FEDERAL DO PARANÁ

EVELIN HERINGER MANOEL KRULIKOVSKI

MATHEMATICAL PROGRAMS WITH CARDINALITY CONSTRAINTS: A UNIFIED APPROACH FOR WEAK STATIONARITY CONDITIONS AND A SEQUENTIAL OPTIMALITY CONDITION

CURITIBA

2021

EVELIN HERINGER MANOEL KRULIKOVSKI

MATHEMATICAL PROGRAMS WITH CARDINALITY CONSTRAINTS: A UNIFIED APPROACH FOR WEAK STATIONARITY CONDITIONS AND A SEQUENTIAL OPTIMALITY CONDITION

Projeto de tese apresentado ao Programa de Pós-Graduação em Matemática, Setor de Ciências Exatas, Universidade Federal do Paraná, como requisito parcial para obtenção do título de Doutora em Matemática.

Orientador: Prof. Dr. Ademir Alves Ribeiro.
Coorientadora: Prof.ª Dr.ª Mael Sachine.

CURITIBA
2021

CATALOGAÇÃO NA FONTE – SIBI/UFPR

K94m Krulikovski, Evelin Heringer Manoel

Mathematical programs with cardinality constraints: a unified approach for weak stationarity conditions and a sequential optimality condition [recurso eletrônico]/ Evelin Heringer Manoel Krulikovski - Curitiba, 2021.

Tese (Doutorado) apresentada no Programa de Pós-Graduação em Matemática, Setor de Ciências Exatas, Universidade Federal do Paraná, com requisito parcial para obtenção do Título de Doutora em Matemática.
Orientador: Prof. Dr. Ademir Alves Ribeiro.
Coorientadora: Prof[a] Dr.a Mael Sachine

1. Matemática. 2. Matemática – Solução de Problemas. I. Ribeiro, Ademir Alves. II. Sachine, Mael. III. Título. IV. Universidade Federal do Paraná. CDD 570

Bibliotecária: Vilma Machado CRB9/1563

MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIENCIAS EXATAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO MATEMÁTICA - 40001016041P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em MATEMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **EVELIN HERINGER MANOEL KRULIKOVSKI** intitulada: **Mathematical Programs with Cardinality Constraints: a unified approach for weak stationarity conditions and a sequential optimality condition.**, sob orientação do Prof. Dr. ADEMIR ALVES RIBEIRO, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.
A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 25 de Janeiro de 2021.

Assinatura Eletrônica
25/01/2021 17:55:41.0
ADEMIR ALVES RIBEIRO
Presidente da Banca Examinadora

Assinatura Eletrônica
25/01/2021 17:29:01.0
JOSÉ ALBERTO RAMOS FLOR
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
25/01/2021 17:40:56.0
LUIZ CARLOS MATIOLI
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
25/01/2021 17:21:44.0
SANDRA AUGUSTA SANTOS
Avaliador Externo (UNIVERSIDADE ESTADUAL DE CAMPINAS)

Assinatura Eletrônica
25/01/2021 17:12:55.0
LEONARDO DELARMELINA SECCHIN
Avaliador Externo (UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO)

MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIENCIAS EXATAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO MATEMÁTICA -
40001016041P1

ATA Nº037

ATA DE SESSÃO PÚBLICA DE DEFESA DE DOUTORADO PARA A OBTENÇÃO DO GRAU DE DOUTOR EM MATEMÁTICA

No dia vinte e cinco de janeiro de dois mil e vinte e um às 14:00 horas, na sala https://join.skype.com/qka71MUQakEc, DMAT - UFPR, foram instaladas as atividades pertinentes ao rito de defesa de tese da doutoranda **EVELIN HERINGER MANOEL KRULIKOVSKI**, intitulada**: Mathematical Programs with Cardinality Constraints: a unified approach for weak stationarity conditions and a sequential optimality condition.**, sob orientação do Prof. Dr. ADEMIR ALVES RIBEIRO. A Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em MATEMÁTICA da Universidade Federal do Paraná, foi constituída pelos seguintes Membros: ADEMIR ALVES RIBEIRO (UNIVERSIDADE FEDERAL DO PARANÁ), JOSÉ ALBERTO RAMOS FLOR (UNIVERSIDADE FEDERAL DO PARANÁ), LUIZ CARLOS MATIOLI (UNIVERSIDADE FEDERAL DO PARANÁ), SANDRA AUGUSTA SANTOS (UNIVERSIDADE ESTADUAL DE CAMPINAS), LEONARDO DELARMELINA SECCHIN (UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO). A presidência iniciou os ritos definidos pelo Colegiado do Programa e, após exarados os pareceres dos membros do comitê examinador e da respectiva contra argumentação, ocorreu a leitura do parecer final da banca examinadora, que decidiu pela APROVAÇÃO. Este resultado deverá ser homologado pelo Colegiado do programa, mediante o atendimento de todas as indicações e correções solicitadas pela banca dentro dos prazos regimentais definidos pelo programa. A outorga de título de doutor está condicionada ao atendimento de todos os requisitos e prazos determinados no regimento do Programa de Pós-Graduação. Nada mais havendo a tratar a presidência deu por encerrada a sessão, da qual eu, ADEMIR ALVES RIBEIRO, lavrei a presente ata, que vai assinada por mim e pelos demais membros da Comissão Examinadora.

CURITIBA, 25 de Janeiro de 2021.

Assinatura Eletrônica
25/01/2021 17:55:41.0
ADEMIR ALVES RIBEIRO
Presidente da Banca Examinadora

Assinatura Eletrônica
25/01/2021 17:29:01.0
JOSÉ ALBERTO RAMOS FLOR
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
25/01/2021 17:40:56.0
LUIZ CARLOS MATIOLI
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
25/01/2021 17:21:44.0
SANDRA AUGUSTA SANTOS
Avaliador Externo (UNIVERSIDADE ESTADUAL DE CAMPINAS)

Assinatura Eletrônica
25/01/2021 17:12:55.0
LEONARDO DELARMELINA SECCHIN
Avaliador Externo (UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO)

ACKNOWLEDGEMENT

Firstly, I would like to express my gratefulness to God, for having done so many works in my life. My family, especially my husband Alex, for his patience, his love and his respect. I love you.

A very special gratitude goes out to my advisors Ademir Alves Ribeiro and Mael Sachine for the friendship that we built during the orientation. Whenever I had a problem, their door stood open and they always managed to come up with an idea, a reference or at least a new angle how to look at the problem. I am greatly indebted to them for their support and guidance. I'm sure there are a lot of them in this work.

It was through a summer course at USP that I had the pleasure of meeting the professor Gabriel Haeser, so I thank for all the help from his, mainly by indicating the theme.

Apart from Gabriel Haeser, many other people influenced this thesis. I would like to thank Professor José Alberto Ramos Flor for always listening to me and advising me not only on research but on life in general. In particular, I would like to thank Professor Leonardo Delarmelina Secchin, for all the moments I requested guidance he helped me. In special, I want to thank you for the real attention he gave me during an event at IMPA. Thanks again.

And from many disciplines, many colleagues would like to thank everyone from the Federal University of Paraná. I won't list all your names here, but I think you know who I am talking about.

I would like to thank my thesis committee, Sandra Augusta Santos (UNICAMP), Leonardo Delarmelina Secchin (UFES), José Alberto Ramos Flor (UFPR) and Luiz Carlos Matioli (UFPR) for their comments, which was a valuable contribution to this work.

Last but not the least, this thesis could not have been created without funding. This work was partially supported by Brazilian Founds through CAPES(Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). They funded not only my research, but also participation at conferences and events during my doctorate.

RESUMO

Nesta tese, estudamos uma classe de problemas de otimização, chamada *Mathematical Programs with Cardinality Constraints* (MPCaC)). Este tipo de problema é geralmente difícil de lidar, porque envolve uma restrição que não é contínua e nem convexa, mas fornece soluções esparsas. Assim, reformulamos o problema de uma forma adequada, modelando-o como um problema inteiro misto e então consideramos a sua contraparte contínua, a qual será referida como problema relaxado. Investigamos o problema relaxado analisando as restrições gerais em dois casos: linear e não linear. No caso linear, propomos uma abordagem geral e apresentamos uma discussão das condições de qualificação de Abadie e Guignard, provando neste caso que todo minimizador do problema relaxado satisfaz as condições de Karush-Kuhn-Tucker (KKT). Por outro lado, no caso não linear, mostramos que as condições de qualificação clássicas podem ser violadas. Motivados por encontrar um minimizador para o problema MPCaC, definimos uma nova condição de estacionariedade, mais fraca do que KKT, propondo uma abordagem unificada que vai da estacionariedade mais fraca até a mais forte (dentro de um certo espectro de condições). Entretanto, estas condições não são condições de otimalidade. Deste modo, propomos também um conceito de estacionariedade fraca aproximada chamado *AW-stationarity* (do inglês, *Approximate Weak stationarity*), desenhado para lidar com problemas MPCaC. Provamos que é uma condição de otimalidade legítima independentemente de qualquer condição de qualificação. Muitas pesquisas em condições sequenciais de otimalidade têm sido feitas para otimização não linear com restrições nos últimos anos, sendo alguns trabalhos no contexto de problemas da classe *Mathematical Programs with Complementarity Constraints* (MPCC). No entanto, até onde sabemos, nenhuma condição sequencial de otimalidade foi proposta para problemas MPCaC. Estabelecemos algumas relações entre a nossa condição AW-stationarity e outras condições sequenciais de otimalidade usuais, tais como AKKT, CAKKT e PAKKT. Ressaltamos que, apesar do apelo computacional das condições sequenciais de otimalidade, nosso objetivo até este momento foi discutir os aspectos teóricos de tais condições para problemas MPCaC. Os aspectos algorítmicos por trás de nossa teoria são temas de pesquisa em andamento.

Palavras-chave: Problemas Matemáticos com restrições de cardinalidade; Soluções esparsas; Condições sequenciais de otimalidade; Estacionariedade fraca; Condição de qualificação; Programação não linear.

ABSTRACT

In this thesis, we study a class of optimization problems, called Mathematical Programs with Cardinality Constraints (MPCaC). This kind of problem is generally difficult to deal with, because it involves a constraint that is not continuous neither convex, but provides sparse solutions. Thereby we reformulate MPCaC in a suitable way, by modeling it as mixed-integer problem and then addressing its continuous counterpart, which will be referred to as relaxed problem. We investigate the relaxed problem by analyzing the general constraints in two cases: linear and nonlinear. In the linear case, we propose a general approach and present a discussion of the Guignard and Abadie constraint qualifications, proving in this case that every minimizer of the relaxed problem satisfies the Karush-Kuhn-Tucker (KKT) conditions. On the other hand, in the nonlinear case, we show that some standard constraint qualifications may be violated. Motivated to find a minimizer for the MPCaC problem, we define new stationarity conditions, weaker than KKT, by proposing a unified approach that goes from the weakest to the strongest stationarity (within a certain range of conditions). However, these conditions are not optimality conditions. Thereby, we also propose an Approximate Weak stationarity (AW-stationarity) concept designed to deal with MPCaC problems. We proved that it is a legitimate optimality condition independently of any constraint qualification. Many research on sequential optimality conditions has been addressed for nonlinear constrained optimization in the last few years, some works in the context of Mathematical Programs with Complementarity Constraints (MPCC). However, as far as we know, no sequential optimality condition has been proposed for MPCaC problems. We also establish some relationships between our AW-stationarity and other usual sequential optimality conditions, such as AKKT, CAKKT and PAKKT. We point out that, despite the computational appeal of the sequential optimality conditions, our aim until this moment was to discuss the theoretical aspects of such conditions for MPCaC problems. The algorithmic aspects behind our theory are subject of ongoing research.

Keywords: Mathematical programs with cardinality constraints; Sparse solutions; Sequential optimality conditions; Weak stationarity; Constraint qualification; Nonlinear programming.

LIST OF SYMBOLS

The main symbols used in this thesis are as follows.

$\lvert J\rvert$	The cardinality of J
$\|x\|_0$	Cardinality of the vector $x \in \mathbb{R}^n$
$I_{00}(x, y)$	Index set $\{i \mid x_i = 0, y_i = 0\}$
$I_{\pm 0}(x, y)$	Index set $\{i \mid x_i \neq 0, y_i = 0\}$
$I_{0\pm}(x, y)$	Index set $\{i \mid x_i = 0, y_i \neq 0\}$
$I_{0+}(x, y)$	Index set $\{i \mid x_i = 0, y_i \in (0, 1)\}$
$I_{0>}(x, y)$	Index set $\{i \mid x_i = 0, y_i > 0\}$
$I_{01}(x, y)$	Index set $\{i \mid x_i = 0, y_i = 1\}$
$I_0(x)$	Index set $\{i \mid x_i = 0\}$
$I_g(x)$	For a function $g : \mathbb{R}^n \to \mathbb{R}^s$, denote the set of active indexes $\{i \mid g_i(x) = 0\}$
x_I	The vector in $\mathbb{R}^{\lvert I\rvert}$ consisting of the components $x_i,\ i \in I$
$T_\Omega(\bar{x})$	The tangent cone to Ω at x
$D_\Omega(\bar{x})$	The linearized cone to Ω at x
S^0	The polar cone to S
$g^+(x)$	Maximum Function, i.e., $g^+(x) = \max\{0, g(x)\}$
$\nabla\xi(x)$	Gradient of the function ξ at point x. If $\xi : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable, $\nabla\xi = (\nabla\xi_1, \ldots, \nabla\xi_m)$ denotes the transpose of the Jacobian of ξ
$x * y$	Hadamard product between $x, y \in \mathbb{R}^n$, i.e., $x * y = (x_1y_1, \ldots, x_ny_n) \in \mathbb{R}^n$
$\mathrm{TNLP}_I(\bar{x}, \bar{y})$	The *Tightened* Nonlinear Problem at $(\bar{x}, \bar{y})$
$\mathcal{L}_I(x, y, \lambda)$	Lagrangian associated with $\mathrm{TNLP}_I(\bar{x}, \bar{y})$
$L(x, \lambda)$	Usual Lagrangian function

Other notations will be introduced throughout the text if they are needed.

CONTENTS

Introduction

In this work we study a class of optimization problems called *Mathematical Programs with Cardinality Constraints* (MPCaC) given by

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X, \\ & \|x\|_0 \leq \alpha, \end{array} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function, $X \subset \mathbb{R}^n$ is a set described by equality and/or inequality constraints, $\alpha > 0$ is a given natural number and $\|x\|_0$ denotes the cardinality of the vector $x \in \mathbb{R}^n$, that is, the number of nonzero components of x. Throughout this work we assume that $\alpha < n$, since otherwise the cardinality constraint would not have any effect. On the other hand, if α is too small, the cardinality constraint may be too restrictive leading to an empty feasible set.

The main difference between problem (1) and a standard nonlinear programming problem is that the cardinality constraint, despite of the notation, is not a norm, nor continuous neither convex. A classical way [1] to deal with this difficult cardinality constraint consists of introducing binary variables and then rewriting the problem as a mixed-integer problem

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & f(x) \\ \text{subject to} & x \in X, \\ & e^T y \geq n - \alpha, \\ & x_i y_i = 0, \ i = 1, \ldots, n, \\ & y_i \in \{0, 1\}, \ i = 1, \ldots, n, \end{array} \tag{2}$$

where $e \in \mathbb{R}^n$ denotes the vector of ones. Note that this reformulation is quite natural by noting that if a vector $x \in \mathbb{R}^n$ is such that $\|x\|_0 = r \leq \alpha$, defining $y \in \mathbb{R}^n$ by $y_i = 0$, if $x_i \neq 0$ and $y_i = 1$, if $x_i = 0$, we have $e^T y = n - r \geq n - \alpha$ and $x_i y_i = 0$ for all $i = 1, \ldots, n$.

Alternatively to the formulation (2), which is still complicated to deal with, in view of the binary variables, one may address its continuous counterpart [2]

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & f(x) \\ \text{subject to} & x \in X, \\ & e^T y \geq n - \alpha, \\ & x_i y_i = 0, \ i = 1, \ldots, n, \\ & 0 \leq y_i \leq 1, \ i = 1, \ldots, n, \end{array} \tag{3}$$

which will be referred to as relaxed problem and plays an important role in this work. Besides, with some abuse of terminology, it will be indicated as MPCaC as well. As we shall see in Section 1.3, the problems (2) and (3) are closely related to (1) in terms of feasible points and solutions.

Despite demanding artificial variables that increase the dimension of the problem, it is precisely due to this augmentation that problem (3) has the feature of being manageable, in the sense of being differentiable, which favors one to discuss stationarity concepts. This approach is common to deal with optimization problems [3–7].

In many areas of applications of optimization we seek to find solutions with a small or a bounded number of nonzero components, namely sparse solutions, such as sampling signals or images, machine learning, subset selection in regression, portfolio problems [8–13]. See also [14–16] and the references therein for some more ideas.

One standard way to obtain sparse solutions consists of employing penalization techniques based on the ℓ_1-norm [17]. Another way is imposing explicitly a cardinality constraint to the problem, as the pioneering work [1]. Here, we follow the approach that considers the cardinality constrained problem MPCaC, as [2,18,19]. Specifically, we focus on the theoretical features of the cardinality problem (1), which may be inferred from the properties of the relaxed problem (3), in view of constraint qualifications (CQs) and stationarity concepts. We stress that in this work we are neither concerned with applications nor with computational aspects or algorithmic consequences, which are subject of ongoing research.

In our analysis, we consider two cases: the set X given by linear constraints in the relaxed problem (3), providing a feasible set consisting of linear (separable) constraints in the variables x and y besides the complementarity constraint; and the set X given by nonlinear constraints.

In the first case, we propose a general approach that allows us to simplify the proofs of the results, as compared with the ones presented in [2], as well as establishing Abadie CQ (ACQ), instead of only Guignard CQ (GCQ). We therefore conclude that every minimizer of the relaxed problem (3) satisfies the KKT conditions. This, however, does not mean that weaker stationarity conditions are unnecessary or less important. They are of interest from both the theoretical and the algorithmic viewpoint, as seen in the context of sparsity constrained optimization (MPCaC problems with $X = \mathbb{R}^n$) [14,20,21].

On the other hand, in the nonlinear case, we show that the most known CQs, namely LICQ and MFCQ, are not satisfied, under the assumption that the feasible point (x, y) has at least one nonzero x-component. We also prove that even a weaker condition, ACQ, fails to hold in a wide range of cardinality problems. Moreover, even GCQ, the weakest CQ, may be violated. Therefore, we cannot assert the convergence results for MPCaC in the same way as we usually have in the context of standard nonlinear programming, i.e. for KKT points.

Motivated to find a minimizer for the MPCaC problem (1) for general constraints, we define new and weaker (than KKT) stationarity conditions for this class of problems. This approach is common in the literature. For example, in the works [22–24] the following stationarity conditions are established for Mathematical Programs with Complementarity Constraints (MPCC): Weak, Clarke, Mordukhovich and Strong; whereas for Mathematical Programs with Vanishing Constraints (MPVC) the concept of T-stationarity is proposed [25].

In order to find stationarity conditions for this class of problems, in the present work we define an auxiliary problem, namely Tightened Nonlinear Problem. Its resulting formulation is similar to that made for Mathematical Programs with Complementarity

Constraints [26]. In this way, based on the stationarity concepts established for MPCC in [27,28], we propose new stationarity concepts for the class of MPCaC problems.

Specifically, we propose a unified approach that goes from the weakest to the strongest stationarity (within a range) for the cardinality problem with general constraints. This approach, which will be called W_I-stationarity, is based on a given set of indices I such that the complementarity constraint is always satisfied. Moreover, different levels of stationarity can be obtained depending on the range for the set I. Besides, we prove that this condition is indeed weaker than the classical KKT condition, that is, every KKT point fulfills W_I-stationarity. We also point out that our definition corresponds to concepts of S- and M-stationarity presented in [2] for a proper choice of the index set I.

We stress that although the relaxed problem (3) resembles an MPCC problem, for which there is a vast literature, and also MPVC and Mathematical Programs with Switching Constraints (MPSC) [29], there are important differences between these classes of optimization problems, which in turn increase the importance of specialized research on MPCaC problems. One of such differences is that here we only require positivity for one term in the complementarity constraint $x_i y_i = 0$. Besides, we establish results that are stronger than the corresponding ones known for MPCC's, as for example, the fulfillment of GCQ in the linear case (see Remark 2.1 ahead). It is also worth mentioning that, besides the usual CQs, the MPCC-tailored CQs are also violated for MPCaC problems in the general case (see [2,30] for a more detailed discussion). On the other hand, we shall conclude that W_I-stationarity is a necessary optimality condition under the MPCaC-tailored CQs proposed in [30].

The proposed condition W_I-stationarity, despite being weaker than KKT, is not a necessary optimality condition. Therefore, we propose in this work an Approximate Weak stationarity (AW-stationarity) concept, which will be proved to be a legitimate optimality condition, independently of any constraint qualification.

Approximate stationarity conditions, also referred to as sequential optimality conditions, has been subject of intense research [6,31–36] and provide strong convergence results when associated with practical methods, such as the augmented Lagrangian method (see [37] and references therein). This is due to the fact that such conditions are necessary for optimality independently of the fulfillment of any constraint qualification (CQ). One of the most popular of these conditions for standard nonlinear programming is the *approximate Karush-Kuhn-Tucker* (AKKT) [32]. Another two of such conditions, both stronger than AKKT, are *positive approximate KKT* (PAKKT) [31] and *complementary approximate KKT* (CAKKT) [35]. Whenever it is proved that an AKKT (or CAKKT or PAKKT) point is indeed a Karush-Kuhn-Tucker (KKT) point under a certain CQ, any algorithm that reaches AKKT (or CAKKT or PAKKT) points (e.g. augmented Lagrangian-type methods) automatically has the theoretical convergence established assuming the same CQ.

Sequential optimality conditions have also been proposed for nonstandard optimization [27,28,38,39]. In the context of *Mathematical Programs with Equilibrium Constraints* (MPECs) and motivated by AKKT, it was introduced in [28] the MPEC-AKKT condition with a geometric appeal and in [27], new conditions were established for *Mathematical Problems with Complementarity Constraints* (MPCCs), namely AW-, AC- and AM-stationarity. The latter one was compared with the sequential condition presents in [28].

Even though there is a considerable literature devoted to sequential conditions for standard nonlinear optimization and even for specific problems (MPCC), to the best of our knowledge, no sequential optimality condition has been proposed for MPCaC problems. Such problems are very degenerate because of the challenging complementarity constraints $x_i y_i = 0$ and therefore the known sequential optimality conditions may not be suitable to deal with them. Thereby, our sequential optimality condition, AW-stationarity, which is associated with W_I-stationarity and designed to deal with MPCaC problems comes to fill this gap. This condition is indeed a necessary optimality condition, without any constraint qualification assumption, and it is based on the one proposed in [27] for MPCC problems. We also establish some relationships between our AW-stationarity and other well known sequential optimality conditions. In particular, and surprisingly, we prove that AKKT fails to detect good candidates for optimality for every MPCaC problem.

Organization of the thesis. This thesis consists of two papers developed during the PhD *On the weak stationarity conditions for Mathematical Programs with Cardinality Constraints: a unified approach* [50] and *A sequential optimality condition for Mathematical Programs with Cardinality Constraints* [51]. In order to facilitate the reading, we start with a short presentation on what the two papers are about. The work is organized as follows: in Chapter 1 we establish some basic definitions, results and examples regarding standard nonlinear programming as well as one of the contributions of this work for cardinality constrained problems. Particularly, in Section 1.1, we establish more general results than the ones presented in [2] (for the linear case) as well as a discussion about ACQ, instead of only GCQ. Section 1.2 brings the known sequential optimality conditions for standard NLP. In Section 1.3 we present relations between the MPCaC and the reformulated problems. Section 1.4 treats some relations between the MPCaC and other classes of optimization problems. We define in Chapter 2 the weak stationarity conditions for Mathematical Programs with Cardinality Constraints. In Section 2.1 we present one of the contributions, by considering the relaxed problem (3) with X given by linear constraints. Section 2.2 is devoted to our main contribution, presenting the analysis of the nonlinear case, including a discussion of the main CQs, with results, examples and counterexamples. In Chapter 3 we present the theoretical results obtained until now concerning sequential optimality conditions for MPCaC problems. Section 3.1 presents our definition of AW-stationarity and, among several results, the proof that it is a legitimate optimality condition without any constraint qualification assumption. In Section 3.2 we provide some relationships between approximate stationarity for standard nonlinear optimization and AW-stationarity. In Section 3.3 we discuss some ideas for future work on MPCaC, in particular, regarding sequential optimality conditions. Finally, concluding remarks are presented in Chapter 3.3 and Appendix A brings some additional examples discussed along the seminars and alternative proofs for some results of this work.

Capítulo 1

Preliminaries

In this chapter we recall some basic definitions, results and examples regarding standard nonlinear programming (NLP) as well as some of the contributions of this work, establishing general results from which we can derive properties for the specific problem (3). We present some relations between the MPCaC and the reformulated problems as well as a short comparison between the MPCaC and other classes of optimization problems.

1.1 NLP and constraint qualifications

Consider the problem

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leq 0, \\ & h(x) = 0, \end{array} \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions. The feasible set of the problem (1.1) is denoted by

$$\Omega = \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}. \tag{1.2}$$

Definition 1.1.1 *We say that $x^* \in \Omega$ is a global solution of the problem (1.1), that is, a global minimizer of f in Ω, if $f(x^*) \leq f(x)$ for all $x \in \Omega$. If $f(x^*) \leq f(x)$ for all $x \in \Omega$ such that $\|x - x^*\| \leq \delta$, for some $\delta > 0$, x^* is said to be a local solution of the problem.*

A feasible point $x^* \in \Omega$ is said to be *stationary* for the problem (1.1) if there exists a vector $\lambda = (\lambda^g, \lambda^h) \in \mathbb{R}^m_+ \times \mathbb{R}^p$ (Lagrange multipliers) such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(x^*) = 0, \tag{1.3a}$$

$$(\lambda^g)^T g(x^*) = 0. \tag{1.3b}$$

The function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ given by

$$L(x, \lambda^g, \lambda^h) = f(x) + (\lambda^g)^T g(x) + (\lambda^h)^T h(x) \tag{1.4}$$

is the *Lagrangian* function associated with the problem (1.1).

The conditions (1.3a)–(1.3b) are known as Karush-Kuhn-Tucker (KKT) conditions and, under certain qualification assumptions, are satisfied at a local minimizer. There are a lot of constraint qualifications, that is, conditions under which every minimizer satisfies KKT [40–46]. In order to discuss some of them, let us recall the definition of cone, which plays an important role in this context.

We say that a nonempty set $C \subset \mathbb{R}^n$ is a *cone* if $td \in C$ for all $t \geq 0$ and $d \in C$. Given a set $S \subset \mathbb{R}^n$, its *polar* is the cone $S^\circ = \{p \in \mathbb{R}^n \mid p^T x \leq 0, \ \forall x \in S\}$. See Figure 1.1.

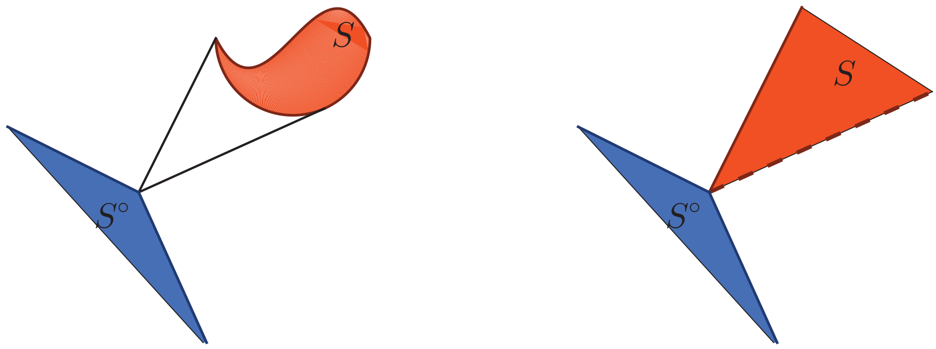

Figura 1.1: Illustration of some sets and their polar cones.

Associated with the feasible set of the problem (1.1), the *tangent cone* at $\bar{x} \in \Omega$ is given by

$$T_\Omega(\bar{x}) = \left\{ d \in \mathbb{R}^n \mid \exists (x^k) \subset \Omega, \, (t_k) \subset \mathbb{R}_+ \, : \, t_k \to 0 \text{ and } \frac{x^k - \bar{x}}{t_k} \to d \right\}$$

and the *linearized cone* at $\bar{x} \in \Omega$ is

$$D_\Omega(\bar{x}) = \left\{ d \in \mathbb{R}^n \mid \nabla g_i(\bar{x})^T d \leq 0, \ i \in I_g(\bar{x}) \text{ and } \nabla h(\bar{x})^T d = 0 \right\}.$$

The following basic result says that we may ignore inactive constraints whenever dealing with the tangent and linearized cones.

Lemma 1.1.2 *Consider a feasible point $\bar{x} \in \Omega$, defined in (1.2), an index set $J \supset I_g(\bar{x})$ and*

$$\Omega' = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, \, i \in J, \ h(x) = 0\}.$$

Then, $T_\Omega(\bar{x}) = T_{\Omega'}(\bar{x})$ and $D_\Omega(\bar{x}) = D_{\Omega'}(\bar{x})$.

Proof. Note first that $\bar{x} \in \Omega'$ since $\Omega \subset \Omega'$. Moreover, since $g_i(\bar{x}) < 0$ for $i \notin J$, there exists $\delta > 0$ such that $B(\bar{x}, \delta) \cap \Omega' = B(\bar{x}, \delta) \cap \Omega$. Thus, $T_{\Omega'}(\bar{x}) = T_\Omega(\bar{x})$ because the conditions $t_k \to 0$ and $(x^k - \bar{x})/t_k \to d$ imply that $x^k \to \bar{x}$. The equality between the linearized cones is straightforward, as the active indices corresponding to Ω and Ω' coincide. $\square$

Now we relate the cones of feasible sets if some variables do not appear in the constraints.

Lemma 1.1.3 *Consider the general feasible set Ω, defined in (1.2), and the set*

$$\Omega' = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid g(x) \leq 0, h(x) = 0\}.$$

Given a feasible point $(\bar{x}, \bar{y}) \in \Omega'$, we have

$$T_{\Omega'}(\bar{x}, \bar{y}) = T_\Omega(\bar{x}) \times \mathbb{R}^m \quad \text{and} \quad D_{\Omega'}(\bar{x}, \bar{y}) = D_\Omega(\bar{x}) \times \mathbb{R}^m.$$

As a consequence,

$$T^\circ_{\Omega'}(\bar{x}, \bar{y}) = T^\circ_\Omega(\bar{x}) \times \{0\} \quad \text{and} \quad D^\circ_{\Omega'}(\bar{x}, \bar{y}) = D^\circ_\Omega(\bar{x}) \times \{0\}.$$

Proof. Consider a tangent direction $d = (\alpha, \beta) \in T_{\Omega'}(\bar{x}, \bar{y})$. Then there exist sequences $(x^k, y^k) \subset \Omega'$ and $t_k \to 0$ such that

$$\frac{(x^k, y^k) - (\bar{x}, \bar{y})}{t_k} \to d.$$

Thus, $\dfrac{x^k - \bar{x}}{t_k} \to \alpha \in T_{\Omega}(\bar{x})$, since $(x^k) \subset \Omega$, proving that $T_{\Omega'}(\bar{x}, \bar{y}) \subset T_{\Omega}(\bar{x}) \times \mathbb{R}^m$. Now, consider a vector $d = (\alpha, \beta)$[1] $\in T_{\Omega}(\bar{x}) \times \mathbb{R}^m$. Then there exist sequences $(x^k) \subset \Omega$ and $t_k \to 0$ such that

$$\frac{x^k - \bar{x}}{t_k} \to \alpha.$$

Defining $y^k = \bar{y} + t_k \beta$, we have $(x^k, y^k) \subset \Omega'$ and

$$\frac{(x^k, y^k) - (\bar{x}, \bar{y})}{t_k} \to d,$$

giving $d \in T_{\Omega'}(\bar{x}, \bar{y})$. The relation between the linearized cones follows from the fact that if $\zeta(x, y) = g(x)$ and $\xi(x, y) = h(x)$ represent the constraints that define Ω' and $d = (\alpha, \beta)$, then

$$\nabla \zeta_i(x, y)^T d = \nabla g_i(x)^T \alpha \quad \text{and} \quad \nabla \xi_j(x, y)^T d = \nabla h_j(x)^T \alpha.$$

Finally, in order to prove the last statement of the lemma, we claim that $(S \times \mathbb{R}^m)^\circ = S^\circ \times \{0\}$. Indeed, if $(u, v) \in (S \times \mathbb{R}^m)^\circ$, then $u^T \alpha + v^T \beta \leq 0$ for all $\alpha \in S$, $\beta \in \mathbb{R}^m$. In particular, for $\beta = tv$, $t \in \mathbb{R}$, we have $u^T \alpha + t\|v\|^2 \leq 0$ for all $t \in \mathbb{R}$. This means that $v = 0$ and hence, $u \in S^\circ$. The reverse inclusion is immediate. □

Two well known constraint qualifications are defined below. Gould and Tolle [41] proved that one of them, namely Guignard CQ, is the weakest CQ that guarantees that local minimality implies KKT. The other one is Abadie CQ, which is stronger than Guignard CQ.

Definition 1.1.4 *We say that Abadie constraint qualification (ACQ) holds at $\bar{x} \in \Omega$ if $T_{\Omega}(\bar{x}) = D_{\Omega}(\bar{x})$. If $T_{\Omega}^\circ(\bar{x}) = D_{\Omega}^\circ(\bar{x})$, we say that Guignard constraint qualification (GCQ) holds at $\bar{x}$.*

We stress that if Ω is defined only by linear constraints, then $T_{\Omega}(\bar{x}) = D_{\Omega}(\bar{x})$ for any $\bar{x} \in \Omega$, that is, every point satisfies ACQ.

From now on, for a better analysis we consider $(\bar{x}, \bar{y})$ a feasible point of the problem (3) and the following sets in Table 1.1. If there is no chance for ambiguity, we sometimes suppress the argument and, for example, write I_{00} for $I_{00}(\bar{x}, \bar{y})$, $I_{0\pm}$ for $I_{0\pm}(\bar{x}, \bar{y})$ and so on.

---

[1] Strictly speaking, we should write α and β in a single column, but we wrote in this manner for notational convenience.

	$\bar{x}_i = 0$	$\bar{x}_i \neq 0$
$\bar{y}_i = 0$	$I_{00}(\bar{x}, \bar{y})$	$I_{\pm 0}(\bar{x}, \bar{y})$
$\bar{y}_i \neq 0$	$I_{0\pm}(\bar{x}, \bar{y})$	does not exist
$\bar{y}_i \in (0, 1)$	$I_{0+}(\bar{x}, \bar{y})$	does not exist
$\bar{y}_i = 1$	$I_{01}(\bar{x}, \bar{y})$	does not exist

Tabela 1.1: Index sets at $(\bar{x}, \bar{y})$ for the problem (3). The index set $I_0(\bar{x}) = \{i \mid \bar{x}_i = 0\}$ is equal to $I_{00} \cup I_{0+} \cup I_{01}$. Moreover, $I_0 \cup I_{\pm 0} = \{1, \ldots, n\}$.

In the following lemma we analyze GCQ for simple complementarity constraints.

Lemma 1.1.5 *Consider the set*

$$\Omega = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid y \geq 0, x * y = 0\}. \tag{1.5}$$

Given $(\bar{x}, \bar{y}) \in \Omega$, *it holds* $T_\Omega^\circ(\bar{x}, \bar{y}) = D_\Omega^\circ(\bar{x}, \bar{y})$.

Proof. Denote the constraints that define Ω by $\zeta(x, y) = -y$ and $\xi(x, y) = x * y$. Given $(u, v) \in D_\Omega(\bar{x}, \bar{y})$, we claim that the vectors $(u, 0)$ and $(0, v)$ belong to $T_\Omega(\bar{x}, \bar{y})$. Indeed,

$$\bar{y}_i u_i + \bar{x}_i v_i = \nabla \xi_i(\bar{x}, \bar{y})^T \begin{pmatrix} u \\ v \end{pmatrix} = 0.$$

for all $i = 1, \ldots, n$. Thus,

$$u_{I_{0+}} = 0 \quad \text{and} \quad v_{I_{\pm 0}} = 0. \tag{1.6}$$

To prove that $(u, 0) \in T_\Omega(\bar{x}, \bar{y})$, define $t_k = 1/k$ and $(x^k, y^k) = (\bar{x} + t_k u, \bar{y})$. Thus,

$$\frac{(x^k, y^k) - (\bar{x}, \bar{y})}{t_k} \to (u, 0).$$

Moreover, $y^k = \bar{y} \geq 0$ and, using (1.6), we see that $x^k_{I_{0+}} = 0$ and $y^k_{I_{\pm 0} \cup I_{00}} = 0$, giving $x^k * y^k = 0$. So, $(x^k, y^k) \subset \Omega$ and then $(u, 0) \in T_\Omega(\bar{x}, \bar{y})$. Let us see that $(0, v) \in T_\Omega(\bar{x}, \bar{y})$. For this we define the sequence (z^k, w^k) by

$$z^k = \bar{x} \quad \text{and} \quad w^k = \bar{y} + t_k v.$$

Analogously to the previous case, we see that

$$\frac{(z^k, w^k) - (\bar{x}, \bar{y})}{t_k} \to (0, v) \quad \text{and} \quad z^k * w^k = 0,$$

where the equality follows from the fact that $z^k_{I_{0+} \cup I_{00}} = 0$ and $w^k_{I_{\pm 0}} = 0$. Furthermore, if $i \in I_{0+}$, we have $\bar{y}_i > 0$, which implies that $w^k_i > 0$ for all sufficiently large k. On the other hand, for $i \in I_{00}$, the constraint ζ_i is active and hence,

$$-v_i = \nabla \zeta_i(\bar{x}, \bar{y})^T \begin{pmatrix} u \\ v \end{pmatrix} \leq 0,$$

which implies that $w^k_i = \bar{y}_i + t_k v_i = t_k v_i \geq 0$. Thus, $(z^k, w^k) \subset \Omega$, giving $(0, v) \in T_\Omega(\bar{x}, \bar{y})$.

Now let us prove GCQ. Since the tangent cone is always a subset of the linearized cone, we only need to prove the inclusion $T_\Omega^\circ(\bar{x}, \bar{y}) \subset D_\Omega^\circ(\bar{x}, \bar{y})$. Consider then $p \in T_\Omega^\circ(\bar{x}, \bar{y})$. Thus, $p^T d \leq 0$ for all $d \in T_\Omega(\bar{x}, \bar{y})$. Given $d = (u, v) \in D_\Omega(\bar{x}, \bar{y})$, we can use the claim just established to write

$$d^1 = (u, 0) \in T_\Omega(\bar{x}, \bar{y}), \quad d^2 = (0, v) \in T_\Omega(\bar{x}, \bar{y}) \quad \text{and} \quad d = d^1 + d^2.$$

Therefore, $p^T d = p^T d^1 + p^T d^2 \leq 0$, proving that $p \in D_\Omega^\circ(\bar{x}, \bar{y})$. $\square$

Remark 1.1 *In fact we can also analyze ACQ for the set given in Lemma 1.1.5. In particular, we give explicit representations for the cones so that the constraint qualification can be precisely described. Given $(\bar{x},\bar{y}) \in \Omega$, we have*

$$T_\Omega(\bar{x},\bar{y}) = \big\{(u,v) \in \mathbb{R}^n \times \mathbb{R}^n \mid u_{I_{0>}} = 0;\ v_{I_{\pm 0}} = 0;\ v_{I_{00}} \geq 0;\ u * v = 0\big\},$$
$$D_\Omega(\bar{x},\bar{y}) = \big\{(u,v) \in \mathbb{R}^n \times \mathbb{R}^n \mid u_{I_{0>}} = 0;\ v_{I_{\pm 0}} = 0;\ v_{I_{00}} \geq 0\big\},$$

where we have denoted $I_{0>} = I_{0>}(\bar{x},\bar{y})$, $I_{0\pm} = I_{0\pm}(\bar{x},\bar{y})$ and so on. Thus, ACQ holds if and only if $I_{00} = \emptyset$.

We present in Proposition A.2 the proof of the claim in the above remark. Besides, part of the statements are revisited and generalized in Theorems 1.1.7 and 1.1.8.

An interesting property obtained under the *strict complementarity* condition $I_{00}(\bar{x},\bar{y}) = \emptyset$ says that the constraint $x * y = 0$ near $(\bar{x},\bar{y})$ can be rewritten only by linear constraints, as proved in the following result and illustrated in the Figure 1.2.

Lemma 1.1.6 *Consider the set*

$$\Omega = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n \mid \varphi(x,y) \leq 0,\ \rho(x,y) = 0,\ x * y = 0\},$$

where $\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ and $\rho : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions. Suppose that a feasible point $(\bar{x},\bar{y}) \in \Omega$ satisfies $I_{00}(\bar{x},\bar{y}) = \emptyset$. Define the set

$$\Omega' = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n \mid \varphi(x,y) \leq 0,\ \rho(x,y) = 0,\ x_{I_{0\pm}} = 0,\ y_{I_{\pm 0}} = 0\},$$

again using the simplified notation $I_{0\pm} = I_{0\pm}(\bar{x},\bar{y})$ and $I_{\pm 0} = I_{\pm 0}(\bar{x},\bar{y})$. Then, we have $T_\Omega(\bar{x},\bar{y}) = T_{\Omega'}(\bar{x},\bar{y})$ and $D_\Omega(\bar{x},\bar{y}) = D_{\Omega'}(\bar{x},\bar{y})$.

Proof. Note first that $(\bar{x},\bar{y}) \in \Omega'$. Besides, since $I_{0\pm} \cup I_{\pm 0} = \{1,\ldots,n\}$, we have $\Omega' \subset \Omega$. We claim that there exists $\delta > 0$ such that

$$B\big((\bar{x},\bar{y}),\delta\big) \cap \Omega = B\big((\bar{x},\bar{y}),\delta\big) \cap \Omega'. \tag{1.7}$$

Indeed, take $\delta = \min\{|\bar{x}_i|,\ i \in I_{\pm 0}, |\bar{y}_i|,\ i \in I_{0\pm}\}$ and consider the $\|\cdot\|_\infty$ norm to define the ball. So, given $(x,y) \in B\big((\bar{x},\bar{y}),\delta\big) \cap \Omega$, if $i \in I_{0\pm}$, then $y_i \neq 0$, implying that $x_i = 0$. Analogously, we see that $y_i = 0$ for $i \in I_{\pm 0}$. Thus we have (1.7) and hence, $T_\Omega(\bar{x},\bar{y}) = T_{\Omega'}(\bar{x},\bar{y})$.

Now, consider the sets

$$\mathcal{G}_{\varphi\rho} = \left\{ \sum_{i \in I_\varphi} \lambda_i^\varphi \nabla \varphi_i(\bar{x},\bar{y}) + \sum_{i=1}^p \lambda_i^\rho \nabla \rho_i(\bar{x},\bar{y}) \mid \lambda_i^\varphi \in \mathbb{R}_+,\ i \in I_\varphi,\ \lambda^\rho \in \mathbb{R}^p \right\},$$

where $I_\varphi = I_\varphi(\bar{x},\bar{y})$,

$$\mathcal{G}_0 = \left\{ \sum_{i \in I_{0\pm}} \lambda_i \begin{pmatrix} \bar{y}_i e_i \\ 0 \end{pmatrix} + \sum_{i \in I_{\pm 0}} \lambda_i \begin{pmatrix} 0 \\ \bar{x}_i e_i \end{pmatrix} \mid \lambda \in \mathbb{R}^n \right\}$$

and

$$\mathcal{G}_0' = \left\{ \sum_{i \in I_{0\pm}} \lambda_i \begin{pmatrix} e_i \\ 0 \end{pmatrix} + \sum_{i \in I_{\pm 0}} \lambda_i \begin{pmatrix} 0 \\ e_i \end{pmatrix} \mid \lambda \in \mathbb{R}^n \right\}.$$

Since $\mathcal{G}_0 = \mathcal{G}_0'$, we have

$$D_\Omega(\bar{x},\bar{y}) = (\mathcal{G}_{\varphi\rho} + \mathcal{G}_0)^\circ = (\mathcal{G}_{\varphi\rho} + \mathcal{G}_0')^\circ = D_{\Omega'}(\bar{x},\bar{y}),$$

completing the proof. □

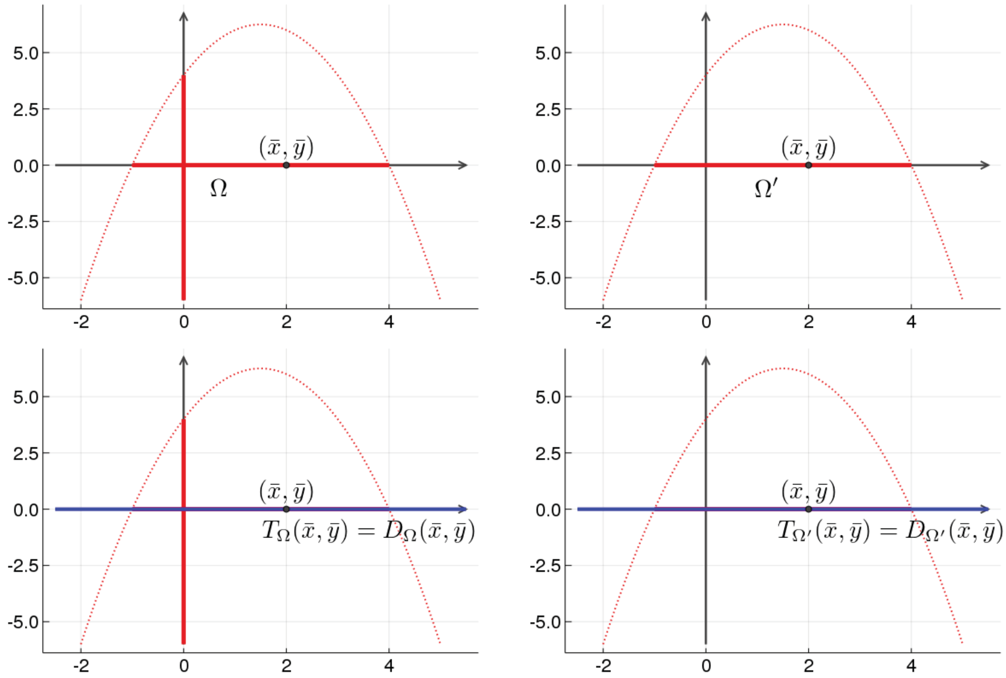

Figura 1.2: Illustration of Lemma 1.1.6. The thick red line is the feasible set $\Omega = \{(x, y) \in \mathbb{R}^2 \mid y \leq (x+1)(4-x), y = 0, \; xy = 0\}$ and the feasible point $(2, 0)$ satisfies $I_{00}(2, 0) = \emptyset$. The constraint $xy = 0$ near $(2, 0)$ can be rewritten by $y = 0$, see the set Ω'. Note that the tangent and linearized cones coincide.

Now, we present one of the contributions of this work, establishing more general results from which we will derive, in Section 2.1, properties for the specific problem (3). Moreover, this general approach enables us to simplify the proofs, as compared with the ones presented in [2], as well as to discuss also ACQ, instead of only GCQ. The major difference between our approach and the strategy used in [2] for proving the Guignard CQ is that they use partitions of the index set I_{00} to construct decompositions of the feasible set and the corresponding cones in terms of simpler sets, whereas we decompose an arbitrary vector of the linearized cone as a sum of two vectors belonging to the tangent cone, making the proof very simple. Furthermore, we also provide here the analysis of the Abadie condition under the strict complementarity condition.

Consider then the set

$$\Omega = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid Ax = b, \; \tilde{A}x \leq \tilde{b}, \; My = r, \; \tilde{M}y \leq \tilde{r}, \; x * y = 0\}, \tag{1.8}$$

where the matrices A, $\tilde{A}$, M and $\tilde{M}$ and the vectors b, $\tilde{b}$, r and $\tilde{r}$ have appropriate dimensions.

We start by giving sufficient conditions for ACQ to be satisfied.

Theorem 1.1.7 *Consider the set Ω, defined in (1.8), and a feasible point $(\bar{x}, \bar{y}) \in \Omega$. If $I_{00}(\bar{x}, \bar{y}) = \emptyset$, then ACQ holds at $(\bar{x}, \bar{y})$.*

Proof. Consider the set Ω' defined in Lemma 1.1.6 associated with the set Ω given in (1.8). Using such a lemma and the fact that Ω' is given by linear constraints, we have

$$T_\Omega(\bar{x}, \bar{y}) = T_{\Omega'}(\bar{x}, \bar{y}) = D_{\Omega'}(\bar{x}, \bar{y}) = D_\Omega(\bar{x}, \bar{y})$$

and therefore ACQ holds at $(\bar{x}, \bar{y})$. ◻

Remark 1.2 *Note that the validity of Theorem 1.1.7 does not depend on the separability of the linear constraints, that is, it is valid for the more general set*

$$\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid Bx + Cy = c,\ \bar{B}x + \bar{C}y \leq \bar{c},\ x * y = 0\}, \tag{1.9}$$

where the matrices B, C, $\bar{B}$ and $\bar{C}$ and the vectors c and $\bar{c}$ have appropriate dimensions. Moreover, in the case $I_{00}(\bar{x}, \bar{y}) \neq \emptyset$, ACQ may or may not be valid. For example, ACQ is satisfied at every (x, y) in the set

$$\Omega = \{(x, y) \in \mathbb{R}^2 \mid x = 0,\ 0 \leq y \leq 1,\ xy = 0\}, \tag{1.10}$$

in particular at the point $(\bar{x}, \bar{y}) = (0, 0)$. On the other hand, if

$$\Omega' = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1,\ 0 \leq y \leq 1,\ xy = 0\}, \tag{1.11}$$

ACQ does not hold at the point $(\bar{x}, \bar{y})$, as illustrated in the Figure 1.3.

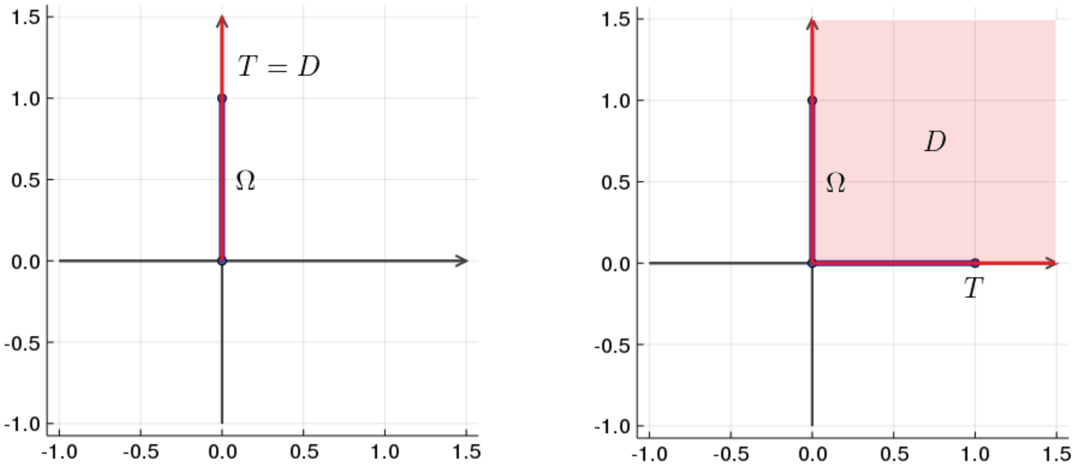

Figura 1.3: Note that in both pictures we have $I_{00}(\bar{x}, \bar{y}) \neq \emptyset$. The first picture illustrates the set Ω given by (1.10) and the fact that ACQ holds at $(\bar{x}, \bar{y})$. The second picture illustrates the set Ω' given by (1.11) where ACQ does not hold.

The next result is more precise and does not depend on strict complementarity if we replace ACQ by GCQ.

Theorem 1.1.8 *Consider the set Ω defined in (1.8). Then, every feasible point $(\bar{x}, \bar{y}) \in \Omega$ satisfies GCQ.*

Proof. We invoke Lemma 1.1.2 to assume without loss of generality that there is no inactive constraint at $(\bar{x}, \bar{y})$. Denote $\rho(x, y) = Ax - b$, $\tilde{\rho}(x, y) = \tilde{A}x - \tilde{b}$, $\zeta(x, y) = My - r$, $\tilde{\zeta}(x, y) = \tilde{M}x - \tilde{r}$ and $\xi(x, y) = x * y$. Whenever there is no chance for ambiguity, we suppress the argument and, for example, write I_{00} for $I_{00}(\bar{x}, \bar{y})$, $I_{0\pm}$ for $I_{0\pm}(\bar{x}, \bar{y})$ and so on. Consider an arbitrary $d = (u, v) \in D_\Omega(\bar{x}, \bar{y})$. We claim that the vectors $(u, 0)$ and $(0, v)$ belong to $T_\Omega(\bar{x}, \bar{y})$. Indeed,

$$\bar{y}_i u_i + \bar{x}_i v_i = \nabla \xi_i(\bar{x}, \bar{y})^T d = 0$$

for all $i = 1, \ldots, n$. Thus,

$$u_{I_{0\pm}} = 0 \quad \text{and} \quad v_{I_{\pm 0}} = 0. \tag{1.12}$$

To prove that $(u,0) \in T_\Omega(\bar{x},\bar{y})$, define $t_k = 1/k$ and the sequence (x^k, y^k) by

$$x^k = \bar{x} + t_k u \quad \text{and} \quad y^k = \bar{y}.$$

Thus, $\dfrac{(x^k,y^k) - (\bar{x},\bar{y})}{t_k} \to (u,0)$. Let us prove that $(x^k,y^k) \subset \Omega$. Using (1.12), we see that $x^k_{I_{0\pm}} = 0$ and $y^k_{I_{\pm 0} \cup I_{00}} = 0$ and then $x^k * y^k = 0$. Moreover, since $(u,v) \in D_\Omega(\bar{x},\bar{y})$, we obtain

$$Au = 0, \quad \tilde{A}u \leq 0, \quad Mv = 0 \quad \text{and} \quad \tilde{M}v \leq 0$$

implying that

$$Ax^k = b, \quad \tilde{A}x^k \leq \tilde{b}, \quad My^k = r \quad \text{and} \quad \tilde{M}y^k \leq \tilde{r}.$$

So, $(x^k,y^k) \subset \Omega$ and then $(u,0) \in T_\Omega(\bar{x},\bar{y})$. The fact that $(0,v) \in T_\Omega(\bar{x},\bar{y})$ can be proved analogously.

Finally, to establish the relation $T^\circ_\Omega(\bar{x},\bar{y}) = D^\circ_\Omega(\bar{x},\bar{y})$, consider arbitrary vectors $p \in T^\circ_\Omega(\bar{x},\bar{y})$ and $d \in D_\Omega(\bar{x},\bar{y})$. As seen above, we can write $d = d^1 + d^2$, with $d^1, d^2 \in T_\Omega(\bar{x},\bar{y})$. Thus, $p^T d = p^T d^1 + p^T d^2 \leq 0$. □

It should be noted that, contrary to what occurs in Theorem 1.1.7, the above result cannot be generalized for the set defined in (1.9), as can be seen in the following example.

Example 1.1.9 *Consider the set*

$$\Omega = \{(x,y) \in \mathbb{R}^2 \mid x \geq 0,\ y \geq 0,\ -x + y \leq 0,\ xy = 0\}.$$

It can be seen that

$$T_\Omega(0,0) = \left\{d \in \mathbb{R}^2 \mid d_1 \geq 0,\ d_2 = 0\right\}$$

and

$$D_\Omega(0,0) = \left\{d \in \mathbb{R}^2 \mid 0 \leq d_2 \leq d_1\right\}.$$

Hence, $T^\circ_\Omega(0,0) \neq D^\circ_\Omega(0,0)$ *(see Figure 1.4).*

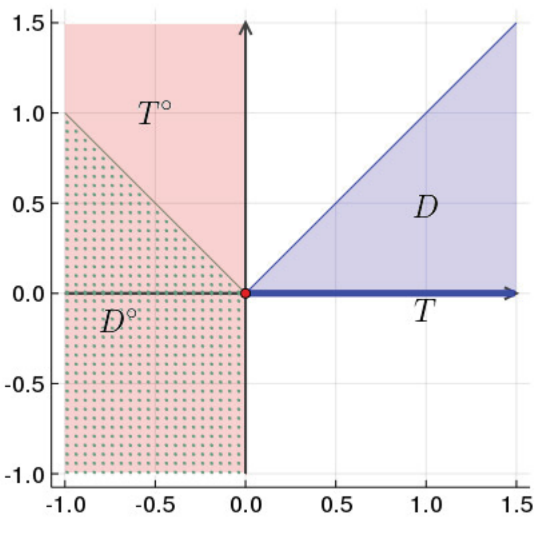

Figura 1.4: Illustration of the cones associated with the set Ω of the Example 1.1.9, showing that GCQ does not hold.

Obviously, we could bring Lemma 1.1.5 and Remark 1.1 back and analyze them in the light of Theorems 1.1.7 and 1.1.8.

Now we recall the most well known constraint qualifications regarding the problem (1.1).

Definition 1.1.10 *We say that linear independence constraint qualification (LICQ) is satisfied at $\bar{x} \in \Omega$ if the set formed by the gradients of the active inequality constraints and the gradients of the equality constraints*

$$\{\nabla g_i(\bar{x}) \mid i \in I_g(\bar{x})\} \cup \{\nabla h_i(\bar{x}), i = 1, \ldots, p\}$$

is linearly independent.

Definition 1.1.11 *We say that Mangasarian-Fromovitz constraint qualification (MFCQ) holds at $\bar{x} \in \Omega$ if the gradient vectors $\nabla h_i(\bar{x}), i = 1, \ldots, p$, are linearly independent and there exists a vector $d \in \mathbb{R}^n$ such that*

$$\nabla g_i(\bar{x})^T d < 0 \text{ and } \nabla h_j(\bar{x})^T d = 0$$

for all $i \in I_g(\bar{x})$ and $j = 1, \ldots, p$.

It is well known that LICQ implies MFCQ, which in turn implies ACQ (see Figure 1.5).

$$\boxed{\text{LICQ}} \longrightarrow \boxed{\text{MFCQ}} \longrightarrow \boxed{\text{ACQ}} \longrightarrow \boxed{\text{GCQ}}$$

Figura 1.5: Some of the constraint qualifications considered in this work. An arrow indicates a strict implication between two conditions.

Moreover, any constraint qualification ensures that a local minimizer is a KKT point, as stated in the following classical result.

Theorem 1.1.12 (Karush-Kuhn-Tucker (KKT)) *Let $x^* \in \mathbb{R}^n$ be a local minimizer of problem (1.1) and suppose that a constraint qualification is satisfied. Then there exists $\lambda = (\lambda^g, \lambda^h) \in \mathbb{R}^m_+ \times \mathbb{R}^p$ such that $(x^*, \lambda^g, \lambda^h)$ satisfies (1.3a) and (1.3b).*

1.2 Sequential optimality conditions for standard NLP

The goal of this section is to present some well known approximate optimality conditions for nonlinear constrained optimization [32–37].

Definition 1.2.1 *[32] Let $\bar{x} \in \mathbb{R}^n$ be a feasible point for the problem (1.1). We say that $\bar{x}$ is an* Approximate KKT *(AKKT) point if there exist sequences $(x^k) \subset \mathbb{R}^n$ and $(\lambda^k) = \left(\lambda^{g,k}, \lambda^{h,k}\right) \subset \mathbb{R}^m_+ \times \mathbb{R}^p$ such that $x^k \to \bar{x}$,*

$$\nabla_x L(x^k, \lambda^{g,k}, \lambda^{h,k}) \to 0, \tag{1.13a}$$

$$\min\{-g(x^k), \lambda^{g,k}\} \to 0. \tag{1.13b}$$

We present in the appendix A two conditions equivalent to the above definition. On the other hand, we have below two conditions stronger than AKKT.

Definition 1.2.2 *[35] Let $\bar{x} \in \mathbb{R}^n$ be a feasible point for the problem (1.1). We say that $\bar{x}$ is a* Complementary Approximate KKT *(CAKKT) point if there exist sequences $(x^k) \subset \mathbb{R}^n$ and $(\lambda^k) = \left(\lambda^{g,k}, \lambda^{h,k}\right) \subset \mathbb{R}^m_+ \times \mathbb{R}^p$ such that $x^k \to \bar{x}$,*

$$\nabla_x L(x^k, \lambda^{g,k}, \lambda^{h,k}) \to 0, \tag{1.14a}$$

$$\lambda^{g,k} * g(x^k) \to 0 \quad \text{and} \quad \lambda^{h,k} * h(x^k) \to 0. \tag{1.14b}$$

Remark 1.3 *Note that if $(\alpha^k) \subset \mathbb{R}_+$ and $(\beta^k) \subset \mathbb{R}$ are sequences satisfying $\alpha^k \beta^k \to 0$ and $\beta^k \to \bar{\beta} \leq 0$, then $\min\{-\beta^k, \alpha^k\} \to 0$. Indeed, if $\bar{\beta} < 0$, we have $\alpha^k \to 0$ and hence $\alpha^k < -\beta^k$ for all k sufficiently large, giving $\min\{-\beta^k, \alpha^k\} = \alpha^k \to 0$. On the other hand, if $\bar{\beta} = 0$, we also conclude that $\min\{-\beta^k, \alpha^k\} \to 0$, since $\alpha^k \geq 0$. This means that condition (1.14b) implies (1.13b), and thus CAKKT implies AKKT.*

Another known sequential optimality condition aims to control the sign of the Lagrange multipliers.

Definition 1.2.3 *[31] Let $\bar{x} \in \mathbb{R}^n$ be a feasible point for the problem (1.1). We say that $\bar{x}$ is a* Positive Approximate KKT *(PAKKT) point if there exist sequences $(x^k) \subset \mathbb{R}^n$ and $(\lambda^k) = \left(\lambda^{g,k}, \lambda^{h,k}\right) \subset \mathbb{R}^m_+ \times \mathbb{R}^p$ such that $x^k \to \bar{x}$,*

$$\nabla_x L(x^k, \lambda^{g,k}, \lambda^{h,k}) \to 0, \tag{1.15a}$$

$$\min\{-g(x^k), \lambda^{g,k}\} \to 0, \tag{1.15b}$$

$$\lambda_i^{g,k} g_i(x^k) > 0 \text{ if } \limsup_{k\to\infty} \frac{\lambda_i^{g,k}}{\delta_k} > 0, \tag{1.15c}$$

$$\lambda_j^{h,k} h_j(x^k) > 0 \text{ if } \limsup_{k\to\infty} \frac{|\lambda_j^{h,k}|}{\delta_k} > 0, \tag{1.15d}$$

where $\delta_k = \|(1, \lambda^k)\|_\infty$.

As well known in the literature, all the three sequential conditions above are necessary optimality conditions without any constraint qualification.

1.3 Relations between the MPCaC and the reformulated problems

In this section we present results that show some properties of the reformulated problems (2) and (3) and the equivalence between their solutions and the solutions of the cardinality problem (1). Such results are based on the ones presented in [2].

We start by noting that it is immediate that every feasible point of the mixed-integer problem (2) is also feasible for the relaxed problem (3), but the converse is clearly false. There is, however, a particular case in which the equivalence holds, as we can see from the next result.

Lemma 1.3.1 *Let $(\bar{x}, \bar{y})$ be a feasible point of the relaxed problem (3) and suppose that $\|\bar{x}\|_0 = \alpha$. Then,*

1. $e^T \bar{y} = n - \alpha$;
2. $\bar{y}_i = \begin{cases} 0, & \text{if } i \notin I_0(\bar{x}) \\ 1, & \text{if } i \in I_0(\bar{x}). \end{cases}$

So, $I_{00}(\bar{x},\bar{y}) = \emptyset$ and, in particular, $(\bar{x},\bar{y})$ is feasible for the mixed-integer problem (2).

Proof. Note first that $|I_0(\bar{x})| = n - \alpha$ and $\bar{y}_i = 0$ for $i \notin I_0(\bar{x})$. Thus, since $\bar{y}_i \leq 1$,

$$n - \alpha \leq e^T\bar{y} = \sum_{i \in I_0(\bar{x})} \bar{y}_i \leq n - \alpha,$$

yielding $e^T\bar{y} = n - \alpha$. Therefore, $\bar{y}_i = 1$ for $i \in I_0(\bar{x})$. □

Now we relate feasible points of the cardinality problem with feasible points of the reformulated ones.

Lemma 1.3.2 *Consider a point $\bar{x} \in \mathbb{R}^n$.*

1. *If $\bar{x}$ is feasible for the cardinality problem (1), then there exists $\bar{y} \in \mathbb{R}^n$ such that $(\bar{x},\bar{y})$ is feasible for the mixed-integer problem (2) and, hence, feasible for the relaxed problem (3). If, in addition, $\|\bar{x}\|_0 = \alpha$, then the vector $\bar{y}$ is unique;*

2. *If $(\bar{x},\bar{y})$ is feasible for (3), then $\bar{x}$ is feasible for (1).*

Proof. Denote $r = \|\bar{x}\|_0$ and $J = \{i \mid \bar{y}_i = 0\}$.

1. Defining $\bar{y} \in \mathbb{R}^n$ by $\bar{y}_i = 0$, if $i \notin I_0(\bar{x})$ and $\bar{y}_i = 1$, if $i \in I_0(\bar{x})$, we have $e^T\bar{y} = n - r \geq n - \alpha$ and $\bar{x}_i\bar{y}_i = 0$ for all $i = 1,\dots,n$, which implies that $(\bar{x},\bar{y})$ is feasible for (2). If $\|\bar{x}\|_0 = \alpha$, the uniqueness follows from Lemma 1.3.1.

2. If $(\bar{x},\bar{y})$ is feasible for (3), then $|J| \geq r$. So, $n - \alpha \leq e^T\bar{y} \leq n - r$, meaning that $\|\bar{x}\|_0 \leq \alpha$. Therefore, $\bar{x}$ is feasible for (1).

□

The following theorem states that the MPCaC problem has a global minimizer if and only if the reformulated problems have global minimizers too.

Theorem 1.3.3 *Consider a point $x^* \in \mathbb{R}^n$.*

1. *If x^* is a global solution of problem (1), then there exists a vector $y^* \in \mathbb{R}^n$ such that (x^*,y^*) is a global solution of problems (2) and (3). Moreover, for each reformulated problem, every feasible pair of the form $(x^*,\bar{y})$ is a global solution;*

2. *If (x^*,y^*) is a global solution of (2) or (3), then x^* is a global solution of (1).*

Proof.

1. Let x^* be a global solution of (1). By Lemma 1.3.2 there exists $y^* \in \mathbb{R}^n$ such that (x^*,y^*) is feasible for (2) and, hence, feasible for (3). Given an arbitrary feasible point (x,y) for (2), it is also feasible for (3) and again by Lemma 1.3.2, we conclude that x is feasible for (1). So, $f(x^*) \leq f(x)$ for such an x, proving that (x^*,y^*) is a global solution of (2) and (3). Since this argument does not depend on y, the second statement is valid as well.

2. If (x^*,y^*) is a global solution of (2), by Lemma 1.3.2 we have that x^* is feasible for (1). Given an arbitrary feasible point x for (1), there exists $y \in \mathbb{R}^n$ such that (x,y) is feasible for (2). So, $f(x^*) \leq f(x)$ for such an x, proving that x^* is a global solution of (1). On the other hand, if (x^*,y^*) is a global solution of (3), we can use the same arguments to see that x^* is a global solution of (1).

□

As a consequence of Theorem 1.3.3, we have that every global solution of (2) is also a global solution of (3). However, the converse is not necessarily true, as we can see in the example below.

Example 1.3.4 *Consider the relaxed problem*

$$\begin{array}{ll} \underset{x,y\in\mathbb{R}^3}{\text{minimize}} & (x_1-1)^2+(x_2-1)^2+x_3^2 \\ \text{subject to} & x_1 \leq 0, \\ & y_1+y_2+y_3 \geq 1, \\ & x_i y_i = 0,\ i=1,2,3, \\ & 0 \leq y_i \leq 1,\ i=1,2,3. \end{array}$$

Given any $t \in [0,1]$, the pair (x^*,y^*), with $x^*=(0,1,0)$ and $y^*=(1-t,0,t)$, is a global solution of the relaxed problem, but for $t \in (0,1)$ this point is not even feasible for the mixed-integer problem (2).

Now, let us discuss a result concerning the existence of global minimizers for the MPCaC and the reformulated problems. For this purpose, note first the closedness of the set defined by the cardinality constraint. Indeed, despite the fact that the function $x \mapsto \|x\|_0$ is not continuous, it is lower semicontinuous and hence, given $\alpha \geq 0$, the level set $\mathcal{C} = \{x \in \mathbb{R}^n \mid \|x\|_0 \leq \alpha\}$ is closed [47]. In Appendix A we give an alternative and direct proof of the closedness of the set defined by the cardinality constraint. Therefore, another consequence of Theorem 1.3.3 is the following result.

Theorem 1.3.5 *[2] Suppose that the feasible set $\Omega_0 = \{x \in X \mid \|x\|_0 \leq \alpha\}$ of the cardinality-constrained problem (1) is nonempty and that X is compact. Then the problems (1), (2) and (3) have a nonempty solution set.*

Now, let us analyze the relations among the problems by considering local solutions. We shall see that, differently from the global case, part of the equivalence is lost, but the relations of MPCaC with the relaxed problem remains valid.

Theorem 1.3.6 *[2] Let $x^* \in \mathbb{R}^n$ be a local minimizer of (1). Then there exists a vector $y^* \in \mathbb{R}^n$ such that the pair (x^*,y^*) is a local minimizer of (3).*

The next example shows that the converse of the above result is not valid.

Example 1.3.7 *Consider the MPCaC and the corresponding relaxed problem*

$$\begin{array}{ll} \underset{x\in\mathbb{R}^3}{\text{minimize}} & x_1^2+(x_2-1)^2+(x_3-1)^2 \\ \text{subject to} & x_1 \leq 0, \\ & \|x\|_0 \leq 2, \end{array}$$

$$\begin{array}{ll} \underset{x,y\in\mathbb{R}^3}{\text{minimize}} & x_1^2+(x_2-1)^2+(x_3-1)^2 \\ \text{subject to} & x_1 \leq 0, \\ & y_1+y_2+y_3 \geq 1, \\ & x_i y_i = 0,\ i=1,2,3, \\ & 0 \leq y_i \leq 1,\ i=1,2,3. \end{array}$$

Fix $t \in (0,1)$ and define $x^*=(0,1,0)$ and $y^*=(1-t,0,t)$. We claim that the pair (x^*,y^*) is a local solution of the relaxed problem. Indeed, if (x,y) is sufficiently close to (x^*,y^*), then $y_1 \neq 0$ and $y_3 \neq 0$, which implies that $x_1=0$ and $x_3=0$. So, the objective value at (x,y) is $(x_2-1)^2+1 \geq 1$, proving the claim. Nevertheless, x^* is not a local minimizer of the MPCaC, because we can consider a point $x_\delta=(0,1,\delta)$ with $\delta \in (0,1)$ as close to $x^*=(0,1,0)$ as we want whose objective value at this point is $(\delta-1)^2<1$.

Note that, in the above example, there are infinitely many vectors y^* such that (x^*, y^*) is a local solution of the relaxed problem. This is the reason why x^* is not a local minimizer of the MPCaC, as we can see from the next result.

Theorem 1.3.8 *Let (x^*, y^*) be a local minimizer of problem (3). Then $\|x^*\|_0 = \alpha$ if and only if y^* is unique, that is, if there is exactly one y^* such that (x^*, y^*) is a local minimizer of (3). In this case, the components of y^* are binary and x^* is a local minimizer of (1).*

Proof. The "only if" part and the claim that y^* is a binary vector follow directly from Lemma 1.3.1. The "if" part and the proof that x^* is a local minimizer of (1) are in [2].

□

1.4 Relations between the MPCaC and other classes of optimization problems

This section provides a brief comparison between MPCaC, MPCC, MPVC and MPSC. Despite the similarities, there are significant differences between these classes of optimization problems.

Considering the set $X \subset \mathbb{R}^n$ defined by inequality and equality constraints, our relaxed problem can be rewritten as

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \leq 0, h(x) = 0, \\ & e^T y \geq n - \alpha, \\ & x_i y_i = 0, \ i = 1, \dots, n, \\ & 0 \leq y_i \leq 1, \ i = 1, \dots, n, \end{array} \tag{1.16}$$

where $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions. Using the same arguments as in [2,30], if the inequality constraint $g(x) \leq 0$ encompasses a positivity constraint $x \geq 0$, say, $g_i(x) = -x_i$, $i = 1, \dots, n$, we see that the problem (1.16) is exactly

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & f(x) \\ \text{subject to} & g_i(x) \leq 0, i = n+1, \dots, m, \\ & h(x) = 0, \\ & e^T y \geq n - \alpha, \\ & x_i \geq 0, \ i = 1, \dots, n, \\ & x_i y_i = 0, \ i = 1, \dots, n, \\ & 0 \leq y_i \leq 1, \ i = 1, \dots, n. \end{array} \tag{1.17}$$

Thus, in this very specific case, the MPCaC problem has the form of an MPCC problem [27,28], which is given by

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \leq 0, h(x) = 0, \\ & G(x) \geq 0, H(x) \geq 0, \\ & G_i(x) H_i(x) = 0, \ i = 1, \dots, n. \end{array} \tag{1.18}$$

Note that we made an abuse of notation, since the variable x of the problem (1.18) plays the role of the pair (x, y), of the problem (1.17). A similar discussion is presented at the end of Section 4.1 of [2] and in Section 5 of [30]. We emphasize, however, that a general

MPCaC problem cannot be stated as an MPCC, since we do not have the constraint $x \geq 0$.

Now, observe that if we replace the constraints $x_i y_i = 0$, $i = 1, \ldots, n$, in the problem (1.17) by $x_i y_i \leq 0$, we have exactly the same problem, since $x \geq 0$ and $y \geq 0$. Therefore this problem can be viewed as an MPVC problem [48], namely,

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \leq 0,\ h(x) = 0, \\ & H_i(x) \geq 0, G_i(x)H_i(x) \leq 0,\ i = 1, \ldots, n. \end{array} \tag{1.19}$$

Finally, consider an MPSC problem [29]

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \leq 0,\ h(x) = 0, \\ & G_i(x)H_i(x) = 0,\ i = 1, \ldots, n. \end{array} \tag{1.20}$$

In contrast to the above discussion, the relaxed problem (1.16), without any further assumption, may be stated as an MPSC problem.

In view of the above relations, we conclude that it is not possible to employ the stationarity concepts (weak and approximate) established for MPCC and MPVC to our MPCaC problem. On the other hand, despite MPCaC is a particular case of an MPSC problem, it may be not suitable to inherit the stationarity concepts of MPSC since this problem is more general than an MPCaC. Therefore, considering the specific features of an MPCaC problem, it becomes necessary to propose tailored stationarity concepts for this class of problems.

Capítulo 2

Weak stationarity conditions for Mathematical Programs with Cardinality Constraints: a unified approach

The aim of this chapter is to analyze constraint qualifications and propose stationarity conditions for MPCaC. For this purpose, we consider two cases: the linear case, if the set X is given by linear constraints in the relaxed problem (3), and the nonlinear case, if the set X is given by nonlinear constraints.

We start by showing that the most well known constraint qualifications, LICQ and MFCQ, are not satisfied almost anywhere.

Proposition 2.1 *Let $(\bar{x}, \bar{y})$ be a feasible point of the problem (3) and suppose that $\bar{x}_\ell \neq 0$ for some index $\ell \in \{1, \ldots, n\}$. Then $(\bar{x}, \bar{y})$ does not satisfy MFCQ and, therefore, it does not satisfy LICQ.*

Proof. Denote $\xi_i(x, y) = x_i y_i$ and $H_i(y) = -y_i$. Given $d = (u, v) \in \mathbb{R}^n \times \mathbb{R}^n$, we have $\nabla \xi_i(\bar{x}, \bar{y})^T d = \bar{y}_i u_i + \bar{x}_i v_i$ and $\nabla H_i(\bar{y})^T v = -v_i$. Since $\bar{x}_\ell \neq 0$, it holds $\bar{y}_\ell = 0$, which implies that the constraint H_ℓ is active at $\bar{y}$ and

$$\nabla \xi_\ell(\bar{x}, \bar{y})^T d = \bar{x}_\ell v_\ell \quad \text{and} \quad \nabla H_\ell(\bar{y})^T v = -v_\ell.$$

So, there is no $d \in \mathbb{R}^n \times \mathbb{R}^n$ satisfying $\nabla \xi_\ell(\bar{x}, \bar{y})^T d = 0$ and $\nabla H_\ell(\bar{y})^T d < 0$ simultaneously. This means that $(\bar{x}, \bar{y})$ cannot satisfy MFCQ and, hence, it does not satisfy LICQ as well.
□

We stress that, for MPCC problems, MFCQ is violated at every feasible point. This constraint qualification has also been analyzed for MPVC [48] and MPSC [29] problems.

In the next section, we show, for the linear case, that every minimizer of the relaxed problem (3) satisfies the KKT conditions.

2.1 MPCaC: the Linear Case

It is well known that a set defined by linear constraints naturally satisfies a constraint qualification in standard NLP. In particular, ACQ holds for this kind of constraints. Now we discuss what happens if we consider the relaxed problem (3) with X given by linear constraints. Note that in this case, we have linear (separable) constraints together with a coupling complementarity constraint.

We obtain in this section, as a direct consequence of the Theorems 1.1.7 and 1.1.8, the constraint qualification analysis for the relaxed problem (3) in the linear case.

Theorem 2.1.1 *Consider the problem (3), with X defined by linear (equality and/or inequality) constraints, and its feasible set*

$$\Omega = \{(x, y) \in X \times \mathbb{R}^n \mid e^T y \geq n - \alpha, \ x * y = 0, \ 0 \leq y \leq e\}.$$

Then, every feasible point $(\bar{x}, \bar{y}) \in \Omega$ satisfies GCQ. Moreover, If $I_{00}(\bar{x}, \bar{y}) = \emptyset$, then ACQ holds at $(\bar{x}, \bar{y})$.

Proof. It follows directly from Theorems 1.1.7 and 1.1.8. □

Remark 2.1 *Note that given an arbitrary function $f : \mathbb{R}^2 \to \mathbb{R}$, the problem of minimizing f over the set Ω of the Example 1.1.9, is an MPCC for which GCQ does not hold at the origin. Therefore, this example points out a significant difference between the class of problems we are considering in this work, MPCaC, and the closely related problems MPCC's. Another MPCC problem in which all but the complementarity constraint are linear and GCQ does not hold is presented in [24]. Hence, Theorem 2.1.1 evidences that MPCaC is less degenerate than MPCC. Concerning the ACQ, a similar relationship is found in [48], in which a variant of the standard MFCQ condition for MPVC, namely VC-MFCQ, implies standard ACQ, provided that $I_{00} = \emptyset$.*

Corollary 2.1.2 *Under the assumptions of Theorem 2.1.1, if $\|\bar{x}\|_0 = \alpha$, then ACQ holds at $(\bar{x}, \bar{y})$.*

Proof. It follows directly from Lemma 1.3.1. □

In real problems, when modeling the problem, one can choose the parameter α in such a way that the cardinality constraint is active. Thus, the above corollary suggests that MPCaC problems are even less degenerate if the hypothesis is not considered.

As we pointed out in Remark 1.2, the condition $I_{00}(\bar{x}, \bar{y}) = \emptyset$ is sufficient but not necessary for ACQ. There is, however, a situation in which the equivalence holds.

Proposition 2.1.3 *Consider the problem (3) with $X = \mathbb{R}^n$ and its feasible set*

$$\Omega = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid e^T y \geq n - \alpha, \ x * y = 0, \ 0 \leq y \leq e\}.$$

Given an arbitrary feasible point $(\bar{x}, \bar{y}) \in \Omega$, ACQ holds at $(\bar{x}, \bar{y})$ if and only if $I_{00}(\bar{x}, \bar{y}) = \emptyset$.

Proof. Denote

$$\theta(y) = n - \alpha - e^T y, \quad H(y) = -y, \quad \tilde{H}(y) = y - e \quad \text{and} \quad \xi(x, y) = x * y.$$

Then, given $d = (u, v) \in \mathbb{R}^n \times \mathbb{R}^n$, we have

$$\nabla \theta(\bar{y})^T v = -e^T v, \tag{2.1a}$$

$$\nabla H_i(\bar{y})^T v = -v_i, \tag{2.1b}$$

$$\nabla \tilde{H}_i(\bar{y})^T v = v_i, \tag{2.1c}$$

$$\nabla \xi_i(\bar{x}, \bar{y})^T d = \bar{y}_i u_i + \bar{x}_i v_i. \tag{2.1d}$$

Assume first that $I_{00}(\bar{x},\bar{y}) \neq \emptyset$, take an index $\ell \in I_{00}(\bar{x},\bar{y})$ and consider the vector $\bar{d} = (e_\ell, e_\ell) \in \mathbb{R}^n \times \mathbb{R}^n$. Let us prove that $\bar{d} \in D_\Omega(\bar{x},\bar{y})$. We have

$$\nabla\theta(\bar{y})^T e_\ell = -1 \quad \text{and} \quad \nabla H_i(\bar{y})^T e_\ell \leq 0$$

for all i. Moreover, the constraint $\tilde{H}_\ell$ is inactive at $(\bar{y})$ and $\nabla H_i(\bar{y})^T e_\ell = 0$ for all $i \neq \ell$. Note also that $\nabla\xi_i(\bar{x},\bar{y})^T \bar{d} = 0$ for all i. Therefore, $\bar{d} \in D_\Omega(\bar{x},\bar{y})$. We claim that $\bar{d} \notin T_\Omega(\bar{x},\bar{y})$. Indeed, given any $d = (u,v) \in T_\Omega(\bar{x},\bar{y})$, there exist sequences $(x^k,y^k) \subset \Omega$ and $t_k \to 0$ such that $\dfrac{(x^k,y^k)-(\bar{x},\bar{y})}{t_k} \to (u,v)$. This implies that

$$\frac{x^k_\ell}{t_k} = \frac{x^k_\ell - \bar{x}_\ell}{t_k} \to u_\ell \quad \text{and} \quad \frac{y^k_\ell}{t_k} = \frac{y^k_\ell - \bar{y}_\ell}{t_k} \to v_\ell.$$

So, $0 = \dfrac{x^k_\ell y^k_\ell}{t_k^2} \to u_\ell v_\ell$, yielding $u_\ell v_\ell = 0$. Therefore, $\bar{d} = (e_\ell, e_\ell) \notin T_\Omega(\bar{x},\bar{y})$ and, hence, ACQ does not hold at $(\bar{x},\bar{y})$. The converse follows directly from Theorem 1.1.7. □

A concluding remark of this section is that, in view of Theorem 2.1.1, with X defined by linear constraints, every minimizer of the relaxed problem (3) satisfies the KKT conditions. This fact, however, does not mean that weaker stationarity conditions (than KKT) are unnecessary or less important. They are of interest from both the theoretical and the practical viewpoint, as in the sparsity constrained optimization (if there is only the cardinality constraint). See [14,20,21] and references therein for a more detailed discussion.

We now turn our attention to the general nonlinear case, to be discussed in the next section.

2.2 MPCaC: the Nonlinear Case

In this section we present one of the main contributions of this work. We propose a unified approach that goes from the weakest to the strongest stationarity for the cardinality problem with general constraints. This approach, which will be called W_I-stationarity, is based on a given set of indices I such that the complementarity constraint is always satisfied. Moreover, different levels of stationarity can be obtained depending on the range for the set I. Besides, we prove that this condition is indeed weaker than the classical KKT condition, that is, every KKT point fulfills W_I-stationarity. We also point out that our definition generalizes the concepts of S- and M-stationarity presented in [2] for a proper choice of the index set I.

For this purpose, consider the MPCaC problem (1) with

$$X = \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}, \tag{2.2}$$

where $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions.

Differently from the linear case, if the set X is given by nonlinear constraints we cannot guarantee that the standard constraints qualifications are satisfied for the relaxed problem (3), making convenient the study of stationarity conditions weaker than KKT.

2.2.1 Constraint qualifications for MPCaC

We have proved at the beginning of this chapter that the most well known constraint qualifications, LICQ and MFCQ, are not satisfied almost anywhere. Even the weaker condition, ACQ, may fail to hold in a wide range of cardinality problems, as we saw in Proposition 2.1.3. In that situation, despite considering the simplest MPCaC problem, without constraints other than the cardinality constraint itself, ACQ does not hold if there is an index i for which $\bar{x}_i = \bar{y}_i = 0$.

Since MPCaC are highly degenerate problems, even GCQ, the weakest constraint qualification, may be violated. Indeed, it can be seen in the following example where the set X is given by a single quadratic constraint.

Example 2.2.1 *Consider the MPCaC and the corresponding relaxed problem*

$$\begin{array}{cl} \underset{x\in\mathbb{R}^2}{\text{minimize}} & x_1 + x_2 \\ \text{subject to} & -x_1 + x_2^2 \le 0, \\ & \|x\|_0 \le 1, \end{array} \qquad \begin{array}{cl} \underset{x,y\in\mathbb{R}^2}{\text{minimize}} & x_1 + x_2 \\ \text{subject to} & -x_1 + x_2^2 \le 0, \\ & y_1 + y_2 \ge 1, \\ & x_i y_i = 0, \ i = 1, 2, \\ & 0 \le y_i \le 1, \ i = 1, 2. \end{array}$$

Note that $x^* = (0, 0)$ is the unique global solution of the cardinality problem and, defining $y^* = (1, 0)$, the pair (x^*, y^*) is a global solution of the relaxed problem. However, this pair does not satisfy GCQ, because otherwise it would be a KKT point, that is, the expression

$$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \mu_1 \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \mu_2 \begin{pmatrix} 0 \\ 0 \\ -1 \\ -1 \end{pmatrix} + \mu_3 \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \end{pmatrix} + \mu_4 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

would vanish for some Lagrange multipliers $\mu \in \mathbb{R}^4$, $\lambda \in \mathbb{R}$, what is impossible due to its second row. Figure 2.1 illustrates this example.

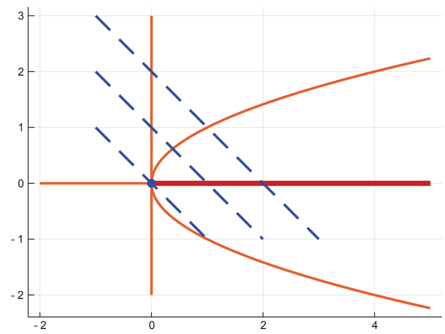

Figura 2.1: Illustration of MPCaC problem in Example 2.2.1. The feasible set is the thick horizontal line and the level curves of the objective function are shown as dashed lines. The global solution x^* is the origin.

The violation of GCQ in the previous example is due to the existence of a nonlinear constraint, since in the linear case GCQ is always satisfied. Moreover, the example shows that the classical stationarity conditions may not be able to detect the solution.

2.2.2 Stationarity conditions for MPCaC

As we have seen before, except in special cases, e.g., if X is polyhedral and convex, we do not have a constraint qualification for the relaxed problem (3). So, even in simple cases the standard KKT conditions are not necessary optimality conditions.

Thus, in this section we define weaker stationarity concepts to deal with this class of problems. In fact, we propose a unified approach that goes from the weakest to the strongest stationarity.

For ease of presentation we consider the functions (some of which have already been used in the proof of Proposition 2.1.3) $\theta : \mathbb{R}^n \to \mathbb{R}$, $H, \tilde{H}, G : \mathbb{R}^n \to \mathbb{R}^n$ given by

$$\theta(y) = n - \alpha - e^T y, \ H(y) = -y, \ \tilde{H}(y) = y - e \text{ and } G(x) = x.$$

Then we can rewrite the relaxed problem (3) as

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \leq 0, h(x) = 0, \\ & \theta(y) \leq 0, \\ & H(y) \leq 0, \tilde{H}(y) \leq 0, \\ & G(x) * H(y) = 0. \end{array} \tag{2.3}$$

Given a feasible point $(\bar{x}, \bar{y})$ for the problem (2.3) and a set of indices I such that

$$I_{0+}(\bar{x}, \bar{y}) \cup I_{01}(\bar{x}, \bar{y}) \subset I \subset I_0(\bar{x}), \tag{2.4}$$

we have that $G_i(\bar{x}) = 0$ for $i \in I$ and $H_i(\bar{y}) = 0$ for $i \in I_{00}(\bar{x}, \bar{y}) \cup I_{\pm 0}(\bar{x}, \bar{y})$. This suggests to consider an auxiliary problem by removing the challenging constraint $G(x) * H(y) = 0$ and including alternative ones that ensure the null product. We then define the *I-Tightened* Nonlinear Problem at $(\bar{x}, \bar{y})$ by

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \leq 0, h(x) = 0, \\ & \theta(y) \leq 0, \\ & \tilde{H}(y) \leq 0, \\ & H_i(y) \leq 0, \ i \in I_{0+}(\bar{x}, \bar{y}) \cup I_{01}(\bar{x}, \bar{y}), \\ & H_i(y) = 0, \ i \in I_{00}(\bar{x}, \bar{y}) \cup I_{\pm 0}(\bar{x}, \bar{y}), \\ & G_i(x) = 0, \ i \in I. \end{array} \tag{2.5}$$

This problem will be also indicated by $\text{TNLP}_I(\bar{x}, \bar{y})$ and, whenever there is no chance for ambiguity, it will be referred to simply as *tightened problem*. Note that we tighten only those constraints that are involved with the complementarity constraint $G(x) * H(y) = 0$, by converting the active inequalities H_i's into equalities and incorporating the equality constraints G_i's. The upper set $I_0(\bar{x})$ in the range given by (2.4) guarantees that we do not incorporate a constraint $G_i(x) = 0$ for some i such that $G_i(\bar{x}) = \bar{x}_i \neq 0$.

The following lemma is a straightforward consequence of the definition of $\text{TNLP}_I(\bar{x}, \bar{y})$.

Lemma 2.2.2 *Consider the tightened problem (2.5). Then,*

1. *all the inequalities defined by H_i, $i \in I_{0+}(\bar{x}, \bar{y}) \cup I_{01}(\bar{x}, \bar{y})$, are inactive at $(\bar{x}, \bar{y})$;*
2. *$(\bar{x}, \bar{y})$ is feasible for $TLNP_I(\bar{x}, \bar{y})$;*
3. *every feasible point of (2.5) is feasible for (2.3);*
4. *if $(\bar{x}, \bar{y})$ is a global (local) minimizer of (2.3), then it is also a global (local) minimizer of $TNLP_I(\bar{x}, \bar{y})$.*

The Lagrangian function associated with $\mathrm{TNLP}_I(\bar{x},\bar{y})$ is the function

$$\mathcal{L}_I : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{|I|} \to \mathbb{R}$$

given by

$$\begin{aligned}\mathcal{L}_I(x,y,\lambda^g,\lambda^h,\lambda^\theta,\lambda^H,\lambda^{\tilde{H}},\lambda^G_I) = {} & f(x) + (\lambda^g)^T g(x) + (\lambda^h)^T h(x) + \lambda^\theta \theta(y) \\ & + (\lambda^H)^T H(y) + (\lambda^{\tilde{H}})^T \tilde{H}(y) + (\lambda^G_I)^T G_I(x).\end{aligned}$$

Note that the tightened problem, and hence its Lagrangian, depends on the index set I, which in turn depends on the point $(\bar{x},\bar{y})$. It should be also noted that

$$\nabla_{x,y}\mathcal{L}_I(x,y,\lambda) = \begin{pmatrix} \nabla_x L(x,\lambda^g,\lambda^h) + \sum_{i\in I} \lambda^G_i e_i \\ -\lambda^\theta e - \lambda^H + \lambda^{\tilde{H}} \end{pmatrix}. \tag{2.6}$$

Weak stationarity

Our weaker stationarity concept for the relaxed problem (2.3) is then defined in terms of the tightened problem as follows.

Definition 2.2.3 *Consider a feasible point $(\bar{x},\bar{y})$ of the relaxed problem (2.3) and a set of indices I satisfying (2.4). We say that $(\bar{x},\bar{y})$ is I-weakly stationary (W_I-stationary) for this problem if there exists a vector*

$$\lambda = (\lambda^g,\lambda^h,\lambda^\theta,\lambda^H,\lambda^{\tilde{H}},\lambda^G_I) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^n_+ \times \mathbb{R}^{|I|}$$

such that

1. $\nabla_{x,y}\mathcal{L}_I(\bar{x},\bar{y},\lambda) = 0$;
2. $(\lambda^g)^T g(\bar{x}) = 0$;
3. $\lambda^\theta \theta(\bar{y}) = 0$;
4. $(\lambda^{\tilde{H}})^T \tilde{H}(\bar{y}) = 0$;
5. $\lambda^H_i = 0$ *for all* $i \in I_{0+}(\bar{x},\bar{y}) \cup I_{01}(\bar{x},\bar{y})$.

Remark 2.2 *In view of (2.6), the first item of Definition 2.2.3 means that*

$$\nabla f(\bar{x}) + \sum_{i=1}^m \lambda^g_i \nabla g_i(\bar{x}) + \sum_{i=1}^p \lambda^h_i \nabla h_i(\bar{x}) + \sum_{i\in I} \lambda^G_i e_i = 0 \tag{2.7}$$

$$-\lambda^\theta e - \lambda^H + \lambda^{\tilde{H}} = 0. \tag{2.8}$$

Items 2, 3 and 4 represent the standard KKT complementarity conditions for the inequality constraints $g(x) \le 0$, $\theta(y) \le 0$ and $\tilde{H}(y) \le 0$, respectively, of the tightened problem (2.5). In view of Lemma 2.2.2(1), the last item also represents the KKT complementarity conditions for the constraints $H_i(y) \le 0$, $i \in I_{0+}(\bar{x},\bar{y}) \cup I_{01}(\bar{x},\bar{y})$, of the tightened problem.

As an immediate consequence of Remark 2.2 we have the following characterization of W_I-stationarity for the relaxed problem in terms of the stationarity for the tightened problem.

Proposition 2.2.4 *Let $(\bar{x}, \bar{y})$ be a feasible point of the relaxed problem (2.3). Then, $(\bar{x}, \bar{y})$ is W_I-stationary if and only if it is a KKT point for the tightened problem (2.5).*

Proof. It follows from the feasibility of $(\bar{x}, \bar{y})$, stated in Lemma 2.2.2(2), the comments in Remark 2.2, the fact that $\mathcal{L}_I$ is the Lagrangian of $\mathrm{TNLP}_I(\bar{x}, \bar{y})$ and that the non-negativeness of the multipliers corresponding to the inequality constraints $H_i(x, y) \leq 0$, $i \in I_{0+}(\bar{x}, \bar{y}) \cup I_{01}(\bar{x}, \bar{y})$, is equivalent to the last item of Definition 2.2.3, because of Lemma 2.2.2(1). □

Note that in view of Proposition 2.2.4 we could have defined W_I-stationarity simply as KKT for the tightened problem (2.5). Nevertheless, we prefer as in Definition 2.2.3 in order to have its last condition 5 explicitly, instead of hiding it in the complementarity condition. This way of stating weak stationarity is also similar to that used in the MPCC setting, see [27,49].

In the next result we justify why Definition 2.2.3 is considered a weaker stationarity concept than KKT for the relaxed problem.

Theorem 2.2.5 *Suppose that $(\bar{x}, \bar{y})$ is a KKT point for the relaxed problem (2.3). Then $(\bar{x}, \bar{y})$ is W_I-stationary for every I satisfying (2.4).*

Proof. Denoting $\xi(x, y) = x * y$, we have $\nabla \xi_i(\bar{x}, \bar{y}) = \begin{pmatrix} \bar{y}_i e_i \\ \bar{x}_i e_i \end{pmatrix}$. By the hypothesis, there exists a vector

$$(\lambda^g, \lambda^h, \lambda^\theta, \mu, \lambda^{\tilde{H}}, \lambda^\xi) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}^n_+ \times \mathbb{R}^n_+ \times \mathbb{R}^n$$

such that

$$\begin{pmatrix} \nabla f(\bar{x}) \\ 0 \end{pmatrix} + \sum_{i=1}^m \lambda_i^g \begin{pmatrix} \nabla g_i(\bar{x}) \\ 0 \end{pmatrix} + \sum_{i=1}^p \lambda_i^h \begin{pmatrix} \nabla h_i(\bar{x}) \\ 0 \end{pmatrix} + \lambda^\theta \nabla \theta(\bar{y})$$
$$+ \sum_{i=1}^n \mu_i \nabla H_i(\bar{y}) + \sum_{i=1}^n \lambda_i^{\tilde{H}} \nabla \tilde{H}_i(\bar{y}) + \sum_{i=1}^n \lambda_i^\xi \nabla \xi_i(\bar{x}, \bar{y}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which means that

$$\nabla f(\bar{x}) + \sum_{i=1}^m \lambda_i^g \nabla g_i(\bar{x}) + \sum_{i=1}^p \lambda_i^h \nabla h_i(\bar{x}) + \sum_{i \in I_{01}} \lambda_i^\xi e_i + \sum_{i \in I_{0+}} \lambda_i^\xi \bar{y}_i e_i = 0, \tag{2.9}$$

$$-\lambda^\theta e - \mu + \lambda^{\tilde{H}} + \sum_{i \in I_{\pm 0}} \lambda_i^\xi \bar{x}_i e_i = 0, \tag{2.10}$$

where, for simplicity, we denoted $I_{0+} = I_{0+}(\bar{x}, \bar{y})$ and $I_{\pm 0} = I_{\pm 0}(\bar{x}, \bar{y})$.

Moreover, we have

$$(\lambda^g)^T g(\bar{x}) = \lambda^\theta \theta(\bar{y}) = \mu^T H(\bar{y}) = (\lambda^{\tilde{H}})^T \tilde{H}(\bar{y}) = 0. \tag{2.11}$$

Defining

$$\lambda_i^G = \begin{cases} \lambda_i^\xi \bar{y}_i, & \text{for } i \in I_{01} \cup I_{0+}, \\ 0, & \text{for } i \in I \setminus I_{01} \cup I_{0+} \end{cases} \quad \text{and} \quad \lambda_i^H = \begin{cases} \mu_i, & \text{for } i \in I_0, \\ \mu_i - \lambda_i^\xi \bar{x}_i, & \text{for } i \in I_{\pm 0}, \end{cases}$$

we conclude immediately that $(\lambda^g, \lambda^h, \lambda^\theta, \lambda^H, \lambda^{\tilde{H}}, \lambda_I^G)$ satisfies (2.7) and items 2–5 of Definition 2.2.3. To finish the proof, note that

$$\begin{aligned}
-\lambda^\theta e - \lambda^H + \lambda^{\tilde{H}} &= -\lambda^\theta e + \lambda^{\tilde{H}} - \sum_{i\in I_0} \mu_i e_i - \sum_{i\in I_{\pm 0}} (\mu_i - \lambda_i^\xi \bar{x}_i) e_i \\
&= -\lambda^\theta e + \lambda^{\tilde{H}} - \mu + \sum_{i\in I_{\pm 0}} \lambda_i^\xi \bar{x}_i e_i
\end{aligned}$$

which, in view of (2.10), gives (2.8). □

As we have discussed before, a minimizer of the relaxed problem does not necessarily satisfy the KKT conditions mostly because of the complementarity constraint, which may prevent the fulfillment of constraint qualifications. This fact was illustrated in Example 2.2.1. Let us revisit this example in the light of our W_I-stationarity concept. Now we can capture the minimizer by means of the KKT conditions for the tightened problem.

Example 2.2.6 *Consider the MPCaC and the corresponding relaxed problem presented in Example 2.2.1.*

$$\begin{array}{ll}
\underset{x\in\mathbb{R}^2}{\text{minimize}} & x_1 + x_2 \\
\text{subject to} & -x_1 + x_2^2 \le 0, \\
& \|x\|_0 \le 1,
\end{array}
\qquad
\begin{array}{ll}
\underset{x,y\in\mathbb{R}^2}{\text{minimize}} & x_1 + x_2 \\
\text{subject to} & -x_1 + x_2^2 \le 0, \\
& y_1 + y_2 \ge 1, \\
& x_i y_i = 0,\ i = 1,2, \\
& 0 \le y_i \le 1,\ i = 1,2.
\end{array}$$

We saw that $x^* = (0,0)$ is the unique global solution of MPCaC and (x^*, y^*), with $y^* = (1,0)$, is a global solution of the relaxed problem. Besides, this pair does not satisfy GCQ and it is not a KKT point. Let us formulate the tightened problem for $I = I_0(x^*)$. We have, suppressing the arguments x^* and (x^*, y^*),

$$I_0 = \{1,2\}, \quad I_{01} = \{1\}, \quad I_{00} = \{2\} \quad \textit{and} \quad I_{\pm 0} = I_{0+} = \emptyset.$$

So, the $\mathrm{TNLP}_I(x^*, y^*)$ is given by

$$\begin{array}{ll}
\underset{x,y\in\mathbb{R}^2}{\text{minimize}} & x_1 + x_2 \\
\text{subject to} & -x_1 + x_2^2 \le 0, \\
& 1 - y_1 - y_2 \le 0, \\
& y_i - 1 \le 0,\ i = 1,2, \\
& -y_1 \le 0, \\
& -y_2 = 0, \\
& x_i = 0,\ i = 1,2.
\end{array}$$

Defining

$$\lambda^g = 0, \quad \lambda^\theta = 1, \quad \lambda^{\tilde{H}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \lambda^G = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \quad \textit{and} \quad \lambda^H = \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

we have

$$\nabla_{x,y} \mathcal{L}_I(x^*, y^*, \lambda) = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -1 \\ -1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1 \end{pmatrix} = 0.$$

Thus, (x^*, y^*) is a KKT point for the tightened problem, that is, it is W_I-stationary for the relaxed problem. Note also that ACQ holds at this point for $\mathrm{TNLP}_I(x^*, y^*)$.

As a matter of fact, we can state W_I-stationarity using only the original variables x. This follows from the next result.

Proposition 2.2.7 *If $(\bar{x}, \bar{y})$ is W_I-stationary for the relaxed problem (2.3) then there exists a vector*

$$(\lambda^g, \lambda^h, \gamma) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^{|I|}$$

such that

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i^g \nabla g_i(\bar{x}) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(\bar{x}) + \sum_{i \in I} \gamma_i e_i = 0, \tag{2.12}$$

$$(\lambda^g)^T g(\bar{x}) = 0. \tag{2.13}$$

Conversely, if $\bar{x}$ satisfies the conditions (2.12) and (2.13) then every feasible point $(\bar{x}, \hat{y})$ for the relaxed problem (2.3) is W_I-stationary.

Proof. Suppose first that $(\bar{x}, \bar{y})$ is W_I-stationary, that is, it satisfies Definition 2.2.3. In view of (2.7), if we define $\gamma = \lambda_I^G$, we obtain (2.12). Moreover, (2.13) follows from Definition 2.2.3(2).

On the other hand, assume that $\bar{x}$ satisfies the conditions (2.12) and (2.13) and $(\bar{x}, \hat{y})$ is feasible for (2.3). Then, defining $\lambda_I^G = \gamma$ and setting $\lambda^\theta = 0$, $\lambda^H = \lambda^{\tilde{H}} = 0$, we obtain (2.7) and (2.8). Therefore, $(\bar{x}, \hat{y})$ is a W_I-stationary point. □

Remark 2.3 *Note that the conditions (2.12) and (2.13) generalize the concepts of S- and M-stationarity presented in [2] if we consider $I = I_{0+}(\bar{x}, \bar{y}) \cup I_{01}(\bar{x}, \bar{y})$ and $I = I_0(\bar{x})$, respectively. However, we stress that here we have different levels of weak stationarity, according to the set I between $I_{0+}(\bar{x}, \bar{y}) \cup I_{01}(\bar{x}, \bar{y})$ and $I_0(\bar{x})$. Moreover, an interesting and direct consequence of Proposition 2.2.7 is that stationarity gets stronger as the index set reduces (cf. proposition below).*

Proposition 2.2.8 *Let $(\bar{x}, \bar{y})$ be feasible for the relaxed problem (2.3). If*

$$I_{0+}(\bar{x}, \bar{y}) \cup I_{01}(\bar{x}, \bar{y}) \subset I' \subset I \subset I_0(\bar{x}),$$

then $W_{I'}$-stationarity implies W_I-stationarity.

Proposition 2.2.7 also makes easier the task of verifying whether a point is or is not W_I-stationary, as we can see in the next example.

Example 2.2.9 *Consider the following MPCaC and the associated relaxed problem.*

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \leq 0, \\ & \|x\|_0 \leq n-1, \end{array} \qquad \begin{array}{ll} \underset{x,y \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \leq 0, \\ & e^T y \geq 1, \\ & x * y = 0, \\ & 0 \leq y \leq e, \end{array}$$

where the functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^{n-1}$ are given by $f(x) = e^T x$ and $g_i(x) = -x_i + x_n^2$, $i = 1, \ldots, n-1$. Note that every feasible point x of MPCaC satisfies $x_n = 0$ because otherwise we would have $\|x\|_0 = n$. Therefore, $x^* = 0$ is the unique global *solution of MPCaC and (x^*, y^*), with $y^* = e_1$, is a global solution of the relaxed problem. Besides, we have*

$$I_0 = \{1, \dots, n\}, \quad I_{01} = \{1\}, \quad I_{00} = \{2, \dots, n\} \quad \text{and} \quad I_{0+} = \emptyset.$$

Consider an arbitrary index set I such that $I_{01} \subset I \subset I_0$. Then, the point (x^*, y^*) is W_I-stationary if and only if $n \in I$.

Indeed, suppose first that $n \in I$. Defining $\lambda_i^g = 1$, $i = 1, \dots, n-1$, $\gamma_n = -1$ and $\gamma_i = 0$, $i \in I \setminus \{n\}$, we have

$$\nabla f(x^*) + \sum_{i=1}^{n-1} \lambda_i^g \nabla g_i(x^*) + \sum_{i \in I} \gamma_i e_i = \begin{pmatrix} \tilde{e} \\ 1 \end{pmatrix} + \begin{pmatrix} -\tilde{e} \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

where $\tilde{e}$ denotes the vector of all ones in $\mathbb{R}^{n-1}$. So, we obtain (2.12) and (2.13). On the other hand, if $n \notin I$, then there is no vector $(\lambda^g, \gamma) \in \mathbb{R}_+^{n-1} \times \mathbb{R}^{|I|}$ such that the expression

$$\nabla f(x^*) + \sum_{i=1}^{n-1} \lambda_i^g \nabla g_i(x^*) + \sum_{i \in I} \gamma_i e_i = \begin{pmatrix} \tilde{e} \\ 1 \end{pmatrix} - \begin{pmatrix} \lambda^g \\ 0 \end{pmatrix} + \begin{pmatrix} \gamma_I \\ 0 \end{pmatrix}$$

vanishes. Therefore, condition (2.12) can never be satisfied.

Another feature of Proposition 2.2.7 is that it suggests that our W_I-stationarity concept can be formulated in terms of x only, for the original cardinality-constrained problem.

Definition 2.2.10 *Consider the MPCaC problem defined in (1) with the set X given by (2.2) and a feasible point $\bar{x}$ for this problem. Given $I \subset I_0(\bar{x})$, we say that the point $\bar{x}$ is I-weakly stationary (W_I-stationary) for problem (1) if there exists a vector $(\lambda^g, \lambda^h, \gamma) \in \mathbb{R}_+^m \times \mathbb{R}^p \times \mathbb{R}^{|I|}$ satisfying the conditions (2.12) and (2.13).*

We can state here a result analogous to Proposition 2.2.4, relating the W_I-stationarity for the cardinality problem with the classical notion of stationarity for some tightened problem. More precisely, given a feasible point $\bar{x}$ of (1), consider the associated tightened problem, indicated by $\mathrm{TNLP}_I(\bar{x})$,

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & g(x) \le 0, \\ & h(x) = 0, \\ & x_i = 0, \ i \in I. \end{array} \tag{2.14}$$

Proposition 2.2.11 *Let $\bar{x}$ be a feasible point of the MPCaC problem (1). Then, $\bar{x}$ is W_I-stationary if and only if it is a KKT point for the tightened problem (2.14).*

Despite the idea of working only on x, instead of adding the artificial variable y, seems to be better and simpler, this approach is not suitable here. Indeed, it should be noted that tightening the MPCaC problem does not necessarily preserve the minimizers. For example, if $|I| < n - \alpha$, then the feasible set of $\mathrm{TNLP}_I(\bar{x})$ is no longer a subset of the feasible set of problem(1). This means that we loose here properties like the last items of Lemma 2.2.2. Another issue in this setting is that, unlike the relaxed formulation, for which we shall define stronger stationarity, we cannot define stronger stationarity concepts for the cardinality problem. We have, however, the following result if $|I| \ge n - \alpha$.

Proposition 2.2.12 *Let x^* be a minimizer of the MPCaC problem (1). If $I \subset I_0(x^*)$ is such that $|I| \geq n - \alpha$, then x^* is also a minimizer of $TNLP_I(x^*)$. In particular, x^* is a minimizer of $TNLP_{I_0(x^*)}(x^*)$.*

We have to mention that Propositions 2.2.11 and 2.2.12, specialized for $I = I_0(\bar{x})$, are presented in [2].

We conclude this section with an important consequence of the Remark 2.3 and Proposition 2.2.8 that refers to the fulfillment of W_I-stationarity under some CQ for MPCaC. In [30] the authors introduced several MPCaC-tailored constraint qualifications (CC-CQ). In particular, they proved that S-stationarity (see Remark 2.3) holds at minimizers under each one of the proposed CC-CQ. Therefore, since S-stationarity at a point $(\bar{x}, \bar{y})$ is equivalent to W_{I_S}-stationarity, with $I_S = I_{0+}(\bar{x}, \bar{y}) \cup I_{01}(\bar{x}, \bar{y})$, and W_{I_S} implies any other W_I in the range $I_S \subset I \subset I_0(\bar{x})$, we have that a local minimizer of the relaxed problem (3), for which a CC-CQ holds, satisfies W_I.

Strong stationarity

In this section we define strong stationarity as a special case of Definition 2.2.3 and prove, among other results, that it coincides with the classical KKT conditions for the relaxed problem.

Definition 2.2.13 *Consider a feasible point $(\bar{x}, \bar{y})$ for the relaxed problem (2.3). We say that $(\bar{x}, \bar{y})$ is strongly stationary (S-stationary) for this problem if it is W_I-stationary with $I = I_{0+}(\bar{x}, \bar{y}) \cup I_{01}(\bar{x}, \bar{y})$, that is, if it satisfies Definition 2.2.3 with this specific index set.*

Let us start by showing the equivalence between S-stationarity and KKT.

Theorem 2.2.14 *[2] Let $(\bar{x}, \bar{y})$ be a feasible point for the relaxed problem (2.3). Then, $(\bar{x}, \bar{y})$ is S-stationary if and only if it satisfies the usual KKT conditions for this problem.*

Proof. Note first that sufficiency follows directly from Theorem 2.2.5. To prove the necessity, note that by Proposition 2.2.7 there exists a vector

$$(\lambda^g, \lambda^h, \gamma) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}^{|I_{0+} \cup I_{01}|}$$

such that

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i^g \nabla g_i(\bar{x}) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(\bar{x}) + \sum_{i \in I_{0+} \cup I_{01}} \gamma_i e_i = 0,$$
$$(\lambda^g)^T g(\bar{x}) = 0.$$

Defining $\lambda^\theta = 0$, $\lambda^H = \lambda^{\tilde{H}} = 0$ and

$$\lambda_i^\xi = \begin{cases} \gamma_i / \bar{y}_i, & \text{for } i \in I_{0+} \cup I_{01}, \\ 0, & \text{for } i \in I_{\pm 0} \cup I_{00}, \end{cases}$$

we conclude that the vector

$$(\lambda^g, \lambda^h, \lambda^\theta, \lambda^H, \lambda^{\tilde{H}}, \lambda^\xi) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}^n_+ \times \mathbb{R}^n_+ \times \mathbb{R}^n$$

fulfills the conditions

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i^g \nabla g_i(\bar{x}) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(\bar{x}) + \sum_{i \in I_{0+} \cup I_{01}} \lambda_i^\xi \bar{y}_i e_i = 0, \tag{2.15}$$

$$-\lambda^\theta e - \lambda^H + \lambda^{\tilde{H}} + \sum_{i \in I_{\pm 0}} \lambda_i^\xi \bar{x}_i e_i = 0, \tag{2.16}$$

$$(\lambda^g)^T g(\bar{x}) = \lambda^\theta \theta(\bar{y}) = (\lambda^H)^T H(\bar{y}) = (\lambda^{\tilde{H}})^T \tilde{H}(\bar{y}) = 0. \tag{2.17}$$

Since we have defined $\xi(x, y) = x * y$, we see that (2.15) and (2.16) are equivalent to

$$\begin{pmatrix} \nabla f(\bar{x}) \\ 0 \end{pmatrix} + \sum_{i=1}^{m} \lambda_i^g \begin{pmatrix} \nabla g_i(\bar{x}) \\ 0 \end{pmatrix} + \sum_{i=1}^{p} \lambda_i^h \begin{pmatrix} \nabla h_i(\bar{x}) \\ 0 \end{pmatrix} + \lambda^\theta \nabla \theta(\bar{y})$$
$$+ \sum_{i=1}^{n} \lambda_i^H \nabla H_i(\bar{y}) + \sum_{i=1}^{n} \lambda_i^{\tilde{H}} \nabla \tilde{H}_i(\bar{y}) + \sum_{i=1}^{n} \lambda_i^\xi \nabla \xi_i(\bar{x}, \bar{y}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and, consequently, $(\bar{x}, \bar{y})$ is a KKT point for the relaxed problem (2.3). □

As an immediate consequence of Theorem 2.2.14 and Proposition 2.2.4 we have the equivalence between the KKT conditions for the relaxed and tightened problems if $I = I_{0+}(\bar{x}, \bar{y}) \cup I_{01}(\bar{x}, \bar{y})$.

Corollary 2.2.15 *Consider a feasible point $(\bar{x}, \bar{y})$ for the relaxed problem (2.3) and $I = I_{0+}(\bar{x}, \bar{y}) \cup I_{01}(\bar{x}, \bar{y})$. Then, $(\bar{x}, \bar{y})$ is a KKT point for (2.3) if and only if it is a KKT point for the tightened problem $TNLP_I(\bar{x}, \bar{y})$, given by (2.5).*

It should be noted that, if $I_{00}(\bar{x}, \bar{y}) = \emptyset$ then all stationarity conditions presented here are the same and correspond to KKT for the relaxed problem. This follows directly from (2.4). In this case, we say that $(\bar{x}, \bar{y})$ satisfies the *strict complementarity*.

On the other hand, given an arbitrary index set I satisfying (2.4), we have the following result.

Proposition 2.2.16 *Suppose that $(\bar{x}, \bar{y})$ is W_I-stationary for the relaxed problem (2.3) with the associated vector of multipliers $\lambda = (\lambda^g, \lambda^h, \lambda^\theta, \lambda^H, \lambda^{\tilde{H}}, \lambda_I^G)$. If $\lambda^G_{I_{00}(\bar{x}, \bar{y})} = 0$, then $(\bar{x}, \bar{y})$ is S-stationary.*

Proof. Note that Definition 2.2.3, relation (2.7) and the condition $\lambda^G_{I_{00}(\bar{x}, \bar{y})} = 0$ allow us to write

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i^g \nabla g_i(\bar{x}) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(\bar{x}) + \sum_{i \in I_{0+} \cup I_{01}} \lambda_i^G e_i = 0,$$
$$(\lambda^g)^T g(\bar{x}) = 0.$$

Using Proposition 2.2.7 with $I = I_{0+} \cup I_{01}$, we see that $(\bar{x}, \bar{y})$ is W$_I$-stationary, which means S-stationary. □

Note that, by Definition 2.2.13, S-stationarity at a point $(\bar{x}, \bar{y})$ is W_{I_S}-stationarity with $I_S = I_{0+}(\bar{x}, \bar{y}) \cup I_{01}(\bar{x}, \bar{y})$. Therefore, in view of Proposition 2.2.8, we have that S-stationarity implies W$_I$-stationarity for any I satisfying $I_S \subset I \subset I_0(\bar{x})$. Moreover, this implication is strict in view of Example 2.2.6.

Capítulo 3

Ongoing research on sequential optimality conditions for MPCaC and future perspectives

In this chapter, we present the theoretical results concerning an ongoing research on sequential optimality condition, associated with the weak stationarity condition presented in Chapter 2. The algorithmic consequences, among other future perspectives, are preliminarily discussed in the Section 3.3.

As we have seen in the previous chapter, W_I-stationarity is weaker than KKT. Thereby, it seems natural to ask if it is a necessary optimality condition. That is, can we ensure that a minimizer of the relaxed problem is W_I-stationary for some index set I satisfying (2.4)? The answer is no, as illustrated in the following example.

Example 3.1 *Consider the MPCaC and the corresponding relaxed problem given below.*

$$\begin{array}{ll} \underset{x\in\mathbb{R}^3}{\text{minimize}} & x_1 \\ \text{subject to} & (1-x_1)^3 + x_3^2 \le 0, \\ & \|x\|_0 \le 2, \end{array} \qquad \begin{array}{ll} \underset{x,y\in\mathbb{R}^3}{\text{minimize}} & x_1 \\ \text{subject to} & (1-x_1)^3 + x_3^2 \le 0, \\ & y_1 + y_2 + y_3 \ge 1, \\ & x_i y_i = 0, \ i = 1,2,3, \\ & 0 \le y_i \le 1, \ i = 1,2,3. \end{array}$$

We claim that $x^* = (1,0,0)$ is a global solution of MPCaC and (x^*, y^*), with $y^* = (0,1,0)$, is a global solution of the relaxed problem. Indeed, denoting the feasible sets of MPCaC and the relaxed problem by Ω_0 and Ω, respectively, given any feasible point $x \in \Omega_0$, we have

$$(x_1 - 1)^3 \ge x_3^2 \ge 0,$$

which means that $x_1 \ge 1$. This proves the first part of the claim. Now, note that $(x^*, y^*) \in \Omega$. Moreover, given $(x,y) \in \Omega$, we have that $x \in \Omega_0$ and hence $x_1 \ge 1$, proving the second statement of the claim. For the points x^* and (x^*, y^*) we have

$$I_0 = \{2,3\}, \ \ I_{01} = \{2\}, \ \ I_{\pm 0} = \{1\}, \ \ I_{00} = \{3\} \quad \text{and} \quad I_{0+} = \emptyset.$$

So, there are two choices for I that satisfy (2.4): $I' = \{2\}$ or $I'' = \{2,3\}$. In view of the Proposition 2.2.8, it is enough to prove that (x^*, y^*) is not W_I-stationary for $I = I''$. Indeed, we have

$$\nabla_x L(x^*, \lambda^g) + \sum_{i\in I} \lambda_i^G e_i = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \lambda_2^G \\ \lambda_3^G \end{pmatrix},$$

which clearly does not vanish.

In view of Example 3.1 and motivated to find a necessary optimality condition for MPCaC problems, we propose in the next section the concept of approximately weak stationarity, which will be satisfied at every minimizer, independently of any constraint qualification.

3.1 Sequential optimality conditions for MPCaC

In order to define our sequential optimality condition, consider the function

$$\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

defined by

$$\begin{aligned}\mathcal{L}(x, y, \lambda^g, \lambda^h, \lambda^\theta, \lambda^G, \lambda^H, \lambda^{\tilde{H}}) = {} & L(x, \lambda^g, \lambda^h) + \lambda^\theta \theta(y) + (\lambda^G)^T G(x) \\ & + (\lambda^H)^T H(y) + (\lambda^{\tilde{H}})^T \tilde{H}(y),\end{aligned}$$

where L is the Lagrangian defined in (1.4).

Note that $\mathcal{L}$ resembles the Lagrangian $\mathcal{L}_I$, associated with $\text{TNLP}_I(\bar{x}, \bar{y})$. The only difference is that the term $(\lambda_I^G)^T G_I(x)$ was replaced by $(\lambda^G)^T G(x)$. Here it will be convenient to see that

$$\nabla_{x,y}\mathcal{L}(x, y, \lambda) = \begin{pmatrix} \nabla_x L(x, \lambda^g, \lambda^h) + \displaystyle\sum_{i=1}^n \lambda_i^G \nabla G_i(x) \\ \lambda^\theta \nabla\theta(y) + \displaystyle\sum_{i=1}^n \lambda_i^H \nabla H_i(y) + \sum_{i=1}^n \lambda_i^{\tilde{H}} \nabla \tilde{H}_i(y) \end{pmatrix}. \tag{3.1}$$

Definition 3.1.1 *Let $(\bar{x}, \bar{y})$ be a feasible point of the relaxed problem (2.3). We say that $(\bar{x}, \bar{y})$ is Approximately Weakly stationary (AW-stationary) for this problem if there exist sequences $(x^k, y^k) \subset \mathbb{R}^n \times \mathbb{R}^n$ and*

$$(\lambda^k) = \left(\lambda^{g,k}, \lambda^{h,k}, \lambda^{\theta,k}, \lambda^{G,k}, \lambda^{H,k}, \lambda^{\tilde{H},k}\right) \subset \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n_+$$

such that

1. $(x^k, y^k) \to (\bar{x}, \bar{y})$;
2. $\nabla_{x,y}\mathcal{L}(x^k, y^k, \lambda^k) \to 0$;
3. $\min\{-g(x^k), \lambda^{g,k}\} \to 0$;
4. $\min\{-\theta(y^k), \lambda^{\theta,k}\} \to 0$;
5. $\min\{|G_i(x^k)|, |\lambda_i^{G,k}|\} \to 0$ *for all* $i = 1, \ldots, n$;
6. $\min\{-H_i(y^k), |\lambda_i^{H,k}|\} \to 0$ *for all* $i = 1, \ldots, n$;
7. $\min\{-\tilde{H}(y^k), \lambda^{\tilde{H},k}\} \to 0$.

Remark 3.1 *Definition 3.1.1 resembles AKKT condition for the relaxed problem (3), where 3, 4 and 7 represent the approximate complementarity conditions for the inequality constraints $g(x) \leq 0$, $\theta(y) \leq 0$ and $\tilde{H}(y) \leq 0$, respectively and 6 is related to the last complementarity condition in W_I-stationarity. As a matter of fact, AW-stationarity is equivalent to AKKT for $TNLP_{I_0}$, as we shall see ahead in Theorem 3.2.4.*

Let us review Example 3.1 in the light of the above definition. We have seen that the minimizer is not W_I-stationary, but now we can see that it is AW-stationary.

Example 3.1.2 *Consider the problem given in Example 3.1. We claim that the global solution of the relaxed problem, (x^*, y^*), is AW-stationary. Indeed, consider the sequences $(x^k, y^k) \subset \mathbb{R}^3 \times \mathbb{R}^3$ and*

$$(\lambda^k) = \left(\lambda^{g,k}, \lambda^{\theta,k}, \lambda^{G,k}, \lambda^{H,k}, \lambda^{\tilde{H},k}\right) \subset \mathbb{R}^3_+ \times \mathbb{R}_+ \times \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3_+$$

given by $x^k = (1 + 1/k, 0, 0)$, $y^k = (0, 1, 0)$, $\lambda^{g,k} = k^2/3$, $\lambda^{\theta,k} = 0$ and $\lambda^{G,k} = \lambda^{H,k} = \lambda^{\tilde{H},k} = 0$. Then, we have $(x^k, y^k) \to (x^*, y^*)$ and

$$\nabla_x L(x^k, \lambda^{g,k}) + \sum_{i=1}^{n} \lambda_i^{G,k} \nabla G_i(x^k) = \begin{pmatrix} 1 - 3\lambda^{g,k}(1 - x_1^k)^2 \\ 0 \\ 2\lambda^{g,k} x_3^k \end{pmatrix} = 0.$$

Therefore, in view of (3.1), we obtain the first two items of Definition 3.1.1. Now, note that $g(x^k) \to g(x^*) = 0$ and $\theta(y^k) \to \theta(y^*) = 0$, giving $\min\{-g(x^k), \lambda^{g,k}\} \to 0$ and $\min\{-\theta(y^k), \lambda^{\theta,k}\} \to 0$, yielding items (3) and (4). The relation $\min\{|G_i(x^k)|, |\lambda_i^{G,k}|\} \to 0$ is immediate. Besides, since $\tilde{H}(y^k) \to \tilde{H}(y^*) \leq 0$, $\lambda^{\tilde{H},k} = 0$, $H(y^k) \to H(y^*) \leq 0$ and $\lambda^{H,k} = 0$, we have $\min\{-\tilde{H}(y^k), \lambda^{\tilde{H},k}\} \to 0$ and $\min\{-H_i(y^k), |\lambda_i^{H,k}|\} \to 0$, obtaining items 5, 6 and 7.

Now we shall prove that the above example reflects a general result, that is, every minimizer of an MPCaC problem is AW-stationary. We start the theoretical analysis with two simple facts. The first one says that the expression $\sum_{i=1}^{n} \lambda_i^{G,k} \nabla G_i(x^k)$ could be replaced by $\sum_{i \in I_0} \lambda_i^{G,k} \nabla G_i(x^k)$. The second fact states that AW-stationarity is weaker than W_I-stationarity, and consequently weaker than KKT, in view of Theorem 2.2.5.

Lemma 3.1.3 *Let $(\bar{x}, \bar{y})$ be an AW-stationary point for the relaxed problem (2.3), with corresponding sequences (x^k, y^k) and (λ^k). Then,*

$$\nabla_x L(x^k, \lambda^{g,k}, \lambda^{h,k}) + \sum_{i \in I_0} \lambda_i^{G,k} \nabla G_i(x^k) \to 0.$$

Proof. In view of (3.1), we have, in particular,

$$\nabla_x L(x^k, \lambda^{g,k}, \lambda^{h,k}) + \sum_{i=1}^{n} \lambda_i^{G,k} \nabla G_i(x^k) \to 0. \tag{3.2}$$

For $i \notin I_0$, we have

$$\lim_{k \to \infty} G_i(x^k) = G_i(\bar{x}) = \bar{x}_i \neq 0.$$

Therefore, we can assume without loss of generality that there exists $\epsilon > 0$ such that $|G_i(x^k)| \geq \epsilon$ for all k. Since $\min\{|G_i(x^k)|, |\lambda_i^{G,k}|\} \to 0$, we obtain $|\lambda_i^{G,k}| \to 0$ and hence,

$$\sum_{i \notin I_0} \lambda_i^{G,k} \nabla G_i(x^k) \to 0.$$

By subtracting this from (3.2), we conclude the proof. □

According to Lemma 3.1.3 we could define our sequential stationarity concept in terms of I_0 and call it AW$_{I_0}$-stationarity. However, since this index set does not appear in Definition 3.1.1, we maintain the more general notation AW-stationarity.

Lemma 3.1.4 *Let $(\bar{x}, \bar{y})$ be a W_I-stationary point for the relaxed problem (2.3), in the sense of Definition 2.2.3. Then $(\bar{x}, \bar{y})$ is AW-stationary for this problem.*

Proof. Consider a vector

$$\lambda = (\lambda^g, \lambda^h, \lambda^\theta, \lambda_I^G, \lambda^H, \lambda^{\tilde{H}}) \in \mathbb{R}_+^m \times \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}^{|I|} \times \mathbb{R}^n \times \mathbb{R}_+^n$$

satisfying Definition 2.2.3. Then, the (constant) sequences $(x^k, y^k) \subset \mathbb{R}^n \times \mathbb{R}^n$ and

$$(\lambda^k) = \left(\lambda^{g,k}, \lambda^{h,k}, \lambda^{\theta,k}, \lambda^{G,k}, \lambda^{H,k}, \lambda^{\tilde{H},k}\right) \subset \mathbb{R}_+^m \times \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+^n,$$

defined by

$$(x^k, y^k) = (\bar{x}, \bar{y})\,,\ \left(\lambda^{g,k}, \lambda^{h,k}, \lambda^{\theta,k}, \lambda_I^{G,k}, \lambda^{H,k}, \lambda^{\tilde{H},k}\right) = (\lambda^g, \lambda^h, \lambda^\theta, \lambda_I^G, \lambda^H, \lambda^{\tilde{H}})$$

and $\lambda_i^{G,k} = 0$ for $i \notin I$ and $k \in \mathbb{N}$, satisfy Definition 3.1.1. Indeed, item (1) is immediate and item (2) follows from the fact that

$$\nabla_{x,y}\mathcal{L}(x^k, y^k, \lambda^k) = \nabla_{x,y}\mathcal{L}_I(\bar{x}, \bar{y}, \lambda) = 0.$$

By the feasibility of $(\bar{x}, \bar{y})$, the nonnegativity of λ^g, λ^θ and $\lambda^{\tilde{H}}$ and the complementarity conditions associated with these multipliers, we have

$$\begin{aligned} \min\{-g(x^k), \lambda^{g,k}\} &= \min\{-g(\bar{x}), \lambda^g\} = 0, \\ \min\{-\theta(y^k), \lambda^{\theta,k}\} &= \min\{-\theta(\bar{y}), \lambda^\theta\} = 0, \\ \min\{-\tilde{H}(y^k), \lambda^{\tilde{H},k}\} &= \min\{-\tilde{H}(\bar{y}), \lambda^{\tilde{H}}\} = 0, \end{aligned}$$

proving (3), (4) and (7). Now, $G_i(x^k) = G_i(\bar{x}) = 0$ for $i \in I$ and $\lambda_i^{G,k} = 0$ for $i \notin I$. In any case, it holds

$$\min\{|G_i(x^k)|, |\lambda_i^{G,k}|\} = \min\{|G_i(\bar{x})|, |\lambda_i^{G,k}|\} = 0,$$

which proves (5). Finally, note that $\lambda_i^{H,k} = \lambda_i^H = 0$ for $i \in I_{0+}(\bar{x}, \bar{y}) \cup I_{01}(\bar{x}, \bar{y})$ and $H_i(y^k) = H_i(\bar{y}) = 0$ for $i \in I_{00}(\bar{x}, \bar{y}) \cup I_{\pm 0}(\bar{x}, \bar{y})$. So,

$$\min\{-H_i(y^k), |\lambda_i^{H,k}|\} = \min\{-H_i(\bar{y}), |\lambda_i^H|\} = 0$$

for all $i = 1, \ldots, n$, proving (6) and completing the proof. □

Remark 3.2 *We point out here that, in contrast to W_I-stationarity, which conveniently depends on the set I, our sequential optimality condition is independent of any set I. This is a desirable feature since AW-stationarity has a certain amount of algorithmic appeal. In practice, one should be able to use such a condition as stopping criterion for an algorithm designed to solve the MPCaC problem.*

Before proving our main sequential optimality results, let us see some preliminary lemmas. To this purpose, consider the augmented problem

$$\begin{array}{ll} \underset{x,y,w}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \leq 0, h(x) = 0, \\ & \theta(y) \leq 0, \\ & w^G - G(x) = 0,\ w^H + H(y) = 0, \\ & \tilde{H}(y) \leq 0, \\ & w \in W, \end{array} \tag{3.3}$$

where $W = \{w = \big(w^G, w^H\big) \in \mathbb{R}^n \times \mathbb{R}^n_+ \mid w^G * w^H = 0\}$.

This problem will be crucial in the analysis. In the next two lemmas we establish the equivalence between the relaxed problem (2.3) and this augmented problem. Moreover, there is a suitable reason to write the constraints $H(y) \leq 0$ and $G(x) * H(y) = 0$ of (2.3) in the format $w \in W$. Such a strategy will enable us to apply Lemma 1.1.5 to obtain Guignard constraint qualification for an auxiliary problem ahead.

Lemma 3.1.5 *Let (x^*, y^*) be a local (global) minimizer of the relaxed problem (2.3). Given $w^* \in \mathbb{R}^n \times \mathbb{R}^n_+$, if the point (x^*, y^*, w^*) is feasible for the augmented problem (3.3), then it is a local (global) minimizer of this problem. In particular, this holds for $w^* = \big(G(x^*), -H(y^*)\big)$.*

Proof. First, let us prove the relation between local minimizers. In view of the equivalence of norms, we consider $\|\cdot\|_\infty$, for convenience[1]. By hypothesis, there exists $\delta > 0$ such that if (x, y) is feasible for (2.3) and $\|(x, y) - (x^*, y^*)\|_\infty \leq \delta$, then $f(x^*) \leq f(x)$. Suppose that (x^*, y^*, w^*) is feasible for the problem (3.3) and consider an arbitrary feasible point (x, y, w) for this problem such that $\|(x, y, w) - (x^*, y^*, w^*)\|_\infty \leq \delta$. Then, the pair (x, y) is feasible for (2.3) and $\|(x, y) - (x^*, y^*)\|_\infty \leq \delta$. Hence, $f(x^*) \leq f(x)$ and, therefore, (x^*, y^*, w^*) is a local minimizer of (3.3). Note that (x^*, y^*, w^*), with $w^* = \big(G(x^*), -H(y^*)\big)$, is trivially feasible. Finally, if we ignore the neighborhoods in the argument above, we obtain the relation between global minimizers. □

For the sake of completeness we establish in the next result the converse of Lemma 3.1.5.

Lemma 3.1.6 *Let (x^*, y^*, w^*) be a local (global) minimizer of (3.3). Then (x^*, y^*) is a local (global) minimizer of (2.3).*

Proof. By the feasibility of (x^*, y^*, w^*) we have that (x^*, y^*) is feasible for (2.3),

$$(w^*)^G = G(x^*) \quad \text{and} \quad (w^*)^H = -H(y^*). \tag{3.4}$$

Consider $\delta_1 > 0$ such that $f(x^*) \leq f(x)$ for all feasible point (x, y, w) of (3.3), satisfying $\|(x, y, w) - (x^*, y^*, w^*)\|_\infty \leq \delta_1$. Let $\delta_2 > 0$ be such that

$$\|G(x) - G(x^*)\|_\infty \leq \delta_1 \quad \text{and} \quad \|H(y) - H(y^*)\|_\infty \leq \delta_1 \tag{3.5}$$

for all $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ with $\|(x, y) - (x^*, y^*)\|_\infty \leq \delta_2$. Define $\delta = \min\{\delta_1, \delta_2\}$ and take (x, y), feasible for (2.3), such that $\|(x, y) - (x^*, y^*)\|_\infty \leq \delta$. Thus we have (3.5), which in

[1] Suppose that there exists $\delta > 0$ such that $\zeta(z^*) \leq \zeta(z)$ if $\|z - z^*\|_\infty \leq \delta$. Given an arbitrary norm $\|\cdot\|$, by the equivalence of norms, there exists a constant $c > 0$ such that $\|z\|_\infty \leq c\|z\|$ for all z. Thus, defining $\delta' = \delta/c$, if $\|z - z^*\| \leq \delta'$, then $\|z - z^*\|_\infty \leq c\|z - z^*\| \leq c\delta' = \delta$. Hence, $\zeta(z^*) \leq \zeta(z)$. This means that the concept of local minimizer does not depend on the norm considered.

view of (3.4) can be rewritten as $\|w - w^*\|_\infty \leq \delta_1$, with $w = \big(G(x), -H(y)\big)$. Therefore, (x, y, w) is feasible for (3.3) and

$$\|(x, y, w) - (x^*, y^*, w^*)\|_\infty \leq \delta_1,$$

implying that $f(x^*) \leq f(x)$.

Now, let us see the global optimality. So, assume that (x^*, y^*, w^*) is a global minimizer of (3.3). Then (x^*, y^*) is feasible for (2.3). Furthermore, given an arbitrary feasible point (x, y), we have that (x, y, w), with $w = \big(G(x), -H(y)\big)$, is feasible for (3.3). Therefore, $f(x^*) \leq f(x)$. □

Lemma 3.1.7 *Suppose that (x^*, y^*) is a local minimizer of the relaxed problem (2.3). Then, given an arbitrary norm $\|\cdot\|$, there exists $\delta > 0$ such that (x^*, y^*, w^*), with $w^* = \big(G(x^*), -H(y^*)\big)$, is the unique global minimizer of the problem*

$$\begin{array}{ll} \underset{x,y,w}{\text{minimize}} & f(x) + \dfrac{1}{2}\|(x, y) - (x^*, y^*)\|_2^2 \\ \text{subject to} & g(x) \leq 0, h(x) = 0, \\ & \theta(y) \leq 0, \\ & w^G - G(x) = 0,\ w^H + H(y) = 0, \\ & \tilde{H}(y) \leq 0, \\ & w \in W, \\ & \|(x, y, w) - (x^*, y^*, w^*)\| \leq \delta. \end{array} \tag{3.6}$$

Proof. By Lemma 3.1.5, we have that (x^*, y^*, w^*) is a local minimizer of (3.3). Consider $\delta > 0$ such that if (x, y, w) is feasible for (3.3) and

$$\|(x, y, w) - (x^*, y^*, w^*)\| \leq \delta, \tag{3.7}$$

then $f(x^*) \leq f(x)$. Note that (x^*, y^*, w^*) is feasible for (3.6). Moreover, given any feasible point (x, y, w), of (3.6), we have that it is also feasible for (3.3) and satisfies (3.7). Hence,

$$f(x^*) + \frac{1}{2}\|(x^*, y^*) - (x^*, y^*)\|_2^2 = f(x^*) \leq f(x) \leq f(x) + \frac{1}{2}\|(x, y) - (x^*, y^*)\|_2^2,$$

proving that (x^*, y^*, w^*) is a global minimizer of (3.6).

Now, suppose that $(\bar{x}, \bar{y}, \bar{w})$ is also a global minimizer of (3.6). Then,

$$f(\bar{x}) + \frac{1}{2}\|(\bar{x}, \bar{y}) - (x^*, y^*)\|_2^2 \leq f(x^*) + \frac{1}{2}\|(x^*, y^*) - (x^*, y^*)\|_2^2 = f(x^*) \leq f(\bar{x}),$$

where the last inequality follows from the fact that $(\bar{x}, \bar{y}, \bar{w})$ is feasible for (3.3) and satisfies (3.7). Therefore, $(\bar{x}, \bar{y}) = (x^*, y^*)$, and hence

$$\bar{w} = \big(G(\bar{x}), -H(\bar{y})\big) = \big(G(x^*), -H(y^*)\big) = w^*,$$

proving the uniqueness. □

The next result shows that our stationarity concept, given in Definition 3.1.1, is a legitimate optimality condition, independently of any constraint qualification. This is a requirement for such a condition to be useful in the analysis of algorithms.

Theorem 3.1.8 *If (x^*, y^*) is a local minimizer of the relaxed problem (2.3), then it is an AW-stationary point, in the sense of Definition 3.1.1.*

Proof. Defining $w^* = \big(G(x^*), -H(y^*)\big)$, we conclude from Lemma 3.1.7 that there exists $\delta > 0$ such that the point (x^*, y^*, w^*) is the unique global minimizer of the problem (3.6), with $\|\cdot\|_2$ in the last constraint. Define the (partial) infeasibility measure associated with this problem as

$$\begin{aligned}\varphi(x,y,w) &= \frac{1}{2}\Big(\|g^+(x)\|_2^2 + \|h(x)\|_2^2 + \|\theta^+(y)\|_2^2 + \|w^G - G(x)\|_2^2 \\ &\quad + \|w^H + H(y)\|_2^2 + \|\tilde{H}^+(y)\|_2^2\Big),\end{aligned}$$

consider a sequence $\rho_k \to \infty$ and let (x^k, y^k, w^k) be a global minimizer of the penalized problem

$$\begin{array}{ll}\underset{x,y,w}{\text{minimize}} & f(x) + \frac{1}{2}\|(x,y) - (x^*,y^*)\|_2^2 + \rho_k\varphi(x,y,w) \\ \text{subject to} & w \in W, \\ & \|(x,y,w) - (x^*,y^*,w^*)\|_2^2 \le \delta^2,\end{array} \tag{3.8}$$

which is well defined because the objective function is continuous and the feasible set is compact. Since $\|(x^k, y^k, w^k) - (x^*, y^*, w^*)\|_2 \le \delta$, we can assume without loss of generality that the sequence (x^k, y^k, w^k) converges to some point $(\bar{x}, \bar{y}, \bar{w})$. We claim that $(\bar{x}, \bar{y}, \bar{w}) = (x^*, y^*, w^*)$. Note first that (x^*, y^*, w^*) is feasible for (3.8) and $\varphi(x^*, y^*, w^*) = 0$. So, by the optimality of (x^k, y^k, w^k) we have

$$f(x^k) + \frac{1}{2}\|(x^k, y^k) - (x^*, y^*)\|_2^2 + \rho_k\varphi(x^k, y^k, w^k) \le f(x^*), \tag{3.9}$$

implying that $\varphi(x^k, y^k, w^k) \to 0$, because $\rho_k \to \infty$. This in turn implies that $\varphi(\bar{x}, \bar{y}, \bar{w}) = 0$, giving $g^+(\bar{x}) = 0$, $h(\bar{x}) = 0$, $\theta^+(\bar{y}) = 0$, $\bar{w}^G = G(\bar{x})$, $\bar{w}^H = -H(\bar{y})$ and $\tilde{H}^+(\bar{y}) = 0$. Moreover, as the sequence (x^k, y^k, w^k) is feasible for (3.8), its limit point $(\bar{x}, \bar{y}, \bar{w})$ satisfies $\bar{w} \in W$, because W is a closed set, and $\|(\bar{x}, \bar{y}, \bar{w}) - (x^*, y^*, w^*)\| \le \delta$. Therefore, $(\bar{x}, \bar{y}, \bar{w})$ is feasible for (3.6). Furthermore, from (3.9) we obtain

$$f(x^k) + \frac{1}{2}\|(x^k, y^k) - (x^*, y^*)\|_2^2 \le f(x^*).$$

Taking the limit, it follows that

$$f(\bar{x}) + \frac{1}{2}\|(\bar{x}, \bar{y}) - (x^*, y^*)\|_2^2 \le f(x^*),$$

which means that $(\bar{x}, \bar{y}, \bar{w})$ is optimal for (3.6). By the uniqueness of the optimal solution of this problem, we conclude that $(\bar{x}, \bar{y}, \bar{w}) = (x^*, y^*, w^*)$, proving the claim. As consequence, we have the first item of Definition 3.1.1.

In order to prove the next item, let us see first that a constraint qualification holds at the minimizer (x^k, y^k, w^k). Since $(x^k, y^k, w^k) \to (x^*, y^*, w^*)$, we may assume without loss of generality that $\|(x^k, y^k, w^k) - (x^*, y^*, w^*)\|_2 < \delta$ for all k. That is, the inequality constraint in the problem (3.8) is inactive at the minimizer. By Lemma 1.1.2, the tangent and linearized cones at this point are the ones taking into account only the constraints in $w \in W$, namely,

$$-w^H \le 0 \quad \text{and} \quad w^G * w^H = 0. \tag{3.10}$$

So, in view of Lemmas 1.1.3 and 1.1.5, Guignard constraint qualification holds at (x^k, y^k, w^k). This implies that this point satisfies the KKT conditions, which means that there exist multipliers $\mu^{H,k} \in \mathbb{R}^n_+$ and $\mu^{0,k} \in \mathbb{R}^n$, associated with the constraints in $w \in W$, such that

$$\nabla f(x^k) + (x^k - x^*) + \rho_k \nabla_x \varphi(x^k, y^k, w^k) = 0 \tag{3.11a}$$

$$(y^k - y^*) + \rho_k \nabla_y \varphi(x^k, y^k, w^k) = 0 \tag{3.11b}$$

$$\rho_k \nabla_{w^G} \varphi(x^k, y^k, w^k) + \mu^{0,k} * w^{H,k} = 0 \tag{3.11c}$$

$$\rho_k \nabla_{w^H} \varphi(x^k, y^k, w^k) - \mu^{H,k} + \mu^{0,k} * w^{G,k} = 0 \tag{3.11d}$$

$$\mu^{H,k} * w^{H,k} = 0. \tag{3.11e}$$

Noting that the partial gradients of φ are given by

$$\nabla_x \varphi(x, y, w) = \nabla g(x) g^+(x) + \nabla h(x) h(x) + \nabla G(x)\big(G(x) - w^G\big), \tag{3.12a}$$

$$\nabla_y \varphi(x, y, w) = \theta^+(y) \nabla \theta(y) + \nabla \tilde{H}(y) \tilde{H}^+(y) + \nabla H(y)\big(w^H + H(y)\big), \tag{3.12b}$$

$$\nabla_{w^G} \varphi(x, y, w) = w^G - G(x) \quad \text{and} \quad \nabla_{w^H} \varphi(x, y, w) = w^H + H(y) \tag{3.12c}$$

and defining the components of λ^k as

$$\lambda^{g,k} = \rho_k g^+(x^k), \;\; \lambda^{h,k} = \rho_k h(x^k), \;\; \lambda^{\theta,k} = \rho_k \theta^+(y^k),$$

$$\lambda^{G,k} = \rho_k\big(G(x^k) - w^{G,k}\big), \;\; \lambda^{H,k} = \rho_k\big(w^{H,k} + H(y^k)\big), \;\; \lambda^{\tilde{H},k} = \rho_k \tilde{H}^+(y^k),$$

we see immediately that $\lambda^{g,k} \geq 0$, $\lambda^{\theta,k} \geq 0$ and $\lambda^{\tilde{H},k} \geq 0$. Moreover, using (3.11a) and (3.12a), we obtain

$$\nabla_x \mathcal{L}(x^k, y^k, \lambda^k) = \nabla f(x^k) + \rho_k \nabla_x \varphi(x^k, y^k, w^k) = x^* - x^k \to 0.$$

Furthermore, from (3.11b) and (3.12b), we have

$$\nabla_y \mathcal{L}(x^k, y^k, \lambda^k) = \rho_k \nabla_y \varphi(x^k, y^k, w^k) = y^* - y^k \to 0,$$

proving item 2.

Let us prove item 3. By the feasibility of (x^*, y^*) we have $g_i(x^*) \leq 0$ for all $i = 1, \ldots, m$. If $g_i(x^*) = 0$, then $\min\{-g_i(x^k), \lambda_i^{g,k}\} \to 0$ since $g_i(x^k) \to 0$ and $\lambda_i^{g,k} \geq 0$. On the other hand, if $g_i(x^*) < 0$, we may assume that $g_i(x^k) < 0$ for all k. Thus, $g_i^+(x^k) = 0$, yielding $\lambda_i^{g,k} = \rho_k g_i^+(x^k) = 0$. Therefore, $\min\{-g_i(x^k), \lambda_i^{g,k}\} = 0$. Items 4 and 7 can be proved by the same reasoning.

Now, note that by (3.11c) ,(3.11d) and (3.12c) we have

$$\lambda^{G,k} = \mu^{0,k} * w^{H,k} \quad \text{and} \quad \lambda^{H,k} = \mu^{H,k} - \mu^{0,k} * w^{G,k}. \tag{3.13}$$

Therefore, using the fact that $w^k \in W$, we obtain

$$\lambda_i^{G,k} w_i^{G,k} = \mu_i^{0,k} w_i^{H,k} w_i^{G,k} = 0$$

for all $i = 1, \ldots, m$. Furthermore, given $i \notin I_0(x^*)$, we have

$$w_i^{G,k} \to (w_i^*)^G = G_i(x^*) = x_i^* \neq 0,$$

implying that $\lambda_i^{G,k} = 0$ for all k large enough. So, $\min\{|G_i(x^k)|, |\lambda_i^{G,k}|\} = 0$. On the other hand, if $i \in I_0(x^*)$, we have $G_i(x^k) \to G_i(x^*) = x_i^* = 0$, and hence, $\min\{|G_i(x^k)|, |\lambda_i^{G,k}|\} \to 0$, proving item 5.

To prove the next item, note that using (3.13), (3.11e) and the fact that $w^k \in W$,

$$\lambda^{H,k} * w^{H,k} = \mu^{H,k} * w^{H,k} - \mu^{0,k} * w^{G,k} * w^{H,k} = 0. \tag{3.14}$$

By the feasibility of (x^*, y^*), we have $H(y^*) \leq 0$. In the case $H_i(y^*) < 0$,

$$w_i^{H,k} \to (w_i^*)^H = -H_i(y^*) > 0,$$

giving $\lambda_i^{H,k} = 0$ for all k large enough. Thus, $\min\{-H_i(y^k), |\lambda_i^{H,k}|\} = 0$. On the other hand, if $H_i(y^*) = 0$, we have $H_i(y^k) \to H_i(y^*) = 0$, and hence, $\min\{-H_i(y^k), |\lambda_i^{H,k}|\} \to 0$, proving item 6 and completing the proof. □

3.2 Relations to other sequential optimality conditions

In this section we discuss the relationships between approximate stationarity for standard nonlinear optimization and AW-stationarity.

As it is well known, every minimizer of an optimization problem is AKKT (see Definition 1.2.1). However, and surprisingly, we start by proving that the AKKT condition fails to detect good candidates for optimality for every MPCaC problem.

Theorem 3.2.1 *Every feasible point $(\bar{x}, \bar{y})$ for the relaxed problem (2.3) is AKKT.*

Proof. We need to prove that there exist sequences $(x^k, y^k) \subset \mathbb{R}^n \times \mathbb{R}^n$ and

$$\left(\mu^{g,k}, \mu^{h,k}, \mu^{\theta,k}, \mu^{H,k}, \mu^{\tilde{H},k}, \mu^{\xi,k}\right) \subset \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}^n_+ \times \mathbb{R}^n_+ \times \mathbb{R}^n$$

such that $(x^k, y^k) \to (\bar{x}, \bar{y})$ and

$$\begin{aligned} &\begin{pmatrix} \nabla_x L(x^k, \mu^{g,k}, \mu^{h,k}) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \mu^{\theta,k} \nabla \theta(y^k) \end{pmatrix} + \sum_{i=1}^n \begin{pmatrix} 0 \\ \mu_i^{H,k} \nabla H_i(y^k) \end{pmatrix} \\ &\quad + \sum_{i=1}^n \begin{pmatrix} 0 \\ \mu_i^{\tilde{H},k} \nabla \tilde{H}_i(y^k) \end{pmatrix} + \sum_{i=1}^n \mu_i^{\xi,k} \begin{pmatrix} H_i(y^k) \nabla G_i(x^k) \\ G_i(x^k) \nabla H_i(y^k) \end{pmatrix} \to 0, \end{aligned} \tag{3.15a}$$

$$\min\{-g(x^k), \mu^{g,k}\} \to 0\,, \quad \min\{-\theta(y^k), \mu^{\theta,k}\} \to 0, \tag{3.15b}$$

$$\min\{-H(y^k), \mu^{H,k}\} \to 0\,, \quad \min\{-\tilde{H}(y^k), \mu^{\tilde{H},k}\} \to 0. \tag{3.15c}$$

Let $b = \nabla f(\bar{x})$ and define $x^k = \bar{x}$, $\mu^{g,k} = 0$, $\mu^{h,k} = 0$, $\mu^{\theta,k} = 0$, $\mu^{\tilde{H},k} = 0$ and

$$y_i^k = \bar{y}_i\,,\ \mu_i^{H,k} = 0\,,\ \mu_i^{\xi,k} = \frac{b_i}{y_i^k} \text{ for } i \in I_{0+}(\bar{x}, \bar{y}) \cup I_{01}(\bar{x}, \bar{y}),$$

$$y_i^k = \frac{b_i}{k}\,,\ \mu_i^{H,k} = 0\,,\ \mu_i^{\xi,k} = k \text{ for } i \in I_{00}(\bar{x}, \bar{y}),$$

$$y_i^k = -\frac{\operatorname{sign}(\bar{x}_i) b_i}{k}\,,\ \mu_i^{\xi,k} = -\operatorname{sign}(\bar{x}_i) k\,,\ \mu_i^{H,k} = -\mu_i^{\xi,k} x_i^k \text{ for } i \in I_{\pm 0}(\bar{x}, \bar{y}).$$

Thus we have $\mu_i^{H,k} \geq 0$, $(x^k, y^k) \to (\bar{x}, \bar{y})$,

$$\nabla_{x_i} L(x^k, \mu^{g,k}, \mu^{h,k}) - \mu_i^{\xi,k} y_i^k = b_i - \mu_i^{\xi,k} y_i^k \to 0,$$

and

$$-\mu^{\theta,k} - \mu_i^{H,k} + \mu_i^{\tilde{H},k} - \mu_i^{\xi,k} x_i^k \to 0$$

for all $i = 1, \ldots, n$, giving (3.15a). Moreover, it is easy to see that (3.15b) and (3.15c) also hold. □

Another sequential optimality condition for standard NLP is PAKKT (Definition 1.2.3). It is stronger than AKKT, but not stronger than AW-stationarity. The next example shows that PAKKT for the relaxed problem does not imply AW-stationarity, even under strict complementarity.

Example 3.2.2 *Consider the MPCaC and the corresponding relaxed problem given below.*

$$\begin{array}{ll} \underset{x\in\mathbb{R}^2}{\text{minimize}} & x_2 \\ \text{subject to} & x_1^2 \le 0, \\ & \|x\|_0 \le 1, \end{array} \qquad \begin{array}{ll} \underset{x,y\in\mathbb{R}^2}{\text{minimize}} & x_2 \\ \text{subject to} & x_1^2 \le 0, \\ & y_1 + y_2 \ge 1, \\ & x_i y_i = 0,\ i = 1,2, \\ & 0 \le y_i \le 1,\ i = 1,2. \end{array}$$

Given $a > 0$, we claim that the point $(\bar{x}, \bar{y})$, with $\bar{x} = (0, a)$ and $\bar{y} = (1, 0)$, is PAKKT but not AW-stationary. Indeed, for the first statement, consider the sequences $(x^k, y^k) \subset \mathbb{R}^2 \times \mathbb{R}^2$ and

$$(\gamma^k) = \left(\lambda^{g,k}, \lambda^{\theta,k}, \mu^k, \lambda^{\tilde{H},k}, \lambda^{\xi,k}\right) \subset \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^2_+ \times \mathbb{R}^2_+ \times \mathbb{R}^2$$

given by $x^k = (1/k^3, a)$, $y^k = (1, -1/k)$, $\lambda^{g,k} = k^2$, $\lambda^{\theta,k} = 0$, $\mu^k = (0, ak)$, $\lambda^{\tilde{H},k} = (0,0)$ and $\lambda^{\xi,k} = (0, k)$. Then we have $(x^k, y^k) \to (\bar{x}, \bar{y})$ and, denoting $\xi(x, y) = x * y$, the gradient of the Lagrangian of the relaxed problem reduces to

$$\begin{pmatrix} \nabla f(x^k) \\ 0 \end{pmatrix} + \lambda^{g,k} \begin{pmatrix} \nabla g(x^k) \\ 0 \end{pmatrix} + \mu_2^k \nabla H_2(y^k) + \lambda_2^{\xi,k} \nabla \xi_2(x^k, y^k)$$

$$= \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 2\lambda^{g,k} x_1^k \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ -\mu_2^k \end{pmatrix} + \begin{pmatrix} 0 \\ \lambda_2^{\xi,k} y_2^k \\ 0 \\ \lambda_2^{\xi,k} x_2^k \end{pmatrix} = \begin{pmatrix} 2/k \\ 0 \\ 0 \\ 0 \end{pmatrix} \to 0,$$

proving (1.15a). Now, note that $g(x^k) \to g(\bar{x}) = 0$ and $\theta(y^k) \to \theta(\bar{y}) = 0$, which in turn implies that

$$\min\{-g(x^k), \lambda^{g,k}\} \to 0 \quad \textit{and} \quad \min\{-\theta(y^k), \lambda^{\theta,k}\} \to 0. \tag{3.17}$$

Moreover, we have $-\tilde{H}(y^k) \to -\tilde{H}(\bar{y}) \ge 0$ and $\lambda^{\tilde{H},k} = (0, 0)$, giving

$$\min\{-\tilde{H}(y^k), \lambda^{\tilde{H},k}\} \to 0. \tag{3.18}$$

Furthermore, since $-H_1(y^k) \to -H_1(\bar{y}) \ge 0$, $\mu_1^k = 0$ and $-H_2(y^k) = y_2^k \to 0$, we have

$$\min\{-H(y^k), \mu^k\} \to 0. \tag{3.19}$$

Conditions (3.17), (3.18) and (3.19) prove the approximate complementarity (1.15b). Moreover, we have $\delta_k = \|(1, \gamma^k)\|_\infty = k^2$ for all k large enough,

$$\limsup_{k\to\infty} \frac{\lambda^{g,k}}{\delta_k} > 0 \quad \textit{and} \quad \lambda^{g,k} g(x^k) > 0.$$

For the remaining multipliers the lim sup *is zero and so we conclude that (1.15c) and (1.15d) hold, proving that Definition 1.2.3 is satisfied, that is, $(\bar{x}, \bar{y})$ is PAKKT.*

Now, let us see that $(\bar{x}, \bar{y})$ is not AW-stationary. For this purpose, assume that the sequences $(x^k, y^k) \subset \mathbb{R}^2 \times \mathbb{R}^2$ and

$$(\lambda^k) = \left(\lambda^{g,k}, \lambda^{\theta,k}, \lambda^{G,k}, \lambda^{H,k}, \lambda^{\tilde{H},k}\right) \subset \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2_+$$

are such that $(x^k, y^k) \to (\bar{x}, \bar{y})$ and $\min\{|G_2(x^k)|, |\lambda_2^{G,k}|\} \to 0$. Then, since $|G_2(x^k)| = |x_2^k| \to a > 0$, we obtain $\lambda_2^{G,k} \to 0$. Therefore, the expression

$$\nabla_x L(x^k, \lambda^{g,k}) + \sum_{i=1}^{2} \lambda_i^{G,k} \nabla G_i(x^k) = \begin{pmatrix} 2\lambda^{g,k} x_1^k + \lambda_1^{G,k} \\ 1 + \lambda_2^{G,k} \end{pmatrix}$$

cannot converge to zero. Thus, taking into account (3.1), item (2) of Definition 3.1.1 does not hold and hence $(\bar{x}, \bar{y})$ is not AW-stationary.

In contrast to AKKT and PAKKT, the other classical sequential optimality condition, CAKKT (Definition 1.2.2), does imply AW-stationarity, as we can see in the next result.

Theorem 3.2.3 *If $(\bar{x}, \bar{y})$ is a CAKKT point for the relaxed problem (2.3), then it is AW-stationary.*

Proof. In view of Definition 1.2.2, there exist sequences $(x^k, y^k) \subset \mathbb{R}^n \times \mathbb{R}^n$ and

$$\left(\lambda^{g,k}, \lambda^{h,k}, \lambda^{\theta,k}, \mu^k, \lambda^{\tilde{H},k}, \lambda^{\xi,k}\right) \subset \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}^n_+ \times \mathbb{R}^n_+ \times \mathbb{R}^n$$

such that $(x^k, y^k) \to (\bar{x}, \bar{y})$,

$$\begin{aligned}
&\begin{pmatrix} \nabla_x L(x^k, \lambda^{g,k}, \lambda^{h,k}) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \lambda^{\theta,k} \nabla \theta(y^k) \end{pmatrix} + \sum_{i=1}^n \begin{pmatrix} 0 \\ \mu_i^k \nabla H_i(y^k) \end{pmatrix} \\
&+ \sum_{i=1}^n \begin{pmatrix} 0 \\ \lambda_i^{\tilde{H},k} \nabla \tilde{H}_i(y^k) \end{pmatrix} + \sum_{i=1}^n \lambda_i^{\xi,k} \begin{pmatrix} H_i(y^k) \nabla G_i(x^k) \\ G_i(x^k) \nabla H_i(y^k) \end{pmatrix} \to 0, \qquad (3.20a)
\end{aligned}$$

$$\lambda^{g,k} * g(x^k) \to 0\,, \quad \lambda^{h,k} * h(x^k) \to 0\,, \quad \lambda^{\theta,k} \theta(y^k) \to 0\,, \qquad (3.20b)$$

$$\mu^k * H(y^k) \to 0\,, \quad \lambda^{\tilde{H},k} * \tilde{H}(y^k) \to 0, \qquad (3.20c)$$

$$\lambda^{\xi,k} * G(x^k) * H(y^k) \to 0. \qquad (3.20d)$$

So, we may define

$$\lambda^{H,k} = \mu^k + \lambda^{\xi,k} * G(x^k) \quad \text{and} \quad \lambda^{G,k} = \lambda^{\xi,k} * H(y^k)$$

to obtain item (2) of Definition 3.1.1 from (3.20a). Items (3), (4) and (7) follow from (3.20b), (3.20c) and Remark 1.3. Let us prove item (5). For $i \in I_0$, it holds $G_i(x^k) \to G_i(\bar{x}) = \bar{x}_i = 0$. Thus, $\min\{|G_i(x^k)|, |\lambda_i^{G,k}|\} \to 0$. If $i \notin I_0$, we have $G_i(x^k) \to \bar{x}_i \neq 0$, which in view of (3.20d) yields

$$\lambda_i^{G,k} = \lambda_i^{\xi,k} H_i(y^k) \to 0.$$

Therefore, $\min\{|G_i(x^k)|, |\lambda_i^{G,k}|\} \to 0$ for all $i = 1, \ldots, n$. Finally, in order to prove item (6), note that (3.20c) and (3.20d) give

$$\lambda_i^{H,k} H_i(y^k) = \mu_i^k H_i(y^k) + \lambda_i^{\xi,k} G_i(x^k) H_i(y^k) \to 0.$$

So, applying the argument of Remark 1.3 with $\alpha^k = |\lambda_i^{H,k}|$ and $\beta^k = H_i(y^k)$, we obtain

$$\min\{-H_i(y^k), |\lambda_i^{H,k}|\} \to 0$$

for all $i = 1, \ldots, n$. Therefore, $(\bar{x}, \bar{y})$ is AW-stationary for the problem (2.3). □

Remark 3.3 *Despite being stronger than AW-stationarity, we emphasize that the sequential optimality condition CAKKT is not so suitable to deal with MPCaC problems as AW-stationarity is. The goal of considering CAKKT is to obtain, under certain constraint qualifications, KKT points for standard nonlinear programming problems. However, as we have been discussed, MPCaC are very degenerate problems because of the challenging complementarity constraint $G(x) * H(y) = 0$. This means that we cannot expect to find strong stationary points for this class of problems and thereby making AW-stationarity a more suitable tool for dealing with such problems.*

We summarize the relations discussed so far in the diagram presented in Figure 3.1.

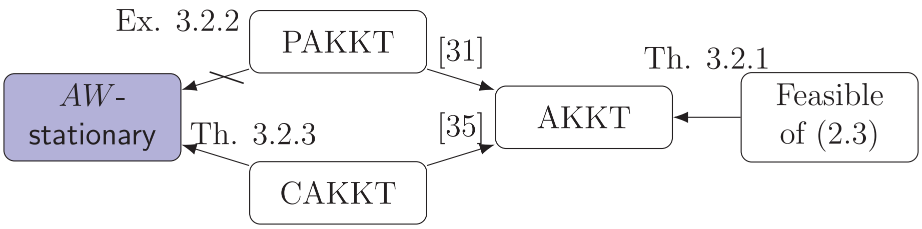

Figura 3.1: Relationships between AW-stationarity and some of the standard sequential optimality conditions. An arrow indicates a strict implication between two conditions.

To finish this section, we relate our sequential optimality condition to the tightened problem. The following result may be viewed as a sequential version of Proposition 2.2.4.

Theorem 3.2.4 *Let $(\bar{x}, \bar{y})$ be a feasible point of the relaxed problem (2.3). Then $(\bar{x}, \bar{y})$ is AW-stationary if and only if it is an AKKT point for the tightened problem $TNLP_{I_0}(\bar{x}, \bar{y})$ defined in (2.5).*

Proof. Suppose first that $(\bar{x}, \bar{y})$ is AW-stationary. Then, in view of Lemma 3.1.3, we conclude that there exist sequences $(x^k, y^k) \subset \mathbb{R}^n \times \mathbb{R}^n$ and

$$(\lambda^k) = \left(\lambda^{g,k}, \lambda^{h,k}, \lambda^{\theta,k}, \lambda^{G,k}, \lambda^{H,k}, \lambda^{\tilde{H},k}\right) \subset \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n_+$$

such that $(x^k, y^k) \to (\bar{x}, \bar{y})$,

$$\nabla_x L(x^k, \lambda^{g,k}, \lambda^{h,k}) + \sum_{i \in I_0} \lambda_i^{G,k} \nabla G_i(x^k) \to 0, \tag{3.21a}$$

$$\lambda^{\theta,k} \nabla \theta(y^k) + \sum_{i=1}^{n} \lambda_i^{H,k} \nabla H_i(y^k) + \sum_{i=1}^{n} \lambda_i^{\tilde{H},k} \nabla \tilde{H}_i(y^k) \to 0, \tag{3.21b}$$

$$\min\{-g(x^k), \lambda^{g,k}\} \to 0\,, \quad \min\{-\theta(y^k), \lambda^{\theta,k}\} \to 0, \tag{3.21c}$$

$$\min\{-H_i(y^k), |\lambda_i^{H,k}|\} \to 0\,,\ i = 1, \ldots, n, \quad \min\{-\tilde{H}(y^k), \lambda^{\tilde{H},k}\} \to 0. \tag{3.21d}$$

For $i \in I_{0+} \cup I_{01}$ we have

$$H_i(y^k) \to H_i(\bar{y}) = -\bar{y}_i < 0.$$

Therefore, we can assume without loss of generality that there exists $\epsilon > 0$ such that $-H_i(y^k) \geq \epsilon$ for all k. So, using (3.21d), we obtain $|\lambda_i^{H,k}| \to 0$, which in turn implies that

$$\sum_{i \in I_{0+} \cup I_{01}} \lambda_i^{H,k} \nabla H_i(y^k) \to 0.$$

By subtracting this from (3.21b), we obtain

$$\lambda^{\theta,k} \nabla \theta(y^k) + \sum_{i \in I_{00} \cup I_{\pm 0}} \lambda_i^{H,k} \nabla H_i(y^k) + \sum_{i=1}^{n} \lambda_i^{\tilde{H},k} \nabla \tilde{H}_i(y^k) \to 0.$$

So, we can redefine $\lambda_i^{H,k}$, $i \in I_{0+} \cup I_{01}$, to be zero, without affecting (3.21b). Therefore, taking into account (3.21a), (3.21c), the second part of (3.21d) and the fact that $\min\{-H_i(y^k), \lambda_i^{H,k}\} = 0$ for $i \in I_{0+} \cup I_{01}$, we conclude that $(\bar{x}, \bar{y})$ is AKKT for $\text{TNLP}_{I_0}(\bar{x}, \bar{y})$, which we recall here for convenience,

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \leq 0, h(x) = 0, \\ & \theta(y) \leq 0, \\ & G_i(x) = 0,\ i \in I_0, \\ & H_i(y) \leq 0,\ i \in I_{0+} \cup I_{01}, \\ & H_i(y) = 0,\ i \in I_{00} \cup I_{\pm 0}, \\ & \tilde{H}(y) \leq 0. \end{array}$$

To prove the converse, suppose that $(\bar{x}, \bar{y})$ is AKKT for $\text{TNLP}_{I_0}(\bar{x}, \bar{y})$. Then there exist sequences $(x^k, y^k) \subset \mathbb{R}^n \times \mathbb{R}^n$ and

$$(\lambda^k) = \left(\lambda^{g,k}, \lambda^{h,k}, \lambda^{\theta,k}, \lambda_{I_0}^{G,k}, \lambda^{H,k}, \lambda^{\tilde{H},k}\right) \subset \mathbb{R}_+^m \times \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}^{|I_0|} \times \mathbb{R}^n \times \mathbb{R}_+^n,$$

with $\lambda_i^{H,k} \geq 0$ for $i \in I_{0+} \cup I_{01}$, such that $(x^k, y^k) \to (\bar{x}, \bar{y})$,

$$\nabla_x L(x^k, \lambda^{g,k}, \lambda^{h,k}) + \sum_{i \in I_0} \lambda_i^{G,k} \nabla G_i(x^k) \to 0, \tag{3.22a}$$

$$\lambda^{\theta,k} \nabla \theta(y^k) + \sum_{i=1}^{n} \lambda_i^{H,k} \nabla H_i(y^k) + \sum_{i=1}^{n} \lambda_i^{\tilde{H},k} \nabla \tilde{H}_i(y^k) \to 0, \tag{3.22b}$$

$$\min\{-g(x^k), \lambda^{g,k}\} \to 0\,, \quad \min\{-\theta(y^k), \lambda^{\theta,k}\} \to 0, \tag{3.22c}$$

$$\min\{-H_i(y^k), \lambda_i^{H,k}\} \to 0,\ i \in I_{0+} \cup I_{01}\,, \quad \min\{-\tilde{H}(y^k), \lambda^{\tilde{H},k}\} \to 0. \tag{3.22d}$$

Extending the sequence $\left(\lambda_{I_0}^{G,k}\right)$ from $\mathbb{R}^{|I_0|}$ to $\mathbb{R}^n$ by setting $\lambda_i^{G,k} = 0$ for $i \notin I_0$, we can rewrite (3.22a) as

$$\nabla_x L(x^k, \lambda^{g,k}, \lambda^{h,k}) + \sum_{i=1}^{n} \lambda_i^{G,k} \nabla G_i(x^k) \to 0. \tag{3.23}$$

Moreover, for $i \in I_0$, it holds $G_i(x^k) \to G_i(\bar{x}) = \bar{x}_i = 0$. Thus,

$$\min\{|G_i(x^k)|, |\lambda_i^{G,k}|\} \to 0 \tag{3.24}$$

for all $i = 1, \ldots, n$. Besides, for $i \in I_{00} \cup I_{\pm 0}$, we have

$$H_i(y^k) \to H_i(\bar{y}) = -\bar{y}_i = 0,$$

which implies $\min\{-H_i(y^k), |\lambda_i^{H,k}|\} \to 0$. Therefore, in view of (3.22d) and the fact that $\lambda_i^{H,k} \geq 0$ for $i \in I_{0+} \cup I_{01}$, we have

$$\min\{-H_i(y^k), |\lambda_i^{H,k}|\} \to 0 \tag{3.25}$$

for all $i = 1, \ldots, n$. Thus, from (3.22b), (3.22c), the second part of (3.22d), (3.23), (3.24) and (3.25), we conclude that $(\bar{x}, \bar{y})$ satisfies the conditions of Definition 3.1.1, that is, $(\bar{x}, \bar{y})$ is an AW-stationary point for the problem (2.3). □

The above relations are summarized in the diagram presented in Figure 3.2.

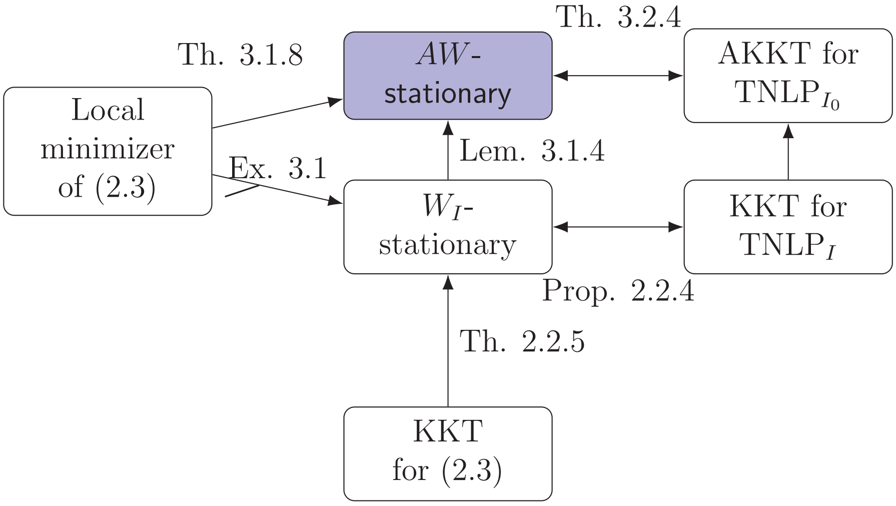

Figura 3.2: Relationships between local optimality and our stationarity concepts. An arrow indicates a strict implication between two conditions.

3.3 Future perspectives

In this chapter, we have defined a new sequential optimality condition for MPCaC and established several results. Despite of that, we are aware that there are open questions to be addressed in future works. Such questions concern both theoretical aspects and algorithmic consequences. Some of them are discussed below.

On the theoretical issues, we believe that it is possible to strengthen the definition of AW-stationarity, either by defining an AW_I-stationarity, associated with a set I in the range (2.4); and/or by adding another requirement in Definition 3.1.1, perhaps involving the product $\lambda^{G,k} * \lambda^{H,k}$, similarly to that made in [27]. This improvement should be able to establish that the new optimality condition must be legitimate, in the sense that every minimizer satisfies it, independently of any constraint qualification. Moreover, it should be at least as strong as PAKKT and CAKKT.

Now, it is known that improving the convergence of an algorithm by means of sequential condition requires stating the associated constraint qualification, because ultimately we want to say something about the exact stationarity. In other words, it is of interest to investigate under which MPCaC-tailored constraint qualifications our sequential condition guarantees W_I-stationary points or even KKT points, similarly to that established in [31,33,34].

Another subject of great interest nowadays concerns to second-order conditions. We think that the sequential condition introduced in this work will support the proposal of a second-order condition for MPCaC problems and the study of convergence to second-order stationary points.

Finally, we turn our attention to the computational aspects that may be derived from the results established in this work. It is necessary to measure the strength and usefulness of our sequential optimality condition by connecting it to practical algorithms with the hope of improving their convergence.

Conclusions

In many areas of applications of optimization we aim to find sparse solutions. In this work, we followed the approach that considers the cardinality constrained problem MPCaC to obtain sparse solutions.

We have presented a new stationarity condition, weaker than KKT, called W_I-stationarity, for the MPCaC problem. We proposed a unified approach that goes from the weakest to the strongest stationarity (within a certain range of conditions). Several theoretical results were presented, such as properties and relations concerning the reformulated problems and the original one. In addition, we discussed the relaxed problem by analyzing the general constraints in two cases, linear and nonlinear, with results, examples and counterexamples.

As we have seen in this work, KKT implies W_I-stationarity. However, despite being weaker than KKT, W_I-stationarity is not a necessary optimality condition. Therefore, another subject of research was the proposal of sequential optimality conditions for MPCaC problems in the hope of obtaining a weaker condition than W_I-stationarity, which would be satisfied at every minimizer, independently of any constraint qualification. Thus, we have presented a sequential optimality condition, namely Approximate Weak stationarity (AW-stationarity), for MPCaC problems.

Several theoretical results were presented, such as: AW-stationarity is a legitimate optimality condition independently of any constraint qualification; every feasible point of MPCaC is AKKT; the equivalence between the AW-stationarity and AKKT for the tightened problem TNLP_{I_0}. In addition, we have established some relationships between our AW-stationarity and other usual sequential optimality conditions, such as AKKT, CAKKT and PAKKT, by means of properties, examples and counterexamples.

This in turn will allow us to discuss algorithmic consequences, once we believe it is worth studying MPCaC, which is an important problem, both theoretically and numerically. It should be mentioned that, despite this computational appeal, our aim until this moment was to discuss the theoretical aspects of such conditions for MPCaC problems. The algorithmic aspects behind our theory are subject of ongoing research, since we consider it is possible to obtain suitable algorithms for solving MPCaC.

Apêndice A

Appendix

In this appendix we present some additional results and examples discussed along the seminars. Besides, we present here alternative proofs for some of the results proved in this work.

The following lemma provides another characterization of the tangent cone.

Lemma A.1 *The tangent cone to Ω at $\bar{x}$ can be expressed alternatively as*

$$T_\Omega(\bar{x}) = \{0\} \cup \left\{ d \in \mathbb{R}^n \mid \exists (x^k) \subset \Omega : x^k \to \bar{x} \text{ and } \frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \to \frac{d}{\|d\|} \right\}.$$

Proof. Consider $d \in \mathbb{R}^n \setminus \{0\}$ and a sequence $(x^k) \subset \Omega$ such that $x^k \to \bar{x}$ and

$$\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \to \frac{d}{\|d\|}.$$

Defining $t_k = \dfrac{\|x^k - \bar{x}\|}{\|d\|}$, we have $t_k \to 0$ and $\dfrac{x^k - \bar{x}}{t_k} \to d$. Conversely, if $(x^k) \subset \Omega$, $t_k \to 0$ and $\dfrac{x^k - \bar{x}}{t_k} \to d$, then $\dfrac{\|x^k - \bar{x}\|}{t_k} \to \|d\|$. Therefore,

$$\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \to \frac{d}{\|d\|}.$$

Moreover, $x^k - \bar{x} = t_k \dfrac{x^k - \bar{x}}{t_k} \to 0$ completing the proof. □

Now we obtain explicit representations for the cones of the Lemma 1.1.5 and describe precisely the constraint qualification ACQ as well.

Proposition A.2 *Consider the set Ω, given in (1.5) and a point $(\bar{x}, \bar{y}) \in \Omega$. Then,*

$$T_\Omega(\bar{x}, \bar{y}) = \big\{ (u, v) \in \mathbb{R}^n \times \mathbb{R}^n \mid u_{I_{0>}} = 0;\ v_{I_{\pm 0}} = 0;\ v_{I_{00}} \geq 0;\ u * v = 0 \big\}$$

and

$$D_\Omega(\bar{x}, \bar{y}) = \big\{ (u, v) \in \mathbb{R}^n \times \mathbb{R}^n \mid u_{I_{0>}} = 0;\ v_{I_{\pm 0}} = 0;\ v_{I_{00}} \geq 0 \big\}$$

Therefore, ACQ holds if and only if $I_{00} = \emptyset$.

Proof. Indeed, given $(u, v) \in T_\Omega(\bar{x}, \bar{y})$, there exist sequences $(x^k, y^k) \subset \Omega$ and $t_k \to 0$ such that

$$\frac{x^k - \bar{x}}{t_k} \to u \quad \text{and} \quad \frac{y^k - \bar{y}}{t_k} \to v.$$

If $i \in I_{0>}$, then $\bar{x}_i = 0$ and $\bar{y}_i > 0$, which implies that $y_i^k > 0$ (for all sufficiently large k) and hence $x_i^k = 0$. So, $u_i = \lim_{k\to\infty} \dfrac{x_i^k - \bar{x}_i}{t_k} = 0$. If $i \in I_{\pm 0}$, then $\bar{x}_i \neq 0$ and $\bar{y}_i = 0$. Hence, $x_i^k \neq 0$ (for all sufficiently large k), giving $y_i^k = 0$. So, $v_i = \lim_{k\to\infty} \dfrac{y_i^k - \bar{y}_i}{t_k} = 0$. For $i \in I_{00}$, we have $v_i \geq 0$, because $\bar{y}_i = 0$, $y^k \geq 0$ and $t_k > 0$. To see that $u * v = 0$, it is enough to prove that $u_i v_i = 0$ for $i \in I_{00}$. But for such an i we have $0 = \dfrac{x_i^k y_i^k}{t_k^2} \to u_i v_i$.

Now, consider a vector $(u, v) \in \mathbb{R}^n \times \mathbb{R}^n$ such that $u_{I_{0>}} = 0$, $v_{I_{\pm 0}} = 0$, $v_{I_{00}} \geq 0$ and $u * v = 0$. Let us prove that $(u, v) \in T_\Omega(\bar{x}, \bar{y})$. For this purpose, define the sequence $(x^k, y^k) \subset \mathbb{R}^n \times \mathbb{R}^n$ by

$$x^k = \bar{x} + \frac{1}{k}u \quad \text{and} \quad y^k = \bar{y} + \frac{1}{k}v.$$

If $i \in I_{\pm 0} \cup I_{00}$, then $\bar{y}_i = 0$, $v_i \geq 0$ and hence, $y_i^k \geq 0$. For $i \in I_{0>}$, we have $\bar{y}_i > 0$, which implies that $y_i^k > 0$ (for sufficiently large k). Moreover, it can be easily verified that

$$\left(\bar{x}_i + \frac{1}{k}u_i\right)\left(\bar{y}_i + \frac{1}{k}v_i\right) = 0$$

for all i by analyzing the three cases $i \in I_{\pm 0}$, $i \in I_{0>}$ and $i \in I_{00}$. Thus we conclude that $(x^k, y^k) \subset \Omega$ and hence, $(u, v) \in T_\Omega(\bar{x}, \bar{y})$.

To analyze the linearized cone, denote the constraints that define Ω by $\zeta(x, y) = -y$ and $\xi(x, y) = x * y$. Given $(u, v) \in D_\Omega(\bar{x}, \bar{y})$, we have

$$\bar{y}_i u_i + \bar{x}_i v_i = \nabla \xi_i(\bar{x}, \bar{y})^T \begin{pmatrix} u \\ v \end{pmatrix} = 0.$$

Therefore, if $i \in I_{0>}$, we have $\bar{x}_i = 0$ and $\bar{y}_i > 0$, which implies that $u_i = 0$. For $i \in I_{\pm 0}$, it holds $\bar{x}_i \neq 0$ and $\bar{y}_i = 0$, giving $v_i = 0$. On the other hand, for $i \in I_{00}$, the constraint ζ_i is active and hence,

$$-v_i = \nabla \zeta_i(\bar{x}, \bar{y})^T \begin{pmatrix} u \\ v \end{pmatrix} \leq 0,$$

which means that $v_i \geq 0$. This proves the inclusion $T_\Omega(\bar{x}, \bar{y}) \subset D_\Omega(\bar{x}, \bar{y})$. To prove the reverse inclusion, consider a vector $(u, v) \in \mathbb{R}^n \times \mathbb{R}^n$ such that $u_{I_{0>}} = 0$, $v_{I_{\pm 0}} = 0$ and $v_{I_{00}} \geq 0$. Then,

$$\nabla \zeta_j(\bar{x}, \bar{y})^T \begin{pmatrix} u \\ v \end{pmatrix} = -v_j \leq 0$$

for $j \in I_{\pm 0} \cup I_{00}$. Moreover, it is easy to see that

$$\nabla \xi_i(\bar{x}, \bar{y})^T \begin{pmatrix} u \\ v \end{pmatrix} = \bar{y}_i u_i + \bar{x}_i v_i = 0$$

for all i by considering the three cases $i \in I_{\pm 0}$, $i \in I_{0>}$ and $i \in I_{00}$, proving that $(u, v) \in D_\Omega(\bar{x}, \bar{y})$.

Finally, if there exists an index $\ell \in I_{00}$, then $(e_\ell, e_\ell) \in D_\Omega(\bar{x}, \bar{y}) \setminus T_\Omega(\bar{x}, \bar{y})$. So, ACQ implies $I_{00} = \emptyset$. Conversely, assume that $I_{00} = \emptyset$. Thus, given $(u, v) \in D_\Omega(\bar{x}, \bar{y})$ and $i \in \{1, \ldots, n\}$, we have $i \in I_{0>}$ or $i \in I_{\pm 0}$. In any case it holds $u_i v_i = 0$, which implies that $D_\Omega(\bar{x}, \bar{y}) \subset T_\Omega(\bar{x}, \bar{y})$, that is, ACQ holds. □

In the next lemma we give an alternative and direct proof of the closedness of the set defined by the cardinality constraint. Note that this is not so obvious because the function $x \mapsto \|x\|_0$ is not continuous.

Lemma A.3 *The set $\mathcal{C} = \{x \in \mathbb{R}^n \mid \|x\|_0 \leq \alpha\}$ is closed.*

Proof. Let $(x^k) \subset \mathcal{C}$ be a sequence such that $x^k \to \bar{x}$. Since $\|x^k\|_0 \leq \alpha$, the set $J_k = \{i \mid x_i^k = 0\}$ satisfies $|J_k| \geq n - \alpha$ for all $k \in \mathbb{N}$. Moreover, there are only finitely many possible choices of index sets J_k and hence at least one index set occurs infinitely often in the sequence, say $J_k = J$ for all $k \in \mathbb{N}' \subset \mathbb{N}$. Thus, $x_j^k = 0$ for all $j \in J$ and $k \in \mathbb{N}'$, which yields $\bar{x}_j = \lim_{k \in \mathbb{N}'} x_j^k = 0$ for all $j \in J$. Therefore, as $|J| \geq n - \alpha$, we conclude that $\|\bar{x}\|_0 \leq \alpha$. □

We give below an alternative and direct proof of Theorem 1.1.7.

Theorem A.4 *[Theorem 1.1.7] Consider the set Ω, defined in (1.8), and a feasible point $(\bar{x}, \bar{y}) \in \Omega$. If $I_{00}(\bar{x}, \bar{y}) = \emptyset$, then ACQ holds at $(\bar{x}, \bar{y})$.*

Proof. In view of Lemma 1.1.2, we can assume without loss of generality that there is no inactive constraint at $(\bar{x}, \bar{y})$. Denote $\rho(x,y) = Ax - b$, $\tilde{\rho}(x,y) = \tilde{A}x - \tilde{b}$, $\zeta(x,y) = My - r$, $\tilde{\zeta}(x,y) = \tilde{M}x - \tilde{r}$ and $\xi(x,y) = x * y$. Given $d = (u,v) \in \mathbb{R}^n \times \mathbb{R}^n$, we have

$$\nabla \rho(\bar{x}, \bar{y})^T d = Au, \tag{A.1a}$$

$$\nabla \tilde{\rho}(\bar{x}, \bar{y})^T d = \tilde{A}u, \tag{A.1b}$$

$$\nabla \zeta(\bar{x}, \bar{y})^T d = Mv, \tag{A.1c}$$

$$\nabla \tilde{\zeta}(\bar{x}, \bar{y})^T d = \tilde{M}v, \tag{A.1d}$$

$$\nabla \xi_i(\bar{x}, \bar{y})^T d = \bar{y}_i u_i + \bar{x}_i v_i. \tag{A.1e}$$

Take an arbitrary vector $d = (u,v) \in D_\Omega(\bar{x}, \bar{y})$. Then, using (A.1e),

$$\bar{y}_i u_i + \bar{x}_i v_i = \nabla \xi_i(\bar{x}, \bar{y})^T d = 0$$

for all $i = 1, \ldots, n$. Thus,

$$u_{I_{0\pm}} = 0 \quad \text{and} \quad v_{I_{\pm 0}} = 0. \tag{A.2}$$

In order to prove that $d \in T_\Omega(\bar{x}, \bar{y})$, define $t_k = 1/k$ and the sequence (x^k, y^k) by

$$x^k = \bar{x} + t_k u \quad \text{and} \quad y^k = \bar{y} + t_k v.$$

Thus, $\dfrac{(x^k, y^k) - (\bar{x}, \bar{y})}{t_k} \to (u,v) = d$. It remains to prove that $(x^k, y^k) \subset \Omega$. By (A.2) we obtain $x^k_{I_{0\pm}} = 0$ and $y^k_{I_{\pm 0}} = 0$. Since $I_{00} = \emptyset$, we have $I_{0\pm} \cup I_{\pm 0} = \{1, \ldots, n\}$ and then $x^k * y^k = 0$. Moreover, as $(u,v) \in D_\Omega(\bar{x}, \bar{y})$, using (A.1a)–(A.1d), we obtain

$$Au = 0, \quad \tilde{A}u \leq 0, \quad Mv = 0 \quad \text{and} \quad \tilde{M}v \leq 0$$

yielding immediately

$$Ax^k = b, \quad \tilde{A}x^k \leq \tilde{b}, \quad My^k = r \quad \text{and} \quad \tilde{M}y^k \leq \tilde{r}.$$

So, $(x^k, y^k) \subset \Omega$ and then $d \in T_\Omega(\bar{x}, \bar{y})$. □

We show the following three equivalent (cf. Proposition A.5, below) conditions for the optimization problem (1.1). Recall that the Lagrangian of this problem is given in (1.4).

Condition 1 *There exist sequences $(x^k) \subset \mathbb{R}^n$ and $(\lambda^k) = \left(\lambda^{g,k}, \lambda^{h,k}\right) \subset \mathbb{R}^m_+ \times \mathbb{R}^p$ such that $x^k \to \bar{x}$,*

$$\nabla_x L(x^k, \lambda^{g,k}, \lambda^{h,k}) \to 0, \tag{A.3a}$$

$$\min\{-g(x^k), \lambda^{g,k}\} \to 0. \tag{A.3b}$$

Condition 2 *There exist sequences* $(x^k) \subset \mathbb{R}^n$ *and* $(\lambda^k) = \left(\lambda^{g,k}, \lambda^{h,k}\right) \subset \mathbb{R}^m_+ \times \mathbb{R}^p$ *such that* $x^k \to \bar{x}$,

$$\nabla_x L(x^k, \lambda^{g,k}, \lambda^{h,k}) \to 0, \tag{A.4a}$$

$$\lambda_i^{g,k} = 0 \text{ for all } i \notin I_g(\bar{x}). \tag{A.4b}$$

Condition 3 *There exist sequences* $(x^k) \subset \mathbb{R}^n$, $(\lambda^k) = \left(\lambda^{g,k}, \lambda^{h,k}\right) \subset \mathbb{R}^m_+ \times \mathbb{R}^p$ *and* $(\varepsilon_k) \subset \mathbb{R}_+$ *such that* $x^k \to \bar{x}$, $\varepsilon_k \to 0$ *and for all* $k \in \mathbb{N}$,

$$\|\nabla_x L(x^k, \lambda^{g,k}, \lambda^{h,k})\| \le \varepsilon_k, \tag{A.5a}$$

$$\|g^+(x^k)\| \le \varepsilon_k, \quad \|h(x^k)\| \le \varepsilon_k, \tag{A.5b}$$

$$\lambda_i^{g,k} = 0 \text{ if } g_i(x^k) < -\varepsilon_k. \tag{A.5c}$$

Proposition A.5 *Conditions 1, 2 and 3 are equivalent.*

Proof. First, assume that the sequences (x^k) and (λ^k) satisfy Condition 1. If $i \in \{1, \ldots, m\}$ is such that $g_i(\bar{x}) < 0$, then there exists $\delta > 0$ satisfying $-g_i(x^k) \ge \delta$ for all $k \in \mathbb{N}$ sufficiently large. So, in view of (A.3b), we have $\lambda_i^{g,k} \to 0$. Since $\nabla g_i(x^k) \to \nabla g_i(\bar{x})$, we have

$$\sum_{i \notin I_g(\bar{x})} \lambda_i^{g,k} \nabla g_i(x^k) \to 0.$$

Subtracting this from (A.3a), we obtain

$$\nabla f(x^k) + \sum_{i \in I_g(\bar{x})} \lambda_i^{g,k} \nabla g_i(x^k) + \nabla h(x^k) \lambda^{h,k} \to 0.$$

Therefore, defining $(\bar{\lambda}^{g,k}) \subset \mathbb{R}^m_+$ by

$$\bar{\lambda}_i^{g,k} = \begin{cases} \lambda_i^{g,k}\,, \text{ if } i \in I_g(\bar{x}) \\ 0\,, \text{ otherwise,} \end{cases}$$

we have

$$\nabla_x L(x^k, \bar{\lambda}^{g,k}, \lambda^{h,k}) \to 0.$$

This means that the sequences (x^k) and $(\bar{\lambda}^{g,k}, \lambda^{h,k})$ satisfy Condition 2.

Now, assume that Condition 2 is true and define

$$\varepsilon_k = \max_{i \in I_g(\bar{x})} \left\{ -g_i(x^k), \|\nabla_x L(x^k, \lambda^{g,k}, \lambda^{h,k})\|, \|g^+(x^k)\|, \|h(x^k)\| \right\}.$$

By (A.4a) and the facts that $\|g^+(x^k)\| \to \|g^+(\bar{x})\| = 0$, $\|h(x^k)\| \to \|h(\bar{x})\| = 0$ and $g_i(x^k) \to g_i(\bar{x}) = 0$, $i \in I_g(\bar{x})$, we conclude that $\varepsilon_k \to 0$. Furthermore, from the definition of ε_k we immediately have (A.5a) and (A.5b). To prove (A.5c), note that if $g_i(x^k) < -\varepsilon_k$, then $i \notin I_g(\bar{x})$ because $-\varepsilon_k \le g_i(x^k)$ for all $i \in I_g(\bar{x})$. Thus, from (A.4b), we have $\lambda_i^{g,k} = 0$, proving Condition 3.

Finally, suppose that Condition 3 is valid. From (A.5a) we get (A.3a). To see (A.3b), consider $i \in \{1, \ldots, m\}$ fixed and assume first that $i \in I_g(\bar{x})$. Defining

$$\mathbb{N}' = \{k \in \mathbb{N} \mid -g_i(x^k) \le \lambda_i^{g,k}\} \quad \text{and} \quad \mathbb{N}'' = \{k \in \mathbb{N} \mid -g_i(x^k) > \lambda_i^{g,k}\},$$

we have

$$\min\{-g_i(x^k), \lambda_i^{g,k}\} = \begin{cases} -g_i(x^k)\,, \text{ if } k \in \mathbb{N}' \\ \lambda_i^{g,k} \in [0, -g_i(x^k)]\,, \text{ if } k \in \mathbb{N}''. \end{cases}$$

Since $g_i(x^k) \to g_i(\bar{x}) = 0$, we obtain (A.3b) for this i. On the other hand, if $i \notin I_g(\bar{x})$, then $g_i(\bar{x}) < 0$, which in turn implies that $g_i(x^k) < -\varepsilon_k$ for all $k \in \mathbb{N}$ sufficiently large. Thus, using (A.5c) we obtain $\lambda_i^{g,k} = 0$ and then $\min\{-g_i(x^k), \lambda_i^{g,k}\} = 0$ for all $k \in \mathbb{N}$ sufficiently large, proving (A.3b) in this case and completing the proof. □

Referências Bibliográficas

[1] Bienstock D. Computational study of a family of mixed-integer quadratic programming problems. Math Program. 1996;74(2):121–140.

[2] Burdakov O, Kanzow C, Schwartz A. Mathematical programs with cardinality constraints: reformulation by complementarity-type conditions and a regularization method. SIAM J Optim. 2016; 26(1):397–425.

[3] Burke J.V. A sequential quadratic programming algorithm for potentially infeasible mathematical programs. J Math Anal Appl. 1989;139:319–351.

[4] Gill P.E, Murray W, Saunders M.A. SNOPT: An SQP algorithm for large-scale constrained optimization. SIAM J Optim. 2002;12(4):979–1006.

[5] Ribeiro A.A, Sachine M, Santos S.A. On the augmented subproblems within sequential methods for nonlinear programming. Comp Appl Math. 2017;36:1255–1272.

[6] Ribeiro A.A, Sachine M, Santos S.A. On the approximate solutions of augmented subproblems within sequential methods for nonlinear programming. Comp Appl Math. 2018;37:6601–6618.

[7] Svanberg K. A class of globally convergent optimization methods based on conservative convex separable approximations. SIAM J Optim. 2002;12:555–573.

[8] Branda M, Bucher M, Červinka M, Schwartz A. Convergence of a Scholtes-type regularization method for cardinality-constrained optimization problems with an application in sparse robust portfolio optimization. Comput Optim Appl. 2018;70(2):503–530.

[9] Candès E.J, Wakin M.B. An introduction to compressive sampling. IEEE Signal Process Mag. 2008; 25(2):21–30.

[10] d'Aspremont A, Bach F, Ghaoui L.E. Optimal solutions for sparse principal component analysis. J Mach Learn Res. 2008;9:1269–1294.

[11] Markowitz H. Portfolio selection. The Journal of Finance. 1952;7:77–91.

[12] Miller A. Subset selection in regression. Florida: Chapman and Hall/CRC Press; 2002.

[13] Tibshirani R. Regression shrinkage and selection via the Lasso. J R Statist Soc. 1996;58:267–288.

[14] Beck A, Eldar Y.C. Sparsity constrained nonlinear optimization: Optimality conditions and algorithms. SIAM J Optim. 2013;23(3):1480–1509.

[15] Ruiz-Torrubiano R, García-Moratilla S, Suárez A. Optimization problems with cardinality constraints. In: Tenne Y, Goh CK, editors. Computational intelligence in optimization. Berlin: Academic Press; 2010. p. 105–130.

[16] Sun X, Zheng X, Li D. Recent advances in mathematical programming with semi-continuous variables and cardinality constraint. J Oper Res Soc China. 2013;1:55–77.

[17] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. New York, NY, USA: Springer New York Inc.; 2001.

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

[19] Burdakov O, Kanzow C, Schwartz A. On a reformulation of mathematical programs with cardinality constraints. In: Gao D, Ruan N, Xing W, editors. Advances in Global Optimization. Springer Proceedings in Mathematics and Statistics; 2015.

[20] Li X, Song W. The first-order necessary conditions for sparsity constrained optimization. J Oper Res Soc China. 2015;3:521–535.

[21] Pan L.L, Xiu N.H, Zhou S.L. On solutions of sparsity constrained optimization. J Oper Res Soc China. 2015;3:421–439.

[22] Dussault J.P, Haddou M, Kadrani A, Migot T. How to compute an M-stationary point of the MPCC. Canada: Université Sherbrooke; 2019.

[23] Outrata J.V. Optimality conditions for a class of mathematical programs with equilibrium constraints. Math Oper Res. 1999;24(3):627–644.

[24] Scheel H, Scholtes S. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. Math Oper Res. 2000;25(1):1–22.

[25] Hoheisel T, Kanzow C, Schwartz A. Convergence of a local regularization approach for mathematical programmes with complementarity or vanishing constraints. Optim Methods Softw. 2012;27(3):483–512.

[26] Izmailov A.F. Mathematical programs with complementarity constraints: Regularity, optimality conditions, and sensitivity. Comput Math and Math Phys. 2004;44(7):1145–1164.

[27] Andreani R, Haeser G, Secchin L.D, Silva P.J.S. New sequential optimality conditions for mathematical problems with complementarity constraints and algorithmic consequences. SIAM J Optim. 2019;29(4):3201–3230.

[28] Ramos A. Mathematical programs with equilibrium constraints: A sequential optimality condition, new constraint qualifications and algorithmic consequences. Optim Methods Softw. 2019;:1–37.

[29] Mehlitz P. Stationarity conditions and constraint qualifications for mathematical programs with switching constraints. Math Program. 2020;181(1):149–186.

[30] Červinka M, Kanzow C, Schwartz A. Constraint qualifications and optimality conditions for optimization problems with cardinality constraints. Math Program. 2016;160:353–377.

[31] Andreani R, Fazzio N.S, Schuverdt M.L, Secchin L.D. A sequential optimality condition related to the quasinormality constraint qualification and its algorithmic consequences. SIAM J Optim. 2019; 29:743–766.

[32] Andreani R, Haeser G, Martínez J.M. On sequential optimality conditions for smooth constrained optimization. Optimization. 2011;60:627–641.

[33] Andreani R, Martínez J.M, Ramos A, Silva P.J.S. A cone-continuity constraint qualification and algorithmic consequences. SIAM J Optim. 2016;26(1):96–110.

[34] Andreani R, Martínez J.M, Ramos A, Silva P.J.S. Strict constraint qualifications and sequential optimality conditions for constrained optimization. Math Oper Res. 2018;43(3):693–717.

[35] Andreani R, Martínez J.M, Svaiter B.F. A new sequential optimality condition for constrained optimization and algorithmic consequences. SIAM J Optim. 2010;6:3533–3554.

[36] Martínez J.M, Svaiter B.F. A practical optimality condition without constraint qualifications for nonlinear programming. J Optim Theory Appl. 2003;118:117–133.

[37] Birgin E.G, Martínez J.M. Practical augmented lagrangian methods for constrained optimization. Philadelphia: SIAM; 2014.

[38] Helou E.S, Santos S.A, Simões L.E.A. Analysis of a new sequential optimality condition applied to mathematical programs with equilibrium constraints. J Optim Theory Appl. 2020;185:433–447.

[39] Helou E.S, Santos S.A, Simões L.E.A. A new sequential optimality condition for constrained nonsmooth optimization. SIAM J Optim. 2020;30(2):1610–1637.

[40] Guignard M. Generalized Lagrangian Kuhn-Tucker conditions for mathematical programming problems in a banach space. SIAM J. Control Optim. 1969;7(2):232–241.

[41] Gould F.J, Tolle J.W. A necessary and sufficient qualification for constrained optimization. SIAM J. Appl. Math. 1971;20:164–172.

[42] Mangasarian O.L, Fromovitz S. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. J Math Anal Appl. 1967;17:37–47.

[43] Bertsekas D.P. Nonlinear programming. Belmont, Massachusetts: Athena Scientific; 1995.

[44] Andreani R, Haeser G, Ramos A, Silva P.J.S. Two new weak constraint qualifications and applications. SIAM J Optim. 2012;22(3):1109–1135.

[45] Nocedal J, Wright S.A. Numerical optimization. Springer-Verlag.Springer Ser. Oper. Res. 1999.

[46] Qi L, Wei Z. On the constant positive linear dependence condition and its application to SQP methods. SIAM J Optim. 2000;10(4):963–981.

[47] Beck A. First-order methods in optimization. Philadelphia: SIAM; 2017.

[48] Achtziger W, Kanzow C. Mathematical programs with vanishing constraints: optimality conditions and constraint qualifications. Math Program. 2008;114:69–99.

[49] Flegel M.L, Kanzow C. On M-stationary points for mathematical programs with equilibrium constraints. J Math Anal Appl. 2005;310(1):286–302.

[50] Krulikovski E.H.M., Ribeiro A.A, Sachine M. On the Weak Stationarity Conditions for Mathematical Programs with Cardinality Constraints: A Unified Approach. Appl Math Optim. 2021.

[51] Krulikovski E.H.M., Ribeiro A.A, Sachine M. A sequential optimality condition for mathematical programs with cardinality constraints. Federal University of Paraná; 2020. arXiv: 2008.03158.