



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de São José do Rio Preto

Vinicius Cabral da Silva

**Condições sequenciais de otimalidade e condições
de qualificação em otimização não-linear**

São José do Rio Preto
2022

Vinicius Cabral da Silva

**Condições sequenciais de otimalidade e condições
de qualificação em otimização não-linear**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Matemática, área de Matemática Aplicada, junto ao Programa de Pós-Graduação em Matemática, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Orientador: Prof. Dr. Geraldo Nunes Silva

Financiadora: CNPq 133467/2020-1

São José do Rio Preto
2022

S586c	Silva, Vinicius Cabral da Condições sequenciais de otimalidade e condições de qualificação em otimização não-linear / Vinicius Cabral da Silva. -- São José do Rio Preto, 2022 100 f. Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto Orientador: Geraldo Nunes Silva 1. Otimização não-linear. 2. Condições de otimalidade. 3. Condições de qualificação. 4. Condições sequenciais. I. Título.
-------	---

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Vinicius Cabral da Silva

**Condições sequenciais de otimalidade e condições
de qualificação em otimização não-linear**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Matemática, área de Matemática Aplicada, junto ao Programa de Pós-Graduação em Matemática, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CNPq 133467/2020-1

Comissão Examinadora

Prof. Dr. Geraldo Nunes Silva
Orientador

Prof. Dr. Roberto Andreani
UNICAMP - Campinas

Prof. Dr. Valeriano Antunes de Oliveira
UNESP - São José do Rio Preto

São José do Rio Preto
30 de março de 2022

*À minha mãe Marta,
minha Avó Maria
e ao meu tio Geraldo,
dedico.*

AGRADECIMENTOS

Agradeço primeiramente a Deus, por ter me dado força e sabedoria nos momentos de dificuldade.

Aos professores Edson Donizete de Carvalho, Ernandes Rocha de Oliveira e Jaime Edmundo Apaza Rodriguez por todo apoio e incentivo em minha vida acadêmica.

Ao meu orientador, Prof. Dr. Geraldo Nunes Silva por todo o apoio e pela excelente orientação durante o mestrado.

Aos membros da comissão examinadora, Buzzi, Igner, Nino e Valeriano pela disponibilidade.

Aos meus familiares e amigos que de alguma forma contribuíram para a realização deste trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo financiamento durante todo o mestrado.

*A imaginação muitas vezes nos leva a mundos que nunca sequer existiram.
Mas sem ela, não vamos a lugar nenhum.*
Carl Sagan [62]

RESUMO

O principal objetivo desta dissertação é estudar métodos numéricos para a resolução de um problema de otimização restrita, utilizando como principais ferramentas as condições sequenciais de otimalidade de primeira e segunda ordem adequadas. É feito um estudo detalhado acerca de cada condição sequencial de otimalidade, mostrando suas consequências na implementação de algoritmos práticos para otimização restrita. Além disso, é apresentado resultados importantes em otimização não linear, como as condições de qualificação conhecidas e suas relações com as condições sequenciais.

Palavras-chave: Otimização não-linear. Condições de otimalidade. Condições de qualificação. Condições sequenciais.

ABSTRACT

The main objective of this dissertation is to study numerical methods for solving a constrained optimization problem, using as main tools the appropriate first and second order sequential optimality conditions. A detailed study is made about each sequential optimality condition, showing its consequences in the implementation of practical algorithms for constrained optimization. In addition, important results in nonlinear optimization are presented, such as the known constraint qualifications and their relationship with the sequential conditions.

Keywords: Nonlinear optimization. Optimality conditions. Constraint qualifications. Sequential conditions.

Lista de Figuras

2.1	As condições KKT falham em um minimizador	22
2.2	KKT não garante a validade de MFCQ	29
2.3	Relações entre as condições de qualificação conhecidas.	37
2.4	As condições CAKKT e PAKKT são independentes uma da outra.	57
2.5	Relações entre as principais condições sequenciais para otimização restrita.	59
2.6	Relações entre as condições de qualificação mínimas	69
3.1	A condição KKT não é suficiente para a otimalidade	71
3.2	MFCQ não garante a condição de segunda ordem	73
3.3	Representação do conjunto de multiplicadores de Lagrange	74

Sumário

Introdução	10
1 Fundamentos	12
1.1 Condições de otimalidade geométricas	15
1.1.1 Geometria variacional	18
2 Condições de Otimalidade de Primeira Ordem	21
2.1 Condições de Qualificação	25
2.2 Condições sequenciais de primeira ordem	38
2.2.1 Condição Aproximadamente KKT	39
2.2.2 Condição Gradiente Projetado Aproximado	42
2.2.3 Condição sequencial SAKKT	47
2.2.4 Condições sequenciais C-AGP e L-AGP	49
2.2.5 Condição sequencial CAKKT	52
2.2.6 Condições sequenciais mais recentes	55
2.3 Condições de qualificação fracas e relações	59
2.3.1 Propriedade de continuidade do cone	60
2.3.2 Condição de qualificação mais fraca associada a AGP	62
2.3.3 Condições de qualificação associadas a CAKKT e SAKKT	63
2.3.4 Condição de qualificação mais fraca associada a PAKKT	65
2.3.5 Condição de qualificação mais fraca associada a PCAKKT	66
3 Condições de Otimalidade de Segunda Ordem	70
3.1 Condições de qualificação de segunda ordem	73
3.2 Condição sequencial AKKT de segunda ordem	81
3.2.1 Propriedade do cone contínuo de segunda ordem	84
4 Conclusões	89
Referências	90
Apêndice A - Pré-requisitos	95
A.1 Definições	95
A.2 Topologia do $\mathbb{R}^n$	96
A.3 Análise convexa	97
Índice Remissivo	101

Introdução

O problema principal a ser considerado neste trabalho é o de minimizar uma função $f(x)$ sujeito a $x \in \Omega$, onde Ω é um subconjunto de $\mathbb{R}^n$ definido através de restrições de igualdades e desigualdades. Assim, nosso objetivo é encontrar os pontos $x^* \in \Omega$ que minimizam a função objetivo f . Esse tipo de problema é denominado problema de otimização restrita, podendo ser linear ou não linear. Problemas desta natureza são frequentemente encontrados no nosso cotidiano. Tais problemas vêm por exemplo da engenharia [18], pesquisa operacional [30, 46], finanças [26], tecnologia espacial [64], agricultura [54], dentre outros. Achar uma solução global para um problema de otimização restrita significa encontrar $x^* \in \Omega$ tal que $f(x^*) \leq f(x)$ para todo $x \in \Omega$. Porém, encontrar soluções globais pode ser muito difícil na prática e, portanto, nosso interesse nesta dissertação é trabalhar com a minimização local.

A maioria dos métodos numéricos existentes na literatura atual são iterativos. Com o desenvolvimento de computadores no Século XX tornou-se possível resolver muitos problemas de otimização restrita por meio de algoritmos iterativos. Entre os diferentes métodos, mencionamos o método Lagrangiano aumentado [22], programação quadrática sequencial [57], método de pontos interiores [24] e restauração inexata [50]. Esses métodos foram amplamente estudados na literatura com muitas aplicações importantes em otimização.

Ao utilizar um algoritmo iterativo para resolver um problema de otimização restrita, precisamos decidir se uma iteração computada está próxima ou não da solução ótima, ou seja, precisamos saber quando encerrar a execução do algoritmo. Uma vez que testar a otimalidade é muito difícil, o ideal é parar o algoritmo quando uma condição necessária de otimalidade é aproximadamente satisfeita. As condições de Karush-Kuhn-Tucker (KKT) [43, 44] que veremos ao longo desta dissertação, fornecem uma base teórica para a definição de critérios de parada adequados para esses algoritmos. Formas aproximadas das condições KKT são usadas para declarar que uma iteração é satisfatória o suficiente para fins de métodos iterativos práticos desde os anos 50, quando surgiram os primeiros algoritmos de otimização restrita.

As condições sequenciais de otimalidade foram introduzidas com o objetivo de estudar a convergência global de algoritmos iterativos. Como veremos ao longo deste trabalho, essas condições fornecem ferramentas teóricas adequadas para justificar critérios de parada para esses algoritmos. Em geral, uma condição sequencial de otimalidade deve cumprir as três propriedades abaixo:

- deve ser uma condição necessária de otimalidade independentemente de qualquer condição de qualificação;
- devem existir métodos numéricos significativos que geram sequências cujos pontos limites satisfazem a condição;

- deve implicar condições de otimalidade da forma “KKT ou não-CQ” para CQs fracas, onde CQ é uma condição de qualificação.

Uma vez que as condições sequenciais de otimalidade são condições necessárias para a otimização restrita, independentemente de condições de qualificação, a questão de sua força passa a ser relevante. Do ponto de vista prático, esse fato justifica a decisão tomada em todo software de otimização não testando as condições de qualificação. Por outro lado, o fato de que as condições sequenciais devem implicar condições de otimalidade da forma “KKT ou não-CQ”, existe um grande interesse em estudar condições de qualificação fracas, pois elas geram condições de otimalidade fortes que servem como uma medida de força para a condição sequencial. Portanto, estamos interessados em estudar condições sequenciais fortes, no sentido de que existam minimizadores que não a satisfaçam; que sejam de simples verificação prática e que possam ser usadas como critério de parada em algoritmos práticos.

A presente dissertação está dividida em quatro capítulos e um Apêndice. No primeiro capítulo, é feito um estudo dos principais conceitos que serão fundamentais para o desenvolvimento deste trabalho, como a teoria de cones e resultados básicos de análise variacional.

O capítulo dois está dividido em três seções. Na primeira seção, apresentamos as condições de qualificação de primeira ordem conhecidas na literatura e que serão de suma importância no estudo das condições sequenciais. Na segunda seção estudamos com detalhes as condições sequenciais de otimalidade de primeira ordem conhecidas atualmente na literatura, mostrando as relações entre elas. Na terceira seção, apresentamos, para cada condição sequencial, a condição de qualificação mínima associada que garante que todo minimizador local é um ponto estacionário. As referências [7, 8, 9, 38, 58] foram as principais utilizadas na elaboração deste capítulo.

No capítulo três, tratamos das condições de otimalidade de segunda ordem e das condições de qualificação de segunda ordem conhecidas atualmente. Em particular, apresentamos uma importante condição sequencial de segunda ordem, conhecida como condição aproximadamente KKT de segunda ordem (AKKT2) [58], apropriada para a análise de algoritmos com convergência a pontos estacionários de segunda ordem.

No último capítulo, apontamos as principais conclusões e possíveis pesquisas futuras envolvendo as condições sequenciais. No Apêndice, apresentamos alguns pré-requisitos fundamentais que foram utilizados durante o desenvolvimento do trabalho. Por fim, a riqueza nos detalhes das demonstrações e exemplos constituem a principal contribuição deste trabalho.

1 Fundamentos

Um problema geral de otimização restrita em dimensão finita consiste em

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && x \in \Omega, \end{aligned} \tag{1.1}$$

onde Ω é um subconjunto de $\mathbb{R}^n$ e $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função continuamente diferenciável num aberto que contém Ω .

O conjunto Ω é chamado de *conjunto viável* do problema, os pontos de Ω são chamados de *pontos viáveis*, e f é chamada *função objetivo*.

Definição 1.1. Dizemos que um ponto $x^* \in \Omega$ é

- (i) um minimizador global de (1.1) se $f(x^*) \leq f(x) \ \forall x \in \Omega$;
- (ii) um minimizador local de (1.1) se existe $\delta > 0$ tal que

$$f(x^*) \leq f(x), \ \forall x \in B(x^*; \delta) \cap \Omega.$$

Um minimizador global pode não existir, é o caso quando Ω é vazio, onde dizemos que o problema é inviável, ou quando existe uma sequência $\{x^k\} \subset \Omega$ tal que $f(x^k) \rightarrow -\infty$, onde dizemos que o problema é ilimitado. Uma condição suficiente para a existência de uma solução é que Ω seja compacto e não vazio. A maioria dos algoritmos não garantem a descoberta de uma solução ótima, mas sim, pontos que são possíveis candidatos à solução. Assim, nosso objetivo é estudar propriedades que sejam satisfeitas por todas as soluções do problema. Tais propriedades são denominadas *condições de otimalidade*.

Definição 1.2. Uma propriedade $\mathcal{P}$ sobre pontos de $\mathbb{R}^n$ é dita uma *condição de otimalidade* se

$$x \text{ é minimizador global} \Rightarrow x \text{ satisfaz } \mathcal{P}.$$

A propriedade de ser um minimizador local ou global são condições de otimalidade. Ser minimizador local é estritamente mais fraco que ser um minimizador global, porém, em geral pode ser muito difícil encontrar soluções locais.

Um exemplo clássico é quando o problema não possui restrições, ou seja, quando $\Omega = \mathbb{R}^n$. Neste caso, se a função objetivo f é diferenciável e se x^* é minimizador local, então $\nabla f(x^*) = 0$. Além disso, se f é duas vezes diferenciável, então $x^T \nabla^2 f(x^*) x \geq 0$ para todo $x \in \mathbb{R}^n$. Estas propriedades são condições de otimalidade que geralmente são simples de serem verificadas por envolverem somente o cálculo das derivadas de f .

Observação 1.3. Qualquer problema de maximização $\max f(x)$ sujeito a $x \in \Omega$, pode ser transformado em um problema de minimização equivalente $\min -f(x)$ sujeito a $x \in \Omega$. Em particular, as soluções locais e globais de ambos os problemas são as mesmas, com sinais opostos para os valores ótimos. Assim, do ponto de vista matemático, não existe nenhuma diferença relevante entre problemas de minimização e maximização. Por esta razão, vamos considerar apenas problemas de minimização.

Tipicamente, o conjunto viável é definido como

$$\Omega = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\},$$

onde $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ são funções continuamente diferenciáveis.

Equivalentemente, os pontos do conjunto Ω satisfazem o seguinte sistema de igualdades e desigualdades:

$$\begin{aligned} h_i(x) &= 0, \quad i = 1, \dots, m, \\ g_j(x) &\leq 0, \quad j = 1, \dots, p, \end{aligned}$$

onde as funções $h_i, g_j : \mathbb{R}^n \rightarrow \mathbb{R}$ são chamadas de *restrições funcionais*.

Denotamos por $A(x) = \{j \in \{1, \dots, p\} : g_j(x) = 0\}$ o conjunto de índices de restrições de desigualdade ativas em x .

A partir de agora, vamos considerar o problema

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && h(x) = 0, \\ &&& g(x) \leq 0, \end{aligned} \tag{1.2}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ são funções continuamente diferenciáveis. Salvo menção explícita, vamos considerar sempre

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

Quando a função objetivo f e as restrições são dadas por funções afins, o problema (1.2) é chamado de *problema de otimização linear*. Se alguma das funções é não linear, então o problema (1.2) é chamado de *problema de otimização não linear*.¹

Muitos problemas científicos e tecnológicos tais como engenharia, física, pesquisa operacional, finanças, entre outros, precisam da solução de problemas desta forma. Podemos encontrar diversos métodos numéricos para resolver um problema de otimização, como por exemplo, o algoritmo simplex no caso de problemas de otimização linear. No caso não linear, é necessário o estudo de condições de otimalidade adequadas que possam ser usadas como critério de parada em algoritmos práticos. As condições sequenciais de otimalidade que estudaremos ao longo desta dissertação são boas ferramentas para esse estudo.

Alguns algoritmos procuram determinar uma direção segundo a qual f decresce. Essas direções são chamadas de *direções de descida*.

Definição 1.4. Dada uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$, um ponto $x \in \mathbb{R}^n$ e uma direção $d \in \mathbb{R}^n \setminus \{0\}$, dizemos que d é uma *direção de descida* para f , a partir de x , quando existe $\delta > 0$ tal que $f(x + td) < f(x)$, para todo $t \in (0, \delta)$.

¹Haeser e Ramos [38], p. 3

Teorema 1.5. Se $\nabla f(x)^T d < 0$, então d é uma direção de descida para f , a partir de x .

Demonstração. Temos que

$$\nabla f(x)^T d = \frac{\partial f}{\partial d}(x) = \lim_{t \rightarrow 0} \frac{f(x + td) - f(x)}{t}.$$

Pela hipótese e pela preservação do sinal, existe $\delta > 0$ tal que

$$\frac{f(x + td) - f(x)}{t} < 0,$$

para todo $t \in (-\delta, \delta) \setminus \{0\}$. Portanto, $f(x + td) < f(x)$, para todo $t \in (0, \delta)$. $\square$

Um resultado importante é Lema de Carathéodory:

Lema 1.6 (Carathéodory). Sejam $u_1, \dots, u_m, v_1, \dots, v_p \in \mathbb{R}^n$. Suponha que

$$x = \sum_{i=1}^m \lambda_i u_i + \sum_{j=1}^p \mu_j v_j$$

com $\mu_j \geq 0$, para todo $j = 1, \dots, p$. Então existem $\mathcal{I} \subset \{1, \dots, m\}$, $\mathcal{J} \subset \{1, \dots, p\}$ e escalares $\tilde{\lambda}_i$, $i \in \mathcal{I}$, $\tilde{\mu}_j \geq 0$, $j \in \mathcal{J}$ tais que

$$x = \sum_{i \in \mathcal{I}} \tilde{\lambda}_i u_i + \sum_{j \in \mathcal{J}} \tilde{\mu}_j v_j$$

e os vetores $\{u_i\}_{i \in \mathcal{I}} \cup \{v_j\}_{j \in \mathcal{J}}$ são linearmente independentes.

Demonstração. Sem perda de generalidade, suponhamos que $\lambda_i \neq 0$ e $\mu_j > 0$ para todo $i = 1, \dots, m$, $j = 1, \dots, p$.

Suponha que os vetores $\{u_i\}_{i=1}^m \cup \{v_j\}_{j=1}^p$ são linearmente dependentes. Logo, existem escalares α_i e β_j , não todos nulos, tais que

$$\sum_{i=1}^m \alpha_i u_i + \sum_{j=1}^p \beta_j v_j = 0.$$

Logo, usando a hipótese, temos

$$x = \sum_{i=1}^m (\lambda_i - t\alpha_i) u_i + \sum_{j=1}^p (\mu_j - t\beta_j) v_j$$

para todo $t \in \mathbb{R}$. Para $t = 0$, nenhum dos coeficientes na igualdade acima se anula. Seja $\tilde{t}$ o t de menor módulo que anula pelo menos um dos coeficientes $\lambda_i - t\alpha_i$ ou $\mu_j - t\beta_j$. Assim,

$$x = \sum_{i=1}^m (\lambda_i - \tilde{t}\alpha_i) u_i + \sum_{j=1}^p (\mu_j - \tilde{t}\beta_j) v_j.$$

Claramente, $\mu_j - \tilde{t}\beta_j \geq 0$ para todo j , mas conseguimos escrever x como combinação linear de, no máximo, $m + p - 1$ vetores. Este processo pode ser repetido até que os vetores que contribuem à combinação linear sejam linearmente independentes. $\square$

1.1 Condições de otimalidade geométricas

Nesta seção, introduzimos uma primeira condição de otimalidade, conhecida como condição de otimalidade geométrica. A idéia é analisar o ângulo de $\nabla f(x^*)$ com direções que apontam aproximadamente para o interior do conjunto viável. Para isso, definiremos um conceito importante no estudo de problemas de otimização, o conceito de cone.

Definição 1.7 (Cone). Um subconjunto $C \subset \mathbb{R}^n$ é um cone, quando, para todo $t \geq 0$ e $d \in C$, tem-se $td \in C$.

Definição 1.8. Seja $C \subset \mathbb{R}^n$ um cone. Denotamos por C° o cone polar de C , dado por

$$C^\circ = \{v \in \mathbb{R}^n : \langle v, d \rangle \leq 0, \forall d \in C\}.$$

Outro objeto geométrico de vital importância é o cone tangente, que será útil para obter relações entre algumas condições de otimalidade e condições de qualificação.

Definição 1.9. Dado $x \in \Omega$, dizemos que um vetor $d \in \mathbb{R}^n$ é uma direção tangente a Ω a partir de x quando for igual ao vetor nulo ou existe uma sequência $\{x^k\} \subset \Omega$ tal que $x^k \neq x$ para todo k ,

$$x^k \rightarrow x \quad \text{e} \quad \frac{x^k - x}{\|x^k - x\|} \rightarrow \frac{d}{\|d\|}.$$

O conjunto formado por todas as direções tangentes é um cone, denominado *cone tangente a Ω em x* e denotado por $T_\Omega(x)$.

Observamos que, de forma equivalente, o cone tangente pode ser definido como

$$T_\Omega(x) = \left\{ d \in \mathbb{R}^n \left| \begin{array}{l} \exists \{t_k\} \subset \mathbb{R}_+, t_k \rightarrow 0^+, \\ \{d^k\} \subset \mathbb{R}^n, d^k \rightarrow d, \text{ tal que } \\ x + t_k d^k \in \Omega, \forall k \in \mathbb{N} \end{array} \right. \right\}.$$

De fato, seja T' o conjunto do lado direito da igualdade acima e considere $0 \neq d \in T'$. Defina a sequência $\{x^k\}$ por $x^k = x + t_k d^k$. Então $x^k \rightarrow x$ e $\{x^k\} \subset \Omega$. Além disso, como $d^k \rightarrow d$, segue que

$$\frac{x^k - x}{\|x^k - x\|} = \frac{t_k d^k}{\|t_k d^k\|} = \frac{d^k}{\|d^k\|} \rightarrow \frac{d}{\|d\|}.$$

Portanto, $d \in T_\Omega(x)$. Para provar a inclusão contrária, basta definir $t_k := \frac{\|x^k - x\|}{\|d\|}$ e $d^k := \frac{\|d\|}{\|x^k - x\|}(x^k - x)$, para todo $k \in \mathbb{N}$.

Teorema 1.10 ([29], TEOREMA 3.12). Uma direção $d \in \mathbb{R}^n$ é uma direção tangente ao conjunto Ω a partir de x se, e somente se,

$$\text{dist}(x + td, \Omega) = o(t), \quad \forall t \geq 0.$$

Teorema 1.11 ([29], LEMA 3.14). O cone tangente é um conjunto fechado.

Lema 1.12. *Dado $C \subset \mathbb{R}^n$, então C° é um cone, convexo e fechado.*

Demonstração. Sejam $d \in C^\circ$ e $t \geq 0$. Logo $\langle d, x \rangle \leq 0$ para todo $x \in C$. Multiplicando por t , obtemos $\langle td, x \rangle \leq 0$ para todo $x \in C$. Portanto, $td \in C^\circ$. Agora vamos mostrar que C° é convexo. Dados $x, y \in C^\circ$ e $t \in [0, 1]$, temos $\langle x, v \rangle \leq 0$ e $\langle y, v \rangle \leq 0$ para todo $v \in C$. Logo,

$$\langle (1-t)x, v \rangle + \langle ty, v \rangle = \langle (1-t)x + ty, v \rangle \leq 0, \quad \forall v \in C.$$

Assim, $(1-t)x + ty \in C^\circ$. Para mostrar que C° é fechado, considere $\{d^k\} \subset C^\circ$ tal que $d^k \rightarrow d$. Logo, para cada $k \in \mathbb{N}$ e $x \in C$, tem-se $\langle d^k, x \rangle \leq 0$. Assim, $\langle d, x \rangle \leq 0$, donde $d \in C^\circ$. $\square$

Lema 1.13. *Se C é um cone convexo fechado não vazio, então $(C^\circ)^\circ = C$.*

Demonstração. Considere qualquer $y \in C^\circ$ e $x \in C$, então $\langle x, y \rangle \leq 0$, o que significa que $x \in (C^\circ)^\circ$. Agora seja $z \in (C^\circ)^\circ$. Como C é convexo fechado, o Lema A.27 garante a existência e unicidade da projeção $\bar{z} = P_C(z)$. O Teorema A.28 implica que

$$\langle z - \bar{z}, x - \bar{z} \rangle \leq 0, \quad \forall x \in C.$$

Tomando $x = 0$ e $x = 2\bar{z}$ na relação anterior, obtemos $\langle z - \bar{z}, \bar{z} \rangle = 0$. Assim, $\langle z - \bar{z}, x \rangle \leq 0$ para todo $x \in C$. Logo $z - \bar{z} \in C^\circ$. Como $z \in (C^\circ)^\circ$, temos que $\langle z - \bar{z}, z \rangle \leq 0$. Portanto,

$$\|z - \bar{z}\|^2 = \langle z - \bar{z}, z - \bar{z} \rangle = \langle z - \bar{z}, z \rangle - \langle z - \bar{z}, \bar{z} \rangle = \langle z - \bar{z}, z \rangle \leq 0,$$

o que implica que $z = \bar{z} \in C$. $\square$

Lema 1.14 ([59], LEMA 7.12). *Sejam $a_1, \dots, a_m$ vetores em $\mathbb{R}^n$. O conjunto*

$$C = \left\{ x \in \mathbb{R}^n : x = \sum_{i=1}^m \mu_i a_i, \mu_i \geq 0, i = 1, \dots, m \right\}$$

é um cone convexo e fechado.

Lema 1.15 (Lema de Farkas). *Dados vetores $a_1, \dots, a_m$ e b em $\mathbb{R}^n$, então exatamente um dos seguintes sistemas têm solução:*

$$\sum_{i=1}^m \beta_i a_i = b, \quad \beta_1, \dots, \beta_m \geq 0; \tag{1.3}$$

$$a_i^T x \leq 0, \quad \forall i \in \{1, \dots, m\}, \quad \text{e } b^T x > 0, \quad x \in \mathbb{R}^n. \tag{1.4}$$

Demonstração. Suponha que o sistema (1.3) tenha solução. Então $\sum_{i=1}^m \beta_i a_i = b$ com $\beta_i \geq 0$ para todo $i = 1, \dots, m$. Logo, dado $x \in \mathbb{R}^n$, tal que $a_i^T x \leq 0$, temos $b^T x = \sum_{i=1}^m \beta_i (a_i^T x) \leq 0$ e, portanto, o sistema (1.4) não tem solução. Agora suponha que (1.3) não tem solução. Considere

$$C = \left\{ x \in \mathbb{R}^n : x = \sum_{i=1}^m \beta_i a_i, \beta_i \geq 0, i = 1, \dots, m \right\}.$$

Logo $b \notin C$. Em virtude do Lema 1.14, temos que C é convexo e fechado. Usando o Lema 1.13, segue-se que $C = (C^\circ)^\circ$. Logo existe $x \in C^\circ$ tal que $b^T x > 0$. Além disso, temos

$$\sum_{i=1}^m \beta_i a_i^T x \leq 0, \quad \beta_i \geq 0, \quad i = 1, \dots, m.$$

Tomando $(\beta_1, \dots, \beta_m) = e_j$, $j = 1, \dots, m$, obtemos $a_j^T x \leq 0$, completando a prova. $\square$

Uma vez que calcular o cone tangente explicitamente pode ser muito difícil, gostaríamos de definir um cone que contenha as informações de primeira ordem do problema e que seja mais simples de ser simulado numericamente.

Definição 1.16. O cone linearizado de Ω em x é definido como

$$L_{\Omega}(x) := \{d \in \mathbb{R}^n : \nabla h_i(x)^T d = 0, i = 1, \dots, m; \nabla g_j(x)^T d \leq 0, j \in A(x)\}.$$

$L_{\Omega}(x)$ é às vezes chamado de “cone de variações de primeira ordem” ou cone de direções viáveis. Note que $L_{\Omega}(x)$ pode ser visto como um conjunto viável, onde linearizamos as restrições de igualdade e as de desigualdade ativas. Quando as restrições do problema são lineares, o cone linearizado coincide com o cone tangente.

Lema 1.17. Dado $x \in \Omega$, considere o conjunto

$$S(x) := \left\{ v \in \mathbb{R}^n : v = \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x)} \mu_j \nabla g_j(x), \mu_j \geq 0, j \in A(x), \lambda_i \in \mathbb{R} \right\}.$$

Então $S(x)$ é um cone convexo fechado e $L_{\Omega}(x)^{\circ} = S(x)$.

Demonstração. Primeiro, mostraremos que $L_{\Omega}(x) = S(x)^{\circ}$. Sejam $d \in L_{\Omega}(x)$ e $v \in S(x)$. Logo, $\nabla h_i(x)^T d = 0$ para todo $i = 1, \dots, m$ e $\nabla g_j(x)^T d \leq 0$ para todo $j \in A(x)$. Assim,

$$d^T v = \sum_{i=1}^m \lambda_i \nabla h_i(x)^T d + \sum_{j \in A(x)} \mu_j \nabla g_j(x)^T d \leq 0.$$

Portanto, $d \in S(x)^{\circ}$. Para provar a inclusão contrária, considere $d \in S(x)^{\circ}$. Então $d^T v \leq 0$ para todo $v \in S(x)$. Como $\pm \nabla h_i(x) \in S(x)$ para todo $i = 1, \dots, m$, temos que

$$\nabla h_i(x)^T d \leq 0 \quad \text{e} \quad (-\nabla h_i(x))^T d \leq 0.$$

Assim, $\nabla h_i(x)^T d = 0$ para todo $i = 1, \dots, m$. Além disso, $\nabla g_j(x) \in S(x)$ para todo $j \in A(x)$ e, portanto, $\nabla g_j(x)^T d \leq 0$, concluindo que $d \in L_{\Omega}(x)$.

O Lema 1.14 garante que $S(x)$ um cone é convexo fechado, basta ver que qualquer elemento de $S(x)$ pode ser escrito como

$$\sum_{i: \lambda_i \geq 0} \lambda_i \nabla h_i(x) + \sum_{i: \lambda_i < 0} (-\lambda_i) (-\nabla h_i(x)) + \sum_{j \in A(x)} \mu_j \nabla g_j(x).$$

Logo, pelo Lema 1.13, temos

$$L_{\Omega}(x)^{\circ} = (S(x)^{\circ})^{\circ} = S(x),$$

completando a demonstração. $\square$

Lema 1.18. Se $x^* \in \Omega$ é um minimizador local do problema (1.2), então para todo $d \in T_{\Omega}(x^*)$ tem-se $\nabla f(x^*)^T d \geq 0$.

Demonstração. Seja $d \in T_{\Omega}(x^*)$, logo existem sequências $\{t_k\} \subset \mathbb{R}_+$ e $\{d^k\}$ tais que $t_k \rightarrow 0$, $d^k \rightarrow d$ e $x^* + t_k d^k \in \Omega$ para todo $k \in \mathbb{N}$. Como $x^* + t_k d^k \rightarrow x^*$ e x^* é minimizador local, temos que $f(x^*) \leq f(x^* + t_k d^k)$ para k suficientemente grande. Aplicando expansão de Taylor, temos

$$f(x^*) \leq f(x^* + t_k d^k) = f(x^*) + t_k \nabla f(x^*)^T d^k + o(t_k)$$

onde $o(t_k)/t_k \rightarrow 0$. Assim, $t_k \nabla f(x^*)^T d^k + o(t_k) \geq 0$. Dividindo ambos os lados da desigualdade por t_k e tomando o limite, obtemos $\nabla f(x^*)^T d \geq 0$ para todo $d \in T_{\Omega}(x^*)$. $\square$

Como consequência do Lema acima temos o seguinte resultado:

Teorema 1.19 (Condição de otimalidade geométrica). *Se $x^* \in \Omega$ é um minimizador local do problema (1.2), então $-\nabla f(x^*) \in T_\Omega(x^*)^\circ$.*

Lema 1.20. *Dado $x^* \in \Omega$, então vale $T_\Omega(x^*) \subset L_\Omega(x^*)$.*

Demonstração. Seja $d \in T_\Omega(x^*)$, $d \neq 0$, por definição, existe uma sequência $\{x^k\} \subset \Omega$ tal que $x^k \rightarrow x^*$ e $\frac{x^k - x^*}{\|x^k - x^*\|} \rightarrow \frac{d}{\|d\|}$. Pela diferenciabilidade de g e h , segue que

$$\begin{aligned} h_i(x^k) &= h_i(x^*) + \nabla h_i(x^*)^T (x^k - x^*) + o(\|x^k - x^*\|), \quad i = 1, \dots, m, \\ g_j(x^k) &= g_j(x^*) + \nabla g_j(x^*)^T (x^k - x^*) + o(\|x^k - x^*\|), \quad j \in A(x^*), \end{aligned}$$

onde $\lim_{k \rightarrow \infty} \frac{o(\|x^k - x^*\|)}{\|x^k - x^*\|} = 0$.

Como $x^k, x^* \in \Omega$ para todo $k \in \mathbb{N}$, temos

$$\begin{aligned} \nabla h_i(x^*)^T \frac{(x^k - x^*)}{\|x^k - x^*\|} + \frac{o(\|x^k - x^*\|)}{\|x^k - x^*\|} &= 0, \quad i = 1, \dots, m \\ \nabla g_j(x^*)^T \frac{(x^k - x^*)}{\|x^k - x^*\|} + \frac{o(\|x^k - x^*\|)}{\|x^k - x^*\|} &\leq 0, \quad j \in A(x^*). \end{aligned}$$

Tomando o limite, obtemos $\nabla h_i(x^*)^T \frac{d}{\|d\|} = 0$ e $\nabla g_j(x^*)^T \frac{d}{\|d\|} \leq 0$. Portanto d pertence a $L_\Omega(x^*)$. $\square$

1.1.1 Geometria variacional

Nesta subseção, faremos uma breve abordagem à geometria variacional [61], através de objetos geométricos de fundamental importância em otimização e em análise variacional. Tal abordagem pode ser elaborada associando cada ponto $x \in \Omega$ a certos cones de vetores tangentes e vetores normais, que generalizam os subespaços tangente e normal da geometria diferencial clássica, o que nos motiva a introduzir o conceito de multifunção.

Definição 1.21. Um mapa $F : \mathbb{R}^s \rightrightarrows \mathbb{R}^d$ é uma *multifunção* se para cada $x \in \mathbb{R}^s$, $F(x)$ é um subconjunto de $\mathbb{R}^d$. Definimos o domínio, a imagem e o gráfico de F por

$$\begin{aligned} \text{dom } F &:= \{x \in \mathbb{R}^s : F(x) \neq \emptyset\}, \quad \text{range } F := \{y \in \mathbb{R}^d : \exists x \in \mathbb{R}^s \text{ com } y \in F(x)\}, \\ \text{graf } F &:= \{(x, y) : y \in F(x)\}, \end{aligned}$$

respectivamente.

Os cones tangente, normal e linearizado são exemplos de multifunções.

Definição 1.22. Dada uma multifunção $F : \mathbb{R}^s \rightrightarrows \mathbb{R}^d$, o *limite exterior sequencial de Painlevé-Kuratowski* de $F(x)$ quando $x \rightarrow x^*$ é denotado por

$$\limsup_{x \rightarrow x^*} F(x) := \{y \in \mathbb{R}^d : \exists (x^k, y^k) \rightarrow (x^*, y) \text{ tal que } y^k \in F(x^k), \forall k \in \mathbb{N}\}.$$

e o limite interior por

$$\liminf_{x \rightarrow x^*} F(x) := \{y \in \mathbb{R}^d : \forall x^k \rightarrow x^*, \exists y^k \rightarrow y \text{ tal que } y^k \in F(x^k), \forall k \in \mathbb{N}\}.$$

Definição 1.23. Dizemos que F é *semi-contínua exteriormente* em x^* quando

$$\limsup_{x \rightarrow x^*} F(x) \subset F(x^*)$$

e F é *semi-contínua interiormente* em x^* quando

$$\liminf_{x \rightarrow x^*} F(x) \supset F(x^*).$$

Dizemos que F é contínua em x^* se F for simultaneamente semi-contínua exteriormente e semi-contínua interiormente em x^* .

Observação 1.24. Dado $x \in \Omega$, o cone tangente a Ω em x pode ser definido como

$$T_\Omega(x) = \limsup_{t \rightarrow 0^+} \frac{\Omega - x}{t}.$$

De fato, dado $d \in \limsup_{t \rightarrow 0^+} \frac{\Omega - x}{t}$, pela definição de limite exterior existe $\{t_k\} \subset \mathbb{R}_+$, $t_k \rightarrow 0^+$, $\{d^k\} \subset \mathbb{R}^n$, $d^k \rightarrow d$ tais que $d^k \in (\Omega - x)/t_k$ para todo $k \in \mathbb{N}$. Logo, para cada k existe $w_k \in \Omega$ tal que $d^k = \frac{w_k - x}{t_k}$. Assim, $x + t_k d^k = w_k \in \Omega$, ou seja, $d \in T_\Omega(x)$.

Para mostrar a outra inclusão, basta seguir os mesmos passos.

Definição 1.25. Seja $x^* \in \Omega$. O *cone normal regular* de Ω em x^* é o conjunto

$$\widehat{N}_\Omega(x^*) := \{d \in \mathbb{R}^n : \langle d, x - x^* \rangle \leq o(\|x - x^*\|), x \in \Omega\}.$$

E o *cone normal limite* de Ω em x^* é dado por

$$N_\Omega(x^*) := \limsup_{x \rightarrow x^*, x \in \Omega} \widehat{N}_\Omega(x).$$

Quando Ω é um conjunto convexo, o cone regular normal e cone normal limite coincidem com a noção clássica de cone normal usado na análise convexa que definimos no Apêndice A (Definição A.39). Para todo $x \in \Omega$, tem-se $\widehat{N}_\Omega(x) \subset N_\Omega(x)$. De fato, sejam $d \in \widehat{N}_\Omega(x^*)$ e $x^* \in \Omega$. Pela definição, segue que $\langle d, x - x^* \rangle \leq o(\|x - x^*\|)$ para todo $x \in \Omega$. Defina $d^k = d$ e $x^k = x^*$ para todo $k \in \mathbb{N}$. Obviamente $d^k \rightarrow d$, $x^k \rightarrow x^*$ e $d^k \in \widehat{N}_\Omega(x^k)$. Portanto $d \in N_\Omega(x^*)$.

Teorema 1.26. Se $x^* \in \Omega$ é minimizador local do problema (1.2), então $-\nabla f(x^*) \in \widehat{N}_\Omega(x^*)$.

Quando Ω é um conjunto convexo, essa condição também pode ser escrita na forma

$$\langle -\nabla f(x^*), x - x^* \rangle \leq 0 \text{ para todo } x \in \Omega.$$

Demonstração. Como x^* é minimizador local, então existe uma vizinhança V de x^* tal que $f(x^*) \leq f(x)$ para todo $x \in V$, ou seja, $f(x) - f(x^*) \geq 0$. Em virtude da diferenciabilidade de f , temos

$$0 \leq f(x) - f(x^*) = \langle \nabla f(x^*), x - x^* \rangle + o(\|x - x^*\|).$$

Logo, $\langle -\nabla f(x^*), x - x^* \rangle \leq o(\|x - x^*\|)$ para todo $x \in \Omega$. Portanto, concluímos que $-\nabla f(x^*) \in \widehat{N}_\Omega(x^*)$. $\square$

A próxima proposição estabelece uma relação entre os cones $\widehat{N}_\Omega(x)$ e $T_\Omega(x)$.

Proposição 1.27. *Sejam $\Omega \subset \mathbb{R}^n$ e $x^* \in \Omega$, então*

$$d \in \widehat{N}_\Omega(x^*) \Leftrightarrow \langle d, w \rangle \leq 0, \text{ para todo } w \in T_\Omega(x^*).$$

Demonstração. $(\Rightarrow)$ Considere $d \in \widehat{N}_\Omega(x^*)$ e $w \in T_\Omega(x^*)$. Pela definição de cone tangente, existem sequências $\{x^k\} \subset \Omega$, $x^k \rightarrow x^*$ e $t_k \rightarrow 0^+$ tal que $w^k = \frac{x^k - x^*}{t_k} \rightarrow w$. Como $d \in \widehat{N}_\Omega(x^*)$, então temos $\langle d, w^k \rangle \leq \frac{o(\|t_k w^k\|)}{t_k} \rightarrow 0$. Portanto, $\langle d, w \rangle \leq 0$.

$(\Leftarrow)$ Suponha que $d \notin \widehat{N}_\Omega(x^*)$. Então $\langle d, x - x^* \rangle > o(\|x - x^*\|)$ para algum $x \in \Omega$. Logo, existe uma sequência $x^k \rightarrow x^*$ com $x^k \neq x^*$ tal que

$$\liminf_{k \rightarrow \infty} \frac{\langle d, x^k - x^* \rangle}{\|x^k - x^*\|} > 0.$$

Defina $w^k = \frac{x^k - x^*}{\|x^k - x^*\|}$. Logo, $\liminf_{k \rightarrow \infty} \langle d, w^k \rangle > 0$ com $\|w^k\| = 1$. Passando a uma subsequência se necessário, podemos supor que $w^k \rightarrow w$. Note que $w \in T_\Omega(x^*)$, pois $\frac{x^k - x^*}{t_k} \rightarrow w$ com $t_k = \|x^k - x^*\| \rightarrow 0^+$. Portanto, $\langle d, w \rangle > 0$ com $w \in T_\Omega(x^*)$, o que contradiz a hipótese. $\square$

Em outras palavras, essa proposição afirma que $\widehat{N}_\Omega(x^*) = T_\Omega(x^*)^\circ$.

Lema 1.28. *Sejam $\Omega \subset \mathbb{R}^n$ convexo e $x^* \in \Omega$. Então $\widehat{N}_\Omega(x^*) = N_\Omega(x^*)$.*

Demonstração. Basta provar que $N_\Omega(x^*) \subset \widehat{N}_\Omega(x^*)$. Seja $d \in N_\Omega(x^*)$ e considere $x \in \Omega$. Pela definição de limite exterior existem sequências $x^k \rightarrow x^*$ e $d^k \rightarrow d$ com $d^k \in \widehat{N}_\Omega(x^k)$. Como Ω é convexo, então o cone $\widehat{N}_\Omega(x^*)$ coincide com o cone da Definição A.39 (Apêndice A). Logo, $\langle d^k, x - x^k \rangle \leq 0$ para todo $k \in \mathbb{N}$. Passando ao limite, obtemos $\langle d, x - x^* \rangle \leq 0$. Portanto $d \in \widehat{N}_\Omega(x^*)$. $\square$

Teorema 1.29 ([61], TEOREMA 6.28). *Considere $\widehat{T}_\Omega(x^*) := \liminf_{t \rightarrow 0^+} \frac{\Omega - x^*}{t}$. Então, $\widehat{T}_\Omega(x^*) = N_\Omega(x^*)^\circ$ para Ω localmente fechado em x^* .*

2 Condições de Otimalidade de Primeira Ordem

As condições de otimalidade de primeira ordem usuais são da forma “KKT ou não-CQ”, onde KKT é o Teorema de Karush-Kuhn-Tucker e CQ é uma condição de qualificação. Em outras palavras, um minimizador satisfaz as condições de KKT ou não satisfaz uma condição de qualificação. Portanto, o natural é estudar condições de qualificação mais fracas para obter condições de otimalidade mais fortes. O Teorema de Karush-Kuhn-Tucker que veremos a seguir, apesar de não ser uma condição de otimalidade é uma ferramenta útil para encontrar possíveis candidatos à solução quando o problema cumpre alguma condição de qualificação. É uma generalização do método dos multiplicadores de Lagrange que conhecemos do cálculo.

Teorema 2.1 (Karush-Kuhn-Tucker, [43, 44]). *Seja $x \in \Omega$ um minimizador local do problema (1.2) e suponha que $T_\Omega(x)^\circ = L_\Omega(x)^\circ$, então existem $\lambda^* \in \mathbb{R}^m$ e $\mu^* \in \mathbb{R}^p$ tais que*

$$\begin{aligned} \nabla f(x) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x) + \sum_{j=1}^p \mu_j^* \nabla g_j(x) &= 0, \\ \mu_j^* &\geq 0, \quad j = 1, \dots, p, \\ \mu_j^* g_j(x) &= 0, \quad j = 1, \dots, p. \end{aligned} \tag{2.1}$$

Demonstração. Em virtude do Teorema 1.19, temos $-\nabla f(x) \in T_\Omega(x)^\circ$. Assim, pela hipótese, obtemos

$$-\nabla f(x) \in T_\Omega(x)^\circ = L_\Omega(x)^\circ.$$

Usando o Lema 1.17, segue que existem $\lambda \in \mathbb{R}^m$ e $\mu_j \geq 0$, $j \in A(x)$, tais que

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

Pondo $\lambda^* = \lambda$, $\mu_j^* = \mu_j$ para $j \in A(x)$ e $\mu_j^* = 0$ para $j \notin A(x)$, concluímos a demonstração. $\square$

Os vetores λ^* e μ^* são denominados multiplicadores de Lagrange. Denotamos por $\Lambda(x)$ o conjunto de multiplicadores de Lagrange associados a x . Os pontos $x \in \Omega$ que cumprem (2.1) são chamados *pontos estacionários* ou pontos KKT do problema. Note que as condições de KKT significa dizer que $-\nabla f(x^*) \in L_\Omega(x^*)^\circ$.

A hipótese sobre os cones $T_\Omega(x)$ e $L_\Omega(x)$ é chamada de *condição de qualificação de Guignard*, introduzida por Monique Guignard [36] para dimensão infinita e reformulada para o caso finito por Gould e Tolle [35]. Essa condição é a condição de qualificação

mais fraca possível, no sentido de que, fixado $x^* \in \Omega$, se para qualquer função objetivo f tal que x^* é solução do problema de minimizar $f(x)$ sujeito a $x \in \Omega$, vale que x^* é um ponto KKT, então a condição de Guignard é satisfeita em x^* . Em outras palavras, a condição de Guignard é necessária e suficiente para termos KKT. Entretanto, pode ser muito difícil obter os cones $T_\Omega(x)$ e $L_\Omega(x)$ e verificar se vale a condição $T_\Omega(x)^\circ = L_\Omega(x)^\circ$. Na próxima seção, veremos que existem outras condições de qualificação mais simples de serem verificadas e que implicam em Guignard, garantindo que um minimizador satisfaz KKT. É importante salientar que o Teorema 2.1 não é suficiente para garantir a otimalidade, podendo existir pontos KKT que não são soluções do problema (1.2). O exemplo a seguir mostra que podem existir minimizadores que não cumprem as condições de KKT, mostrando a importância de uma condição de qualificação.

Exemplo 2.2. Considere o problema de otimização:

$$\begin{aligned} \text{minimizar} \quad & f(x) = x_1 \\ \text{sujeito a} \quad & g_1(x) = -x_1^3 + x_2 \leq 0 \\ & g_2(x) = -x_2 \leq 0. \end{aligned}$$

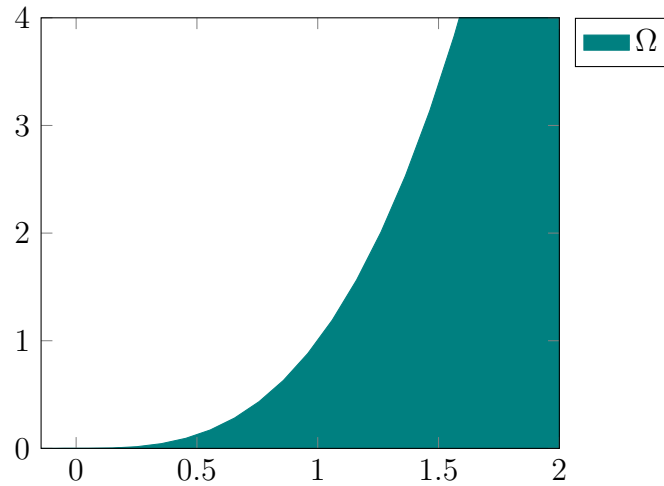


Figura 2.1: As condições KKT falham em um minimizador

Conforme a Figura 2.1, podemos ver que o ponto $x^* = (0,0)$ é o minimizador deste problema. De fato, de $0 \leq x_2 \leq x_1^3$, segue que $f(x) = x_1 \geq 0 = f(x^*)$ para todo x viável. Por outro lado, calculando os gradientes, segue que

$$\nabla f(x^*) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \nabla g_1(x^*) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{e} \quad \nabla g_2(x^*) = \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

o que significa que não vale KKT em x^* .

Quando f e g_j , $j = 1, \dots, p$, são funções convexas e h é afim, ou seja, Ω é um conjunto convexo, as condições de KKT são suficientes para a minimização global. De fato, suponha que $x^* \in \Omega$ satisfaz KKT. Logo, pelo Teorema 1.19 e pela Proposição 1.27, segue que $-\nabla f(x^*) \in L_\Omega(x^*)^\circ = T_\Omega(x^*)^\circ = N_\Omega(x^*)$ e, portanto, $\langle -\nabla f(x^*), x - x^* \rangle \leq 0$ para todo $x \in \Omega$. Pelo Corolário A.36, segue que x^* é minimizador global de f em Ω .

Quando as restrições são lineares, as condições de KKT sempre valem em um minimizador, como mostra o próximo teorema.

Teorema 2.3. *Seja x^* um minimizador local do problema*

$$\begin{aligned} & \text{minimizar} && f(x), \\ & \text{sujeito a} && Ax = b, \\ & && Mx \leq c, \end{aligned}$$

onde $A \in \mathbb{R}^{m \times n}$, $M \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^m$ e $c \in \mathbb{R}^p$, então x^* satisfaz as condições KKT.

Demonstração. Vamos provar que vale a condição de qualificação de Abadie. Em virtude do Lema 1.20, basta provar que $L_\Omega(x^*) \subset T_\Omega(x^*)$. Seja $d \in L_\Omega(x^*)$, então $Ad = 0$ e $Md \leq 0$. Se $d = 0$, o resultado segue trivialmente. Suponha $d \neq 0$ e defina $x^k = x^* + \frac{1}{k}d$. Logo,

$$Ax^k = b, \quad Mx^k \leq c, \quad x^k \rightarrow x^* \quad \text{e} \quad \frac{x^k - x^*}{\|x^k - x^*\|} \rightarrow \frac{d}{\|d\|}.$$

Portanto, $d \in T_\Omega(x^*)$ e, como consequência do Teorema 2.1, concluímos que x^* satisfaz KKT. $\square$

O próximo Teorema fornece uma condição de otimalidade conhecida como *Condição de Otimalidade de Fritz John*, que generaliza o Teorema de Karush-Kuhn-Tucker, e foi proposta pela primeira vez em 1948 por John [42]. A condição de Fritz John, além de ser uma condição de otimalidade simples de ser verificada, não depende de nenhuma hipótese adicional sobre as restrições. Ela introduz um multiplicador auxiliar não-negativo ao gradiente da função objetivo.

Teorema 2.4 (Condição necessária de Fritz John). *Se x^* é um minimizador local, então existe um vetor $(\lambda_0, \lambda, \mu) \in \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}_+^p$ satisfazendo as condições:*

$$(i) \quad \lambda_0 \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) = 0, \quad \lambda_0 \geq 0;$$

$$(ii) \quad (\lambda_0, \lambda, \mu) \neq 0;$$

$$(iii) \quad \text{para cada vizinhança } V \text{ de } x^* \text{ existe } x \in V \text{ tal que } \lambda_i h_i(x) > 0 \text{ para todo } i \text{ com } \lambda_i \neq 0 \text{ e } \mu_j g_j(x) > 0 \text{ para todo } j \text{ com } \mu_j > 0.$$

Demonstração. A demonstração apresentada a seguir também pode ser encontrada em Bertsekas [21]. Considere as funções $g_j^+(x) = \max\{0, g_j(x)\}$ para $j = 1, \dots, p$ e uma sequência $\{\rho_k\} \subset \mathbb{R}_+$ com $\rho_k \rightarrow +\infty$. Para cada k , seja x^k uma solução global problema penalizado (ver Seção 2.2)

$$\begin{aligned} & \text{minimizar} && f(x) + \frac{1}{2}\|x - x^*\|^2 + \frac{\rho_k}{2} \left(\sum_{i=1}^m h_i(x)^2 + \sum_{j=1}^p g_j^+(x)^2 \right) \\ & \text{sujeito a} && x \in X, \end{aligned}$$

onde $X = \{x \in \mathbb{R}^n : \|x - x^*\| \leq \delta\}$ e $\delta > 0$ é dado pela definição de minimizador local. Sendo X um conjunto compacto, a sequência $\{x^k\}$ gerada pelo algoritmo de penalidade externa está bem definida. Como consequência do Teorema 2.38 todos seus pontos limites são soluções globais. Como x^* é a única solução do problema, segue que $x^k \rightarrow x^*$. Logo,

$\|x^k - x^*\| < \delta$ para todo k suficientemente grande, ou seja, x^k é ponto interior de X . Assim, o gradiente da função objetivo deve ser zero em $x = x^k$, isto é,

$$\nabla f(x^k) + \sum_{i=1}^m \gamma_i^k \nabla h_i(x^k) + \sum_{j=1}^p \zeta_j^k \nabla g_j(x^k) + (x^k - x^*) = 0, \quad (2.2)$$

$$\text{onde } \gamma_i^k := \rho_k h_i(x^k) \quad \text{e} \quad \zeta_j^k := \rho_k g_j^+(x^k). \quad (2.3)$$

Denote

$$\delta^k := \sqrt{1 + \sum_{i=1}^m (\gamma_i^k)^2 + \sum_{j=1}^p (\zeta_j^k)^2}, \quad (2.4)$$

$$\lambda_0^k := \frac{1}{\delta^k}, \quad \lambda_i^k := \frac{\gamma_i^k}{\delta^k}, \quad i = 1, \dots, m, \quad \mu_j^k := \frac{\zeta_j^k}{\delta^k}, \quad j = 1, \dots, p. \quad (2.5)$$

Dividindo a equação (2.2) por δ^k , obtemos

$$\lambda_0^k \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) + \frac{1}{\delta^k} (x^k - x^*) = 0. \quad (2.6)$$

Por construção, temos que

$$(\lambda_0^k)^2 + \sum_{i=1}^m (\lambda_i^k)^2 + \sum_{j=1}^p (\mu_j^k)^2 = 1. \quad (2.7)$$

Logo, a sequência $\{(\lambda_0^k, \lambda^k, \mu^k)\} \subset \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}_+^p$ é limitada e, portanto, existe uma subsequência convergente, digamos $(\lambda_0^k, \lambda^k, \mu^k) \xrightarrow{\mathbb{N}'} (\lambda_0, \lambda, \mu)$. Passando ao limite em (2.6), e usando a continuidade dos gradientes, obtemos

$$\lambda_0 \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) = 0, \quad \lambda_0 \geq 0,$$

provando a primeira condição. A segunda condição segue claramente de (2.7). Para provar a terceira condição, considere $I = \{i : \lambda_i \neq 0\}$ e $J = \{j : \mu_j > 0\}$. Assim, para todo $k \in \mathbb{N}'$ suficientemente grande, temos que $\lambda_i \lambda_i^k > 0$ para todo $i \in I$ e $\mu_j \mu_j^k > 0$ para todo $j \in J$. Logo, usando as relações (2.3) e (2.5), segue que para todo $k \in \mathbb{N}'$, $\lambda_i h_i(x^k) > 0$ para todo $i \in I$ e $\mu_j g_j(x^k) > 0$ para todo $j \in J$, o que prova a última condição. $\square$

Os vetores λ_0 , λ e μ são chamados de multiplicadores de Fritz John. A condição (iii) implica que $\mu_j g_j(x^*) = 0$ para todo j , pois se $\mu_j > 0$, a j -ésima restrição de desigualdade pode ser violada arbitrariamente perto de x^* , o que implica que $g_j(x^*) = 0$. Esta condição é tradicionalmente conhecida como *condição de complementaridade*.

A condição de Fritz John é equivalente às condições de KKT no caso $\lambda_0 \neq 0$ ($\lambda_0 = 1$). Quando $\lambda_0 = 0$, essa condição é equivalente à violação da condição de Mangasarian-Fromovitz (MFCQ) que veremos na próxima seção. Em outras palavras, a condição de Fritz John é equivalente à condição “KKT ou não-MFCQ”.

O exemplo a seguir mostra que a condição de Fritz John não é suficiente para garantir a otimalidade, mesmo no caso linear.

Exemplo 2.5. Considere o seguinte problema

$$\begin{aligned} &\text{minimizar} && f(x) = -x_1 \\ &\text{sujeito a} && g_1(x) = x_1 + x_2 \leq 0 \\ &&& g_2(x) = -x_1 - x_2 \leq 0 \\ &&& g_3(x) = -x_2 \leq 0. \end{aligned}$$

Neste caso, o conjunto viável são os pontos que estão sobre a reta $x_1 + x_2 = 0$ com $x_2 \geq 0$. Assim, a única solução é o ponto $x^* = (0, 0)$. Além disso, todo ponto viável satisfaz a condição de Fritz John, basta definir $\lambda_0 = \mu_3 = 0$ e $\mu_1 = \mu_2 = \alpha > 0$.

2.1 Condições de Qualificação

Na seção anterior, vimos que podem existir minimizadores que não cumprem as relações de KKT, ou seja, KKT não é uma condição de otimalidade. Para garantir a validade das condições de KKT num ponto, é necessário alguma hipótese adicional sobre as restrições do problema, as chamadas *condições de qualificação*.

Definição 2.6. Uma *condição de qualificação (CQ)* é uma proposição acerca das restrições, tal que sob essa proposição, todo minimizador local é um ponto estacionário.

Embora KKT não seja uma condição de otimalidade, podemos associar condições de otimalidade às condições KKT. Note que toda condição de qualificação define uma condição de otimalidade da forma “KKT ou não-CQ”. De fato, a proposição “ $\mathcal{P} \wedge \mathcal{Q} \Rightarrow \mathcal{R}$ ” é equivalente a “ $\mathcal{P} \Rightarrow \mathcal{R} \vee \neg \mathcal{Q}$ ”. Logo, quanto mais fraca for a condição de qualificação, mais forte será a condição de otimalidade associada. De fato, sejam $\mathcal{P}$ e $\mathcal{Q}$ duas condições de qualificação tais que $\mathcal{P} \Rightarrow \mathcal{Q}$. Assim, $\neg \mathcal{Q} \Rightarrow \neg \mathcal{P}$ e, portanto, a condição de otimalidade “ $\text{KKT} \vee \neg \mathcal{Q}$ ” é mais forte do que a condição de otimalidade “ $\text{KKT} \vee \neg \mathcal{P}$ ”. Logo, condições de qualificação mais fracas implicam em condições de otimalidade mais fortes.

A condição $T_\Omega(x)^\circ = L_\Omega(x)^\circ$ dada no Teorema 2.1, conhecida como condição de qualificação de Guignard, pode ser muito difícil de ser verificada na prática. Não existe um algoritmo capaz de gerar pontos que satisfazem a condição “KKT ou não-CQ”, onde CQ é a condição de Guignard. Assim, o objetivo é estudar condições de qualificação mais fortes que Guignard e que permitam a prova da convergência global de algoritmos. Outra condição envolvendo os cones tangente e linearizado, mais exigente que Guignard, é a condição de Abadie [1].

Definição 2.7. Dizemos que $x \in \Omega$ satisfaz a *condição de Abadie* se $T_\Omega(x) = L_\Omega(x)$.

Claramente, temos $T_\Omega(x) = L_\Omega(x) \Rightarrow T_\Omega(x)^\circ = L_\Omega(x)^\circ$ e, portanto, a condição de Abadie é uma condição de qualificação mais forte que Guignard. Entretanto, estas condições, além da dificuldade na verificação, não são suficientes para a convergência global de algoritmos. Veremos agora outras condições de qualificação, mais simples de serem verificadas, que também garantem que um minimizador satisfaz as condições KKT.

No Exemplo 2.2 vimos que as condições de KKT falham em um minimizador. Além disso, neste exemplo, os gradientes das restrições de igualdade e desigualdade ativas em x^* são linearmente dependentes. Logo, é natural conjecturar que se o conjunto $\{\nabla h_i(x^*), \nabla g_j(x^*) : i = 1, \dots, m, j \in A(x^*)\}$ é linearmente independente, então as condições KKT valem em x^* . De fato, veremos que essa propriedade é uma condição de qualificação, conhecida como *regularidade* ou *Condição de Independência Linear*.

Definição 2.8. Diremos que um ponto $x \in \Omega$ é *regular*, quando os gradientes das restrições de igualdade e desigualdade ativas em x são linearmente independentes. Ou equivalentemente, diremos que a *condição de independência linear (LICQ)* vale no ponto viável x , se o conjunto $\{\nabla h_i(x), \nabla g_j(x) : i = 1, \dots, m, j \in A(x)\}$ for linearmente independente.

Um fato importante sob a regularidade é a existência e unicidade dos multiplicadores de Lagrange. De fato, para a unicidade, suponha que existem multiplicadores (λ, μ) e $(\tilde{\lambda}, \tilde{\mu})$ associado ao problema (1.2). Logo, temos

$$\begin{aligned} \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x)} \mu_j \nabla g_j(x) &= 0 \quad \text{e} \\ \nabla f(x) + \sum_{i=1}^m \tilde{\lambda}_i \nabla h_i(x) + \sum_{j \in A(x)} \tilde{\mu}_j \nabla g_j(x) &= 0. \end{aligned}$$

Subtraindo as expressões, obtemos:

$$\sum_{i=1}^m (\tilde{\lambda}_i - \lambda_i) \nabla h_i(x) + \sum_{j \in A(x)} (\tilde{\mu}_j - \mu_j) \nabla g_j(x) = 0.$$

Como LICQ vale em x , temos que os gradientes $\{\nabla h_i(x)\}_{i=1}^m \cup \{\nabla g_j(x)\}_{j \in A(x)}$ são linearmente independentes. Logo, $\tilde{\lambda}_i - \lambda_i = 0$ para todo $i = 1, \dots, m$ e $\tilde{\mu}_j - \mu_j = 0$ para todo $j \in A(x)$. Portanto, $\lambda = \tilde{\lambda}$ e $\mu = \tilde{\mu}$.

Para provar a existência, seja $x^* \in \Omega$ um minimizador local e assuma que LICQ seja satisfeita em x^* . Logo, pelo Teorema 2.4 existe um multiplicador de Fritz John $0 \neq (\lambda_0, \lambda, \mu) \in \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}_+^p$ com $\lambda_0 \geq 0$ e $\mu_j = 0$ para $j \notin A(x^*)$ tal que

$$\lambda_0 \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*) = 0. \quad (2.8)$$

Se $\lambda_0 = 0$, obtemos

$$\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*) = 0,$$

com $(\lambda, \mu) \neq 0$, contradizendo a definição de LICQ. Logo, $\lambda_0 \neq 0$. Dividindo a equação (2.8) por λ_0 , obtemos

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j^* \nabla g_j(x^*) = 0,$$

onde $\lambda^* = \frac{\lambda}{\lambda_0}$ e $\mu^* = \frac{\mu}{\lambda_0}$. Portanto vale KKT em x^* . Desse modo, podemos concluir que LICQ é de fato uma condição de qualificação.

Proposição 2.9. Se $x \in \Omega$ satisfaz LICQ, então existe uma vizinhança V de x tal que LICQ é satisfeita em todo ponto $y \in V$.

Demonstração. Suponha que não exista uma vizinhança V de x tal que LICQ é satisfeita em todo $y \in V$. Então existem sequências $\{y^k\} \subset \mathbb{R}^n$ com $y^k \rightarrow x$ e $\|(\lambda^k, \mu^k)\| = 1$ tais que

$$\sum_{i=1}^m \lambda_i^k \nabla h_i(y^k) + \sum_{j \in A(x)} \mu_j^k \nabla g_j(y^k) = 0.$$

Sem perda de generalidade, suponha que $\lambda^k \rightarrow \lambda$ e $\mu^k \rightarrow \mu$. Portanto,

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

com $\|(\lambda, \mu)\| = 1$, o que contradiz a definição de LICQ. $\square$

Embora LICQ seja uma condição simples de ser verificada na prática, muitos problemas de programação matemática não podem ser analisados sob essa condição. Além disso, LICQ não permite redundância na descrição de Ω , por exemplo, escrever uma restrição de igualdade como duas de desigualdade ou quando os gradientes de duas restrições são coincidentes. Porém, outras condições de qualificação mais fracas que LICQ podem ser encontradas na literatura, como por exemplo a condição de Mangasarian-Fromovitz (MFCQ) [49] e a condição de posto contante (CRCQ) [41]. Antes de definir a condição de Mangasarian-Fromovitz precisamos da noção de dependência linear positiva.

Definição 2.10. Dados vetores $u_1, \dots, u_{m+p}$ em $\mathbb{R}^n$, não necessariamente distintos, dizemos que $(\{u_1, \dots, u_m\}, \{u_{m+1}, \dots, u_{m+p}\})$ é *positivo linearmente dependente (PLD)* se existem escalares $\alpha_1, \dots, \alpha_{m+p}$ não todos nulos, com $\alpha_i \geq 0$ para $i = m+1, \dots, m+p$ tais que

$$\sum_{i=1}^{m+p} \alpha_i u_i = 0.$$

Caso contrário, dizemos que $(\{u_1, \dots, u_m\}, \{u_{m+1}, \dots, u_{m+p}\})$ é *positivo-linearmente independente (PLI)*.

A condição MFCQ foi introduzida primeiramente por Arrow-Hurwitz-Uzava [16] para o caso em que as restrições são apenas de desigualdade e, posteriormente, generalizada por Mangasarian-Fromovitz para o caso em que o conjunto viável também possui restrições de igualdade. A idéia é considerar os multiplicadores de Lagrange associados às restrições de desigualdade ativas como não-negativos.

Definição 2.11 (Condição de Mangasarian-Fromovitz). Dizemos que a *Condição de Mangasarian-Fromovitz (MFCQ)* vale em $x \in \Omega$ quando

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

implica que $\lambda_i = 0$, $\mu_j = 0$ para todo $i = 1, \dots, m$, $j \in A(x)$.

Equivalentemente, diremos que os gradientes

$$\{\nabla h_i(x^*), \nabla g_j(x^*) : i = 1, \dots, m, j \in A(x^*)\}$$

são PLI.

Embora sob MFCQ não se possa garantir a unicidade dos multiplicadores, sabe-se que ela é equivalente ao fato do conjunto de multiplicadores de Lagrange ser um conjunto compacto não vazio. Note que a condição de otimalidade associada à condição MFCQ, isto é, “KKT ou não-MFCQ” é precisamente a condição de Fritz John.¹

A condição de Mangasarian-Fromovitz é equivalente a seguinte Proposição:

¹Haeser e Ramos [38], p. 15

Proposição 2.12. *Um ponto $x \in \Omega$ satisfaz MFCQ se, e somente se, são satisfeitas as seguintes condições:*

- $\{\nabla h_i(x)\}_{i=1}^m$ é linearmente independente, e
- existe um vetor $d \in \mathbb{R}^n$ tal que $\nabla h_i(x)^T d = 0$ e $\nabla g_j(x)^T d < 0$ para todo $i = 1, \dots, m$ e $j \in A(x)$.

Observe que a segunda condição diz que $d \in L_\Omega(x)$.

Demonstração. ($\Leftarrow$) Suponha por absurdo que x não satisfaz MFCQ. Logo existe $\lambda \in \mathbb{R}^m$ e $\mu_j \geq 0$, $j \in A(x)$, não todos nulos, tais que

$$\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x)} \mu_j \nabla g_j(x) = 0.$$

Logo,

$$\sum_{i=1}^m \lambda_i \nabla h_i(x)^T d + \sum_{j \in A(x)} \mu_j \nabla g_j(x)^T d = 0.$$

Pela hipótese, temos $\nabla h_i(x)^T d = 0$ para todo $i = 1, \dots, m$, logo $\sum_{j \in A(x)} \mu_j \nabla g_j(x)^T d = 0$,

e como $\nabla g_j(x)^T d < 0$, segue que $\mu_j = 0$ para todo $j \in A(x)$. Portanto, $\sum_{i=1}^m \lambda_i \nabla h_i(x) = 0$ com $\lambda \neq 0$, o que é um absurdo já que $\{\nabla h_i(x)\}_{i=1}^m$ é linearmente independente.

($\Rightarrow$) Se x satisfaz MFCQ, então os gradientes $\{\nabla h_i(x), \nabla g_j(x) : i = 1, \dots, m, j \in A(x)\}$ são PLI, de modo que $\{\nabla h_i(x)\}_{i=1}^m$ é LI. Para provar a segunda condição, considere o problema de otimização linear:

$$\begin{aligned} &\text{minimizar} && F(d, z) = z, \\ &\text{sujeito a} && \nabla h_i(x)^T d = 0, \quad i = 1, \dots, m, \\ &&& \nabla g_j(x)^T d \leq z, \quad j \in A(x). \end{aligned} \tag{2.9}$$

Suponha que o problema possui uma solução. Como as restrições são lineares, as condições de KKT valem para o problema (2.9), isto é, existem $\lambda \in \mathbb{R}^m$ e $\mu_j \geq 0$ com $\mu_j = 0$ para $j \notin A(x)$, tais que

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \sum_{i=1}^m \lambda_i \begin{pmatrix} \nabla h_i(x) \\ 0 \end{pmatrix} + \sum_{j \in A(x)} \mu_j \begin{pmatrix} \nabla g_j(x) \\ -1 \end{pmatrix} = 0.$$

Logo,

$$\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x)} \mu_j \nabla g_j(x) = 0 \quad \text{e} \quad \sum_{j \in A(x)} \mu_j = 1.$$

Portanto, $\mu \neq 0$ com

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

contradizendo a definição de MFCQ. Assim, o problema é ilimitado. Segue que existe $d \in \mathbb{R}^n$ tal que

$$\nabla h_i(x)^T d = 0, \quad i = 1, \dots, m \quad \text{e} \quad \nabla g_j(x)^T d < 0, \quad j \in A(x).$$

□

Esta é a maneira como a condição de Mangasarian-Fromovitz foi originalmente proposta em [49].

Como LI implica PLI, podemos ver claramente que a condição de independência linear implica em Mangasarian-Fromovitz.

Existem muitos problemas em que temos KKT sem que MFCQ seja satisfeita, como mostra o exemplo a seguir.

Exemplo 2.13. Considere o seguinte problema:

$$\begin{aligned} \text{minimizar} \quad & f(x) = x_1, \\ \text{sujeito a} \quad & g_1(x) = -x_1^3 + x_2 \leq 0, \\ & g_2(x) = -x_1^3 - x_2 \leq 0, \\ & g_3(x) = -x_1 \leq 0. \end{aligned}$$

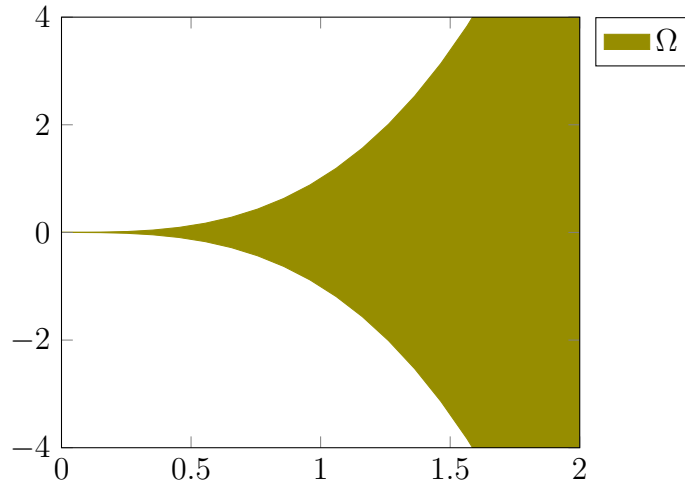


Figura 2.2: KKT não garante a validade de MFCQ

Conforme a Figura 2.2, o ponto $x^* = 0$ é a solução deste problema. Temos $A(x^*) = \{1, 2, 3\}$. As condições de KKT valem em x^* com $\mu_1 = \mu_2 = \alpha > 0$ e $\mu_3 = 1$. Porém, x^* não satisfaz Mangasarian-Fromovitz, pois os gradientes $\nabla g_1(x^*), \nabla g_2(x^*), \nabla g_3(x^*)$ são positivo linearmente dependentes.

Definição 2.14 (Condição de qualificação de Slater). Dizemos que $x \in \Omega$ satisfaz a condição de qualificação de Slater se h é afim, g_j é convexa para todo j e existe $\tilde{x} \in \Omega$ tal que $g_j(\tilde{x}) < 0$ para todo $j \in A(x)$.

Teorema 2.15. A condição de Slater é uma condição de qualificação.

Demonstração. Seja $x^* \in \Omega$. Mostraremos que $T_\Omega(x^*) = L_\Omega(x^*)$. Em virtude do Lema 1.20, basta mostrar que $L_\Omega(x^*) \subset T_\Omega(x^*)$. Considere $d \in L_\Omega(x^*)$. Defina $\tilde{d} = \tilde{x} - x^*$, onde $\tilde{x} \in \Omega$ satisfaz $g_j(\tilde{x}) < 0$ para todo $j \in A(x^*)$. Como g_j é convexa para todo j , temos

$$0 > g_j(\tilde{x}) \geq g_j(x^*) + \nabla g_j(x^*)^T \tilde{d}.$$

Dado $t \in (0, 1)$, defina $\hat{d} = (1 - t)d + t\tilde{d}$. Vamos provar que $\hat{d} \in T_\Omega(x^*)$ para todo $t \in (0, 1)$. Dado $j \in A(x^*)$, temos $\nabla g_j(x^*)^T d \leq 0$ e $\nabla g_j(x^*)^T \tilde{d} < 0$. Logo,

$$\nabla g_j(x^*)^T \hat{d} = (1 - t)\nabla g_j(x^*)^T d + t\nabla g_j(x^*)^T \tilde{d} < 0.$$

Defina a sequência $\{x^k\}$ por $x^k := x^* + \frac{1}{k}\hat{d}$. Pelo Teorema 1.5, segue que

$$g_j(x^k) = g_j(x^* + \frac{1}{k}\hat{d}) < g_j(x^*) = 0,$$

para todo k suficientemente grande. Por outro lado, se $j \notin A(x^*)$, então $g_j(x^*) < 0$. Assim, pela continuidade de g_j , temos que $g_j(x^k) < 0$ para todo k suficientemente grande. Como h é afim, digamos, $h(x) = Ax - b$ e $d \in L_\Omega(x^*)$, temos $Ad = \nabla h(x^*)^T d = 0$. Além disso, temos $A\tilde{d} = A(\tilde{x} - x^*) = (A\tilde{x} - b) - (Ax^* - b) = h(\tilde{x}) - h(x^*) = 0$. Portanto,

$$\begin{aligned} h(x^k) &= Ax^k - b = A\left(x^* + \frac{1}{k}\hat{d}\right) - b \\ &= (Ax^* - b) + \frac{1}{k}A\hat{d} = 0. \end{aligned}$$

Assim, concluímos que $\{x^k\}$ é uma sequência de pontos viáveis. Como

$$\frac{x^k - x^*}{\|x^k - x^*\|} = \frac{\hat{d}}{\|\hat{d}\|},$$

segue que $\hat{d} \in T_\Omega(x^*)$. Fazendo $t \rightarrow 0$ e usando o fato de $T_\Omega(x^*)$ ser um conjunto fechado, obtemos que $d \in T_\Omega(x^*)$, completando a prova. $\square$

Proposição 2.16. *A condição de Slater implica MFCQ.*

Demonstração. Suponha que vale a condição de Slater em $x^* \in \Omega$. Como h é afim, podemos escrever h como $h(x) = Ax + b$, onde $A \in \mathbb{R}^{m \times n}$ e $b \in \mathbb{R}^m$. Além disso, g_j é convexa para todo j e existe $\tilde{x} \in \Omega$ tal que $g_j(\tilde{x}) < 0$ para todo $j \in A(x^*)$. Usando a convexidade de g_j , segue que

$$0 > g_j(\tilde{x}) \geq g_j(x^*) + \nabla g_j(x^*)^T (\tilde{x} - x^*).$$

Tome $d = \tilde{x} - x^*$. Dessa forma, para cada $j \in A(x^*)$, temos

$$\nabla g_j(x^*)^T d < 0.$$

Como $\tilde{x}, x^* \in \Omega$, temos $h(\tilde{x}) = h(x^*) = 0$. Logo,

$$\nabla h(x^*)^T d = A(\tilde{x} - x^*) = (A\tilde{x} + b) - (Ax^* + b) = 0$$

e, portanto, $\nabla h_i(x^*)^T d = 0$ para todo $i = 1, \dots, m$.

Pela linearidade das funções h_i , podemos assumir sem perda de generalidade que os gradientes $\nabla h_i(x)$ são LI, caso contrário podemos eliminar um número suficiente de igualdades para obter um conjunto LI. $\square$

Em 1984 Fujiwara, Han e Mangasarian [33] introduziram uma nova condição de qualificação mais forte que MFCQ e mais fraca que LICQ, e que chamamos de *Condição de Mangasarian-Fromovitz estrita SMFCQ*.

Definição 2.17 (Condição de Mangasarian-Fromovitz estrita). Suponha que $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ seja um multiplicador de Lagrange associado a $x \in \Omega$. Dizemos que a *Condição de Mangasarian-Fromovitz estrita (SMFCQ)* vale em x quando os gradientes das restrições $\{\nabla h_i(x), \nabla g_j(x) : i = 1, \dots, m, j \in A(x), \mu_j > 0\}$ e $\{\nabla g_j(x) : j \in A(x), \mu_j = 0\}$ são PLI.

Em muitos casos de interesse, a unicidade dos multiplicadores de Lagrange simplifica a análise de convergência. Assim, em [45] é provado que SMFCQ é uma condição necessária e suficiente para a unicidade dos multiplicadores de Lagrange, como mostra a próxima proposição.

Proposição 2.18. *Suponha que $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ seja um multiplicador de Lagrange associado a $x \in \Omega$. Então x satisfaz SMFCQ se, e somente se, (λ, μ) é único.*

Demonstração. $(\Rightarrow)$ Suponha que $(\tilde{\lambda}, \tilde{\mu})$ seja outro multiplicador de Lagrange associado a $x \in \Omega$. Então, pelas condições KKT, temos

$$\sum_{i=1}^m (\tilde{\lambda}_i - \lambda_i) \nabla h_i(x) + \sum_{j \in A(x): \mu_j > 0} (\tilde{\mu}_j - \mu_j) \nabla g_j(x) + \sum_{j \in A(x): \mu_j = 0} \tilde{\mu}_j \nabla g_j(x) = 0.$$

Como $(\tilde{\lambda}, \tilde{\mu}) \neq (\lambda, \mu)$, então algum coeficiente na combinação linear acima é não nulo, contradizendo a definição de SMFCQ.

$(\Leftarrow)$ Seja $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ o multiplicador de Lagrange. Suponha que não vale a condição SMFCQ. Logo existe $0 \neq (\tilde{\lambda}, \tilde{\mu}) \in \mathbb{R}^m \times \mathbb{R}_+^p$ com $\tilde{\mu}_j > 0$ para $j \in A(x)$ e $\tilde{\mu}_j = 0$ tal que

$$\sum_{i=1}^m \tilde{\lambda}_i \nabla h_i(x) + \sum_{j \in A(x)} \tilde{\mu}_j \nabla g_j(x) = 0.$$

Claramente, temos que

$$\nabla f(x) + \sum_{i=1}^m (\tilde{\lambda}_i + \lambda_i) \nabla h_i(x) + \sum_{j \in A(x)} (\tilde{\mu}_j + \mu_j) \nabla g_j(x) = 0.$$

Tomando $(\tilde{\lambda}, \tilde{\mu})$ suficientemente pequeno de modo que $\tilde{\mu}_j + \mu_j \geq 0$ para $j \in A(x)$, segue-se que $(\tilde{\lambda} + \lambda, \tilde{\mu} + \mu)$ é um multiplicador de Lagrange diferente de (λ, μ) , contradizendo a hipótese de (λ, μ) ser único. $\square$

Sabemos que a condição de independência linear garante a existência e unicidade dos multiplicadores de Lagrange independente da função objetivo f . Portanto, se para qualquer função objetivo f , existe um multiplicador de Lagrange (λ, μ) associado a $x \in \Omega$ tal que SMFCQ é satisfeita em x , então LICQ vale em x .

Outra maneira de enfraquecer LICQ é, ao invés de requerer posto completo para o conjunto de gradientes, permitir que o posto seja deficiente em uma vizinhança do ponto. Esta condição é chamada *condição de posto constante (CRCQ)*, introduzida por Janin [41] em 1984. Ela é forte no seguinte sentido: se vale CRCQ em $x \in \Omega$, então existe uma vizinhança V de x tal que CRCQ vale em todo $y \in V$. Ao contrário de LICQ e MFCQ, a condição CRCQ permite redundância na descrição de Ω .

Definição 2.19. Dizemos que a *Condição de Posto Constante, Constant Rank Constraint Qualification (CRCQ)* é satisfeita em $x \in \Omega$ se existe uma vizinhança V de x tal que para todo $\mathcal{I} \subset \{1, \dots, m\}$ e todo $\mathcal{J} \subset A(x)$ o posto de $\{\nabla h_i(y), \nabla g_j(y) : i \in \mathcal{I}, j \in \mathcal{J}\}$ se mantém constante para todo $y \in V$.

Equivalentemente, a condição de posto constante pode ser substituída pela *condição de dependência linear constante*, como mostra a próxima proposição.

Proposição 2.20. *Um ponto viável $x \in \Omega$ satisfaz CRCQ se, e somente se, para todos $\mathcal{I} \subset \{1, \dots, m\}$ e $\mathcal{J} \subset A(x)$ existe uma vizinhança V de x tal que, se $\{\nabla h_i(x), \nabla g_j(x) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é linearmente dependente, então $\{\nabla h_i(y), \nabla g_j(y) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é linearmente dependente para todo $y \in V$.*

Demonstração. ($\Rightarrow$) Sejam $\mathcal{I} \subset \{1, \dots, m\}$ e $\mathcal{J} \subset A(x)$ e suponha que $\{\nabla h_i(x), \nabla g_j(x) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é linearmente dependente. Logo, pela hipótese, existe uma vizinhança V de x tal que o posto de $\{\nabla h_i(y), \nabla g_j(y) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é constante para todo $y \in V$, isto é,

$$\text{posto}(\{\nabla h_i(y), \nabla g_j(y) : i \in \mathcal{I}, j \in \mathcal{J}\}) = \text{posto}(\{\nabla h_i(x), \nabla g_j(x) : i \in \mathcal{I}, j \in \mathcal{J}\}).$$

Portanto, $\{\nabla h_i(y), \nabla g_j(y) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é LD para todo $y \in V$.

($\Leftarrow$) Sejam $\mathcal{I} \subset \{1, \dots, m\}$ e $\mathcal{J} \subset A(x)$ e suponha que o posto de $\{\nabla h_i(x), \nabla g_j(x) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é igual a r . Tome $\mathcal{I}_0 \subset \mathcal{I}$ e $\mathcal{J}_0 \subset \mathcal{J}$ com $|\mathcal{I}_0| + |\mathcal{J}_0| = r$ tais que $\{\nabla h_i(x), \nabla g_j(x) : i \in \mathcal{I}_0, j \in \mathcal{J}_0\}$ seja LI. Pela Proposição 2.9, existe uma vizinhança V de x tal que $\{\nabla h_i(y), \nabla g_j(y) : i \in \mathcal{I}_0, j \in \mathcal{J}_0\}$ é LI para todo $y \in V$. Logo, $\{\nabla h_i(y), \nabla g_j(y) : i \in \mathcal{I}, j \in \mathcal{J}\}$ tem posto maior ou igual a r para todo $y \in V$. Queremos provar que o posto de $\{\nabla h_i(y), \nabla g_j(y) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é igual a r . Então, suponha por absurdo que exista uma sequência $y^k \rightarrow x$ tal que

$$\text{posto}(\{\nabla h_i(y^k), \nabla g_j(y^k) : i \in \mathcal{I}, j \in \mathcal{J}\}) > r.$$

Logo, para cada k , existe $\mathcal{I}_k \subset \mathcal{I}$ e $\mathcal{J}_k \subset \mathcal{J}$ com $|\mathcal{I}_k| + |\mathcal{J}_k| > r$ tal que $\{\nabla h_i(y^k), \nabla g_j(y^k) : i \in \mathcal{I}_k, j \in \mathcal{J}_k\}$ é LI. Podemos tomar uma subsequência tal que $\mathcal{I}_k = \mathcal{I}$ e $\mathcal{J}_k = \mathcal{J}$ para todo k . Assim, $\{\nabla h_i(x), \nabla g_j(x) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é LD com $\{\nabla h_i(y^k), \nabla g_j(y^k) : i \in \mathcal{I}, j \in \mathcal{J}\}$ LI, contradizendo a hipótese. $\square$

A condição CRCQ sempre é satisfeita em problemas com restrições lineares, o que fornece uma justificativa de que as condições KKT sempre valem para problemas de otimização linear. Os dois contra-exemplos a seguir mostram que, de uma maneira geral, as condições MFCQ e CRCQ não estão relacionadas.

Contra-Exemplo 2.21 (MFCQ não implica CRCQ). Considere as restrições de desigualdade $g_1(x) = -x_2$ e $g_2(x) = x_1^2 - x_2$. Claramente MFCQ é satisfeita no ponto viável $x^* = (0, 0)$. Porém, CRCQ não vale em x^* , pois $\text{posto}(\{\nabla g_1(y), \nabla g_2(y)\}) = 1$ em $y = x^*$ e $\text{posto}(\{\nabla g_1(y), \nabla g_2(y)\}) = 2$ para y arbitrariamente próximo de x^* com $y_1 \neq 0$.

Contra-Exemplo 2.22 (CRCQ não implica MFCQ). Considere as restrições de desigualdade $g_1(x) = x_1 + x_2$, $g_2(x) = -x_1 - x_2$, $g_3(x) = -x_2$ e o ponto $x^* = (0, 0)$. Como todas as restrições são lineares, segue que CRCQ vale em x^* . Por outro lado, como $A(x^*) = \{1, 2, 3\}$, temos $\{\nabla g_j(x^*) : j \in A(x^*)\} = \{\nabla g_1(x^*), \nabla g_2(x^*), \nabla g_3(x^*)\}$. Portanto não vale MFCQ em x^* .

Outra condição de qualificação bastante conhecida e utilizada para analisar a convergência de métodos de programação quadrática sequencial é a *condição de dependência linear positiva constante* (CPLD), introduzida por Qi e Wei em [57]. Entretanto, os autores apenas conjecturaram que CPLD era uma condição de qualificação, aparecendo como um problema em aberto na Seção 2 de [57]. Esta questão foi resolvida anos mais tarde quando Andreani, Martínez e Schuverdt provam em [14] que CPLD é de fato uma condição de qualificação mais fraca que MFCQ e CRCQ. Schuverdt também mostrou em

[63] que CPLD pode ser utilizada para analisar a convergência de algoritmos do tipo Lagrangiano aumentado. Assim como MFCQ enfraquece LICQ olhando para o sinal dos multiplicadores, a condição CPLD enfraquece CRCQ no mesmo sentido.

Definição 2.23. Dizemos que a *Condição de Dependência Linear Positiva Constante (CPLD)* vale no ponto $x^* \in \Omega$ se para todos $\mathcal{I} \subset \{1, \dots, m\}$ e $\mathcal{J} \subset A(x^*)$ tais que o conjunto $\{\nabla h_i(x^*), \nabla g_j(x^*) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é positivo-linearmente dependente, existir uma vizinhança V de x^* tal que $\{\nabla h_i(x), \nabla g_j(x) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é positivo-linearmente dependente para todo $x \in V$.

Teorema 2.24. *MFCQ implica CPLD.*

Demonstração. Basta observar que não existem $\mathcal{I}' \subset \{1, \dots, m\}$ e $\mathcal{J}' \subset A(x^*)$ tais que o conjunto $\{\nabla h_i(x^*), \nabla g_j(x^*) : i \in \mathcal{I}', j \in \mathcal{J}'\}$ seja PLD, uma vez que os gradientes $\{\nabla h_i(x^*), \nabla g_j(x^*) : i = 1, \dots, m, j \in A(x^*)\}$ são PLI. $\square$

O contra-exemplo a seguir mostra que a recíproca deste Teorema não é verdadeira.

Contra-Exemplo 2.25 (CPLD não implica MFCQ). Considere em $\mathbb{R}^2$ o ponto $x^* = (0, 0)$ e o conjunto Ω definido pelas restrições $g_1(x) = x_1 + x_2 \leq 0$, $g_2(x) = -x_1 \leq 0$ e $g_3(x) = -x_2 \leq 0$. Temos $A(x^*) = \{1, 2, 3\}$. Calculando os gradientes, obtemos

$$\nabla g_1(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \nabla g_2(x) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \nabla g_3(x) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

e, portanto, não vale MFCQ em x^* . Para mostrar que vale CPLD, basta observar que para $\mathcal{J} = A(x^*)$, temos $\{\nabla g_j(x^*) : j \in \mathcal{J}\} = \{\nabla g_j(x) : j \in \mathcal{J}\}$ PLD para todo x numa vizinhança da origem.

Podemos encontrar outro tipo de enfraquecimento dessas condições na literatura, considerando apenas certos subconjuntos de restrições ao invés de todos.

Definição 2.26 (RCRCQ, [53]). Dizemos que a *condição Relaxed-CRCQ (RCRCQ)* vale no ponto viável $x^* \in \Omega$, se existe uma vizinhança V de x^* tal que para todo $\mathcal{J} \subset A(x^*)$, o posto de $\{\nabla h_i(x), \nabla g_j(x) : i = 1, \dots, m; j \in \mathcal{J}\}$ permanece constante para todo $x \in V$.

Comparado a CRCQ, este relaxamento trata o conjunto de restrições de igualdade como um todo, sem a necessidade de impor restrições a todos os seus subconjuntos. Assim, RCRCQ é claramente menos restritiva do que CRCQ e mais simples de ser verificada. Em [52], os autores provaram que RCRCQ é uma condição de qualificação, mostrando que implica a condição de qualificação de Abadie. Ainda mais, no próximo capítulo veremos que RCRCQ garante a validade da condição de otimalidade de segunda ordem.

Teorema 2.27 ([3], TEOREMA 1). *Seja $\mathcal{I} \subset \{1, \dots, m\}$ tal que $\{\nabla h_i(x)\}_{i \in \mathcal{I}}$ forma uma base para $\text{span}\{\nabla h_i(x)\}_{i=1}^m$. Um ponto viável $x \in \Omega$ satisfaz RCRCQ se, e somente se, existir uma vizinhança V de x tal que*

- $\{\nabla h_i(y)\}_{i=1}^m$ têm posto constante para todo $y \in V$;
- Para todo $\mathcal{J} \subset A(x)$, se $\{\nabla h_i(x), \nabla g_j(x) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é linearmente dependente, então $\{\nabla h_i(y), \nabla g_j(y) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é linearmente dependente para todo $y \in V$.

Quando o conjunto viável é descrito por meio de restrições de desigualdades, a preservação do posto do conjunto dos gradientes não é uma condição muito útil, como veremos no próximo contra-exemplo. A preservação do posto é uma ferramenta mais adequada para lidar com os gradientes das restrições de igualdade, pois estes geram um subespaço contido em $L_\Omega(x)^\circ$, onde a noção de dimensão pode ser aplicada.

Em [3], os autores introduziram uma relaxação semelhante à RCRCQ para a condição CPLD, chamada relaxed-CPLD (RCPLD).

Definição 2.28 (RCPLD, [3]). Seja $\mathcal{I} \subset \{1, \dots, m\}$ tal que $\{\nabla h_i(x)\}_{i \in \mathcal{I}}$ forma uma base para $\text{span}\{\nabla h_i(x)\}_{i=1}^m$. Dizemos que a *condição Relaxed-CPLD (RCPLD)* vale no ponto viável $x \in \Omega$ se existir uma vizinhança V de x satisfazendo as condições:

- $\{\nabla h_i(y)\}_{i=1}^m$ têm posto constante para todo $y \in V$;
- Para todo $\mathcal{J} \subset A(x)$, se $\{\nabla h_i(x), \nabla g_j(x) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é positivo-linearmente dependente, então $\{\nabla h_i(y), \nabla g_j(y) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é positivo-linearmente dependente para todo $y \in V$.

Note que a definição de RCPLD não depende da escolha do conjunto de índices que seleciona a base para $\text{span}\{\nabla h_i(x)\}_{i=1}^m$, como mostra em [3].

Em [9], os autores introduziram uma outra condição que unifica todas as demais, no sentido que é implicada pelas condições apresentadas anteriormente. Essa condição é chamada de *condição de posto constante da componente subespaço*, em inglês, *Constant Rank of the Subspace Component (CRSC)*.

Considere o polar do cone linearizado,

$$L_\Omega(x)^\circ = \left\{ v \in \mathbb{R}^n : v = \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x)} \mu_j \nabla g_j(x), \mu_j \geq 0, \lambda_i \in \mathbb{R} \right\},$$

e considere o maior subespaço contido em $L_\Omega(x)^\circ$, chamado de *subespaço componente*. Neste caso, o subespaço componente é dado por $L_\Omega(x)^\circ \cap -L_\Omega(x)^\circ$. Pela definição de $L_\Omega(x)^\circ$, os gradientes $\nabla h_i(x)$ devem pertencer ao conjunto de geradores do subespaço componente. O outro termo na soma associado aos gradientes das restrições de desigualdade deve ser um cone pontudo. A maioria dos problemas surgem quando os gradientes das restrições de desigualdade pertencem ao subespaço componente do polar do cone linearizado. Formalmente, isso acontece sempre que o conjunto

$$J_-(x) := \{j \in A(x) : -\nabla g_j(x) \in L_\Omega(x)^\circ\} \neq \emptyset.$$

Observe que

$$\text{span}\{\nabla h_i(x), \nabla g_j(x) : i = 1, \dots, m; j \in J_-(x)\} = L_\Omega(x)^\circ \cap -L_\Omega(x)^\circ.$$

Definição 2.29 (CRSC, [9]). Dizemos que $x \in \Omega$ satisfaz a condição *Constant Rank of the Subspace Component (CRSC)* se existe uma vizinhança V de x tal que os gradientes $\{\nabla h_i(y), \nabla g_j(y) : i = 1, \dots, m; j \in J_-(x)\}$ têm posto constante para todo $y \in V$.

Um fato interessante, é que, sob CRSC as restrições de desigualdade com índices em $J_-(x)$ atuam localmente como igualdades, isto é, existe uma vizinhança V de x tal que $g_j(y) = 0$ para todo $y \in V \cap \Omega$ e para todo $j \in J_-(x)$.

Antes de definir a condição de qualificação *Constant Positive Generator (CPG)* [9], vejamos alguns conceitos e resultados apresentados em [9, 38].

Definição 2.30. Seja $\mathcal{V} = \{v_1, \dots, v_{m+p}\}$ um conjunto de vetores em $\mathbb{R}^n$ e considere o par ordenado $(\mathcal{I}, \mathcal{J})$, onde $\mathcal{I} = \{1, \dots, m\}$ e $\mathcal{J} = \{m+1, \dots, m+p\}$. Definimos o conjunto $\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V})$ por

$$\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}) := \left\{ v \in \mathbb{R}^n : v = \sum_{i \in \mathcal{I}} \lambda_i v_i + \sum_{j \in \mathcal{J}} \mu_j v_j, \mu_j \geq 0, \forall j \in \mathcal{J} \right\}.$$

Claramente $\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V})$ é um cone, chamado de *cone positivamente gerado*. Quando os vetores $v_i, i \in \mathcal{I} \cup \mathcal{J}$ são PLD, diremos que o par ordenado $(\mathcal{I}, \mathcal{J})$ é PLD. Caso contrário, diremos que $(\mathcal{I}, \mathcal{J})$ é PLI. Observe que quando $v_i = \nabla h_i(x)$, para $i = 1, \dots, m$, e $v_j = \nabla g_j(x)$, para $j \in A(x)$, temos que $\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}) = L_\Omega(x)^\circ$.

Para esses tipos de cones, podemos definir um conceito de base.

Definição 2.31. Considere $\mathcal{I} = \{1, \dots, m\}$ e $\mathcal{J} = \{m+1, \dots, m+p\}$. Sejam $\mathcal{V} = \{v_1, \dots, v_{m+p}\} \subset \mathbb{R}^n$ e $\mathcal{I}', \mathcal{J}' \subset \{1, \dots, m+p\}$. Dizemos que o par $(\mathcal{I}', \mathcal{J}')$ forma uma base para $\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V})$ quando

$$\text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}) = \text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V})$$

e os vetores $v_i, i \in \mathcal{I}'$ e $v_j, j \in \mathcal{J}'$ são PLI.

Podemos mostrar que, de fato, existe uma base nos termos da definição acima. Se os vetores $v_i, i \in \mathcal{I} \cup \mathcal{J}$ são PLD, então existem $\lambda_i, i \in \mathcal{I}$ e $\mu_j \geq 0, j \in \mathcal{J}$ não todos nulos tais que

$$\sum_{i \in \mathcal{I}} \lambda_i v_i + \sum_{j \in \mathcal{J}} \mu_j v_j = 0. \quad (2.10)$$

Se o conjunto $\{v_i : i \in \mathcal{I}\}$ for LD, defina $\mathcal{I}' \subset \mathcal{I}$ um subconjunto próprio de $\mathcal{I}$ tal que $\text{span}\{v_i : i \in \mathcal{I}'\} = \text{span}\{v_i : i \in \mathcal{I}\}$ e $\mathcal{J}' = \mathcal{J}$. Claramente, temos $\text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}) = \text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V})$. Caso contrário, temos que existe $j' \in \mathcal{J}$ tal que $\mu_{j'} > 0$. Dividindo a equação (2.10) por $\mu_{j'}$, segue que

$$\sum_{i \in \mathcal{I}} \frac{\lambda_i}{\mu_{j'}} v_i + \sum_{j \in \mathcal{J} \setminus \{j'\}} \frac{\mu_j}{\mu_{j'}} v_j = -v_{j'}.$$

Definindo $\mathcal{I}' := \mathcal{I} \cup \{j'\}$ e $\mathcal{J}' := \mathcal{J} \setminus \{j'\}$, claramente temos $\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}) \subset \text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V})$. Por outro lado, seja $v = \sum_{i \in \mathcal{I}'} \bar{\lambda}_i v_i + \sum_{j \in \mathcal{J}'} \bar{\mu}_j v_j \in \text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V})$. Logo,

$$v = \sum_{i \in \mathcal{I}} \bar{\lambda}_i v_i + \bar{\lambda}_{j'} v_{j'} + \sum_{j \in \mathcal{J} \setminus \{j'\}} \bar{\mu}_j v_j \in \text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}), \text{ se } \bar{\lambda}_{j'} \geq 0.$$

Caso contrário, segue que

$$\begin{aligned} v &= \sum_{i \in \mathcal{I}'} \bar{\lambda}_i v_i + \sum_{j \in \mathcal{J}'} \bar{\mu}_j v_j = \sum_{i \in \mathcal{I}} \bar{\lambda}_i v_i + \bar{\lambda}_{j'} v_{j'} + \sum_{j \in \mathcal{J}'} \bar{\mu}_j v_j \\ &= \sum_{i \in \mathcal{I}} \bar{\lambda}_i v_i + |\bar{\lambda}_{j'}| \left(\sum_{i \in \mathcal{I}} \frac{\lambda_i}{\mu_{j'}} v_i + \sum_{j \in \mathcal{J} \setminus \{j'\}} \frac{\mu_j}{\mu_{j'}} v_j \right) + \sum_{j \in \mathcal{J}'} \bar{\mu}_j v_j \in \text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}), \end{aligned}$$

o que mostra que $\text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}) = \text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V})$. Podemos repetir essa construção um número finito de vezes até que $(\mathcal{I}', \mathcal{J}')$ seja PLI.

Definição 2.32 (CPG, [9]). Sejam $\mathcal{V}(y) = \{\nabla h_i(y), \nabla g_j(y) : i \in \mathcal{I}, j \in \mathcal{J}\}$, $\mathcal{I} = \{1, \dots, m\}$ e $\mathcal{J} = A(x)$. Dizemos que $x \in \Omega$ satisfaz a condição *Constant Positive Generator (CPG)* quando existe uma base $(\mathcal{I}', \mathcal{J}')$ de $\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}(x))$ e uma vizinhança V de x tais que

$$\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}(y)) \subset \text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}(y)), \quad \forall y \in V.$$

Teorema 2.33. *CRSC implica CPG.*

Demonstração. Seja $x \in \Omega$ um ponto que satisfaz CRSC e considere $(\mathcal{I}', \mathcal{J}')$ uma base para $\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}(x))$. Como $\mathcal{I}' \subset J_-(x)$ corresponde a uma base para $\text{span}\{\nabla f_i(x) : i \in J_-(x)\}$, temos que esta base deve permanecer uma base em uma vizinhança de x , devido ao fato de que o posto do conjunto $\{\nabla f_i(y) : i \in J_-(x)\}$ é constante. $\square$

Contra-Exemplo 2.34 (CPG não implica CRSC). Considere em $\mathbb{R}^2$, o conjunto viável

$$\Omega := \{x \in \mathbb{R}^2 : g_1(x) = x_1^3 - x_2 \leq 0, g_2(x) = x_1^3 + x_2 \leq 0, g_3(x) = x_1 \leq 0\}.$$

No ponto $x = (0, 0)$, temos

$$\nabla g_1(x) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \nabla g_2(x) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \nabla g_3(x) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Neste caso, temos $\mathcal{I} = \emptyset$, $\mathcal{J} = A(x) = \{1, 2, 3\}$, e o polar do cone linearizado,

$$L_\Omega(x)^\circ = \left\{ v : v = \sum_{j \in A(x)} \mu_j \nabla g_j(x), \mu_j \geq 0 \right\}.$$

Logo, temos que $J_-(x) = \{1, 2\}$. Tome $\mathcal{I}' = \{1\}$ e $\mathcal{J}' = \{3\}$. Assim, temos que os gradientes $\{\nabla g_1(x), \nabla g_3(x)\}$ são PLI e $\text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}(x)) = \text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}(x))$. Além disso, para todo y em uma vizinhança de x , temos

$$\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}(y)) \subset \text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}(y)).$$

Portanto, x cumpre a condição CPG. Entretanto, x não satisfaz CRSC, pois

$$\text{posto}(\{\nabla g_1(x), \nabla g_2(x)\}) = 1,$$

enquanto que $\text{posto}(\{\nabla g_1(y), \nabla g_2(y)\}) = 2$ para todo y em uma vizinhança de x .

A estrutura de um conjunto viável está localmente relacionada com a *propriedade de Error Bound*, que consiste em estimar a distância de um determinado ponto ao conjunto viável em uma vizinhança, utilizando uma medida analítica de inviabilidade, e pode desempenhar um papel central na análise de convergência de algoritmos.

Definição 2.35. Dizemos que $x \in \Omega$ cumpre a condição de *Error Bound* se existe uma vizinhança V de x e uma constante $\alpha > 0$ tal que para todo $y \in V$, vale:

$$\min_{z \in \Omega} \|z - y\| \leq \alpha \max\{|h_1(y)|, \dots, |h_m(y)|, g_1^+(y), \dots, g_p^+(y)\}. \quad (2.11)$$

Teorema 2.36. *A condição de Error Bound é uma condição de qualificação.*

Demonstração. Mostremos que vale a condição de qualificação de Abadie. Seja $d \in L_\Omega(x)$. Aplicando expansão de Taylor, temos que

$$\begin{aligned} h_i(x + td) &= t \nabla h_i(x)^T d + o(t) = o(t), \quad i = 1, \dots, m, \\ g_j(x + td) &= t \nabla g_j(x)^T d + o(t) \leq o(t), \quad j \in A(x), \\ g_j(x + td) &< 0, \quad j \notin A(x), \end{aligned}$$

com $\lim_{t \rightarrow 0} \frac{o(t)}{t} = 0$ e $t > 0$ suficientemente pequeno. Logo, usando a propriedade (2.11), obtemos $\text{dist}(x + td, \Omega) = o(t)$. Portanto, $d \in T_\Omega(x)$, concluindo a demonstração. $\square$

Contra-Exemplo 2.37 (CPG não implica Error Bound). Considere em $\mathbb{R}^2$, o ponto $x = (0, 0)$ e Ω definido pelas restrições de desigualdade $g_1(x) = x_1^3 - x_2$, $g_2(x) = x_1^3 + x_2$, $g_3(x) = x_1$ e $g_4(x) = x_2^3$. Neste caso, temos $\mathcal{I} = \emptyset$, $\mathcal{J} = A(x) = \{1, 2, 3, 4\}$ e

$$\nabla g_1(x) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \nabla g_2(x) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \nabla g_3(x) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \nabla g_4(x) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Tomando $\mathcal{I}' = \{1\}$ e $\mathcal{J}' = \{3\}$, temos $\{\nabla g_1(x), \nabla g_3(x)\}$ PLI e $\text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}(x)) = \text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}(x))$. Logo, $(\mathcal{I}', \mathcal{J}')$ forma uma base para $\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}(x))$. Não é difícil ver que $\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}(y)) \subset \text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}(y))$ para todo y em uma vizinhança da origem. Portanto vale CPG em x . Para mostrar que x não satisfaz Error Bound, considere $x^k = (-(1/k)^{1/3}, 1/k)$. Logo temos $\text{dist}(x^k, \Omega) = 1/k$, enquanto que a medida de inviabilidade é $1/k^3$.

Embora CPG não implica Error Bound, provou-se em [9] que CPG implica Abadie.

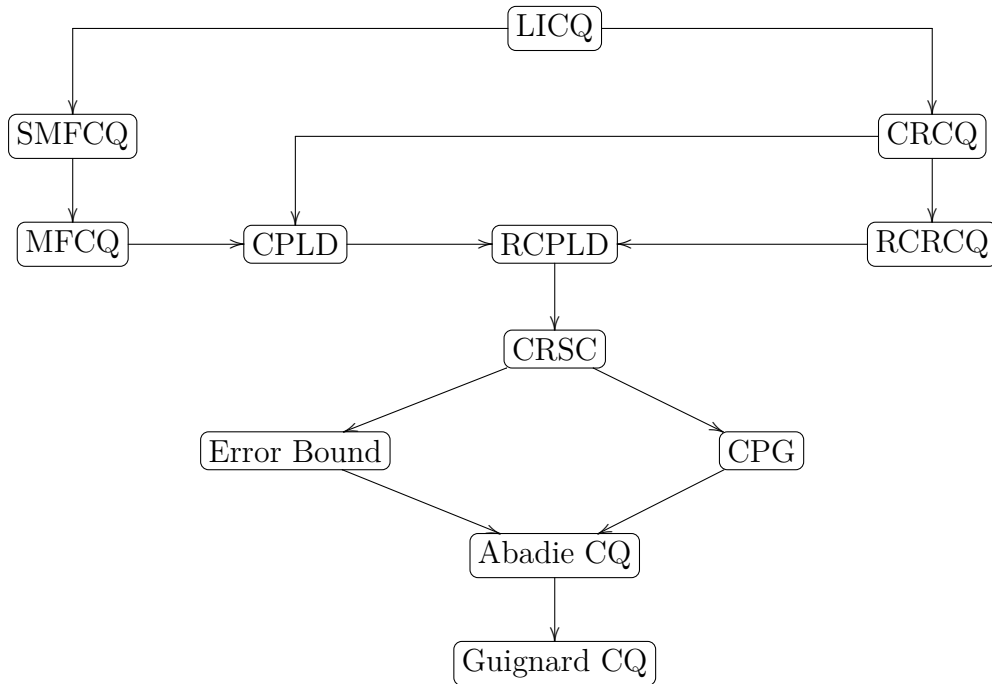


Figura 2.3: Relações entre as condições de qualificação conhecidas.

2.2 Condições sequenciais de primeira ordem

Os resultados a seguir baseiam-se essencialmente em [5, 7, 8, 11, 15, 37, 38, 51, 58].

As condições sequenciais de otimalidade foram introduzidas com o objetivo de estudar a convergência global de algoritmos iterativos. Estas condições fornecem ferramentas teóricas adequadas para justificar critérios de parada em muitos algoritmos práticos. Tipicamente, o critério de parada que corresponde à condição sequencial de otimalidade está associado a um parâmetro $\varepsilon_k \rightarrow 0$, que pode ser considerado como tolerância ao erro. Assim, o natural é terminar a execução do algoritmo quando a quantidade ε_k é suficientemente pequena. Como foi dito na introdução, condições sequenciais são necessárias para a otimalidade local sem exigir uma condição de qualificação. Nesta seção, estudaremos as condições sequenciais de otimalidade de primeira ordem que podem ser usadas como critério de parada em algoritmos. Antes de introduzirmos formalmente as condições sequenciais de otimalidade, consideremos o algoritmo de penalidade externa para o problema

$$\text{minimizar } f(x) \quad \text{sujeito a } h(x) = 0, \quad g(x) \leq 0, \quad x \in X$$

onde $X \subset \mathbb{R}^n$ é um conjunto fechado e não vazio. Considere a função $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por

$$\Phi(x) := \frac{1}{2}(\|h(x)\|^2 + \|g^+(x)\|^2),$$

onde $g^+(x) = (g_1^+(x), \dots, g_p^+(x))$ e $g_j^+(x) = \max\{0, g_j(x)\}$.

Teorema 2.38. *Dada uma sequência $\{\rho_k\} \subset \mathbb{R}$ com $\rho_k \rightarrow +\infty$, para cada k , seja x^k a solução global, caso exista, do problema*

$$\begin{aligned} &\text{minimizar } f(x) + \rho_k \Phi(x), \\ &\text{sujeito a } x \in X \end{aligned}$$

Então todo ponto limite da sequência $\{x^k\}$ é minimizador global do problema original.

Demonstração. Seja x^* um ponto limite da sequência $\{x^k\}$, caso exista. Como X é fechado, segue que $x^* \in X$. Para mostrar que x^* é solução global de $\Phi(x)$ sujeito a $x \in X$, considere $\Phi(y) < \Phi(x^*)$, $y \in X$. Segue que $\Phi(y) + c < \Phi(x^k)$ para alguma subsequência apropriada e algum $c > 0$. Multiplicando ambos os lados por ρ_k , obtemos $\rho_k \Phi(y) + \rho_k c < \rho_k \Phi(x^k)$, o que podemos reescrever como

$$f(y) + \rho_k \Phi(y) + (\rho_k c + f(x^k) - f(y)) < f(x^k) + \rho_k \Phi(x^k).$$

Logo, $(\rho_k c + f(x^k) - f(y)) > 0$ para k suficientemente grande e, portanto, $f(y) + \rho_k \Phi(y) < f(x^k) + \rho_k \Phi(x^k)$, o que contradiz a definição de x^k .

Para qualquer $y \in X$, temos $f(x^k) \leq f(x^k) + \rho_k \Phi(x^k) \leq f(y) + \rho_k \Phi(y)$. Supondo que o problema admite um ponto viável $y \in X$ tal que $\Phi(y) = 0$, segue que $f(x^k) \leq f(y)$. Passando ao limite em uma subsequência, obtemos $f(x^*) \leq f(y)$, concluindo a demonstração. $\square$

Definição 2.39. Dizemos que um ponto x^* satisfaz uma *condição sequencial de otimalidade* associada à proposição matemática $\mathcal{P}$ se existe uma sequência $x^k \rightarrow x^*$ tal que $\mathcal{P}(\{x^k\})$ é satisfeita.

2.2.1 Condição Aproximadamente KKT

A condição sequencial de otimalidade de primeira ordem mais frequentemente utilizada na prática é a condição Aproximadamente KKT [11].

Definição 2.40 (AKKT, [11]). Dizemos que um ponto viável $x \in \Omega$ cumpre a condição *Aproximadamente KKT* (AKKT), se existem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}_+^p$ com $\mu_j^k = 0$, $j \notin A(x)$ e $x^k \rightarrow x$ tais que

$$\lim_{k \rightarrow \infty} \left(\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x)} \mu_j^k \nabla g_j(x^k) \right) = 0.$$

Neste caso, dizemos que a sequência $\{x^k\}$ é uma sequência AKKT.

O seguinte Teorema mostra que a condição Aproximadamente KKT é uma condição sequencial de otimalidade.

Teorema 2.41. *Se $x^* \in \Omega$ é minimizador local, então x^* satisfaz AKKT.*

Demonstração. Como x^* é minimizador local, existe $\delta > 0$ tal que $f(x^*) \leq f(x)$ para todo x com $\|x - x^*\| \leq \delta$. Considere a aplicação do algoritmo de penalidade externa para o problema

$$\text{minimizar } f(x) + \frac{1}{2}\|x - x^*\|^2 \quad \text{sujeito a } h(x) = 0, g(x) \leq 0, \|x - x^*\| \leq \delta.$$

Como o conjunto $X = \{x \in \mathbb{R}^n : \|x - x^*\| \leq \delta\}$ é compacto, a sequência $\{x^k\}$ gerada pelo algoritmo de penalidade externa está bem definida. Pelo Teorema 2.38 todos seus pontos limites são soluções globais. Pela compacidade de X temos a existência de pelo menos um ponto limite. Como x^* é a única solução do problema, segue que $x^k \rightarrow x^*$. Logo, $\|x^k - x^*\| < \delta$ para k suficientemente grande, ou seja, x^k é ponto interior de X . Dessa forma, temos que o gradiente da função $f(x) + \frac{1}{2}\|x - x^*\|^2 + \rho_k \Phi(x)$ em $x = x^k$ deve ser igual a zero, ou seja,

$$\nabla f(x^k) + \sum_{i=1}^m \rho_k h_i(x^k) \nabla h_i(x^k) + \sum_{j=1}^p \rho_k g_j^+(x^k) \nabla g_j(x^k) + (x^k - x^*) = 0.$$

Definindo $\lambda_i^k = \rho_k h_i(x^k)$ para todo $i = 1, \dots, m$ e $\mu_j^k = \rho_k g_j^+(x^k)$ para todo $j = 1, \dots, p$, segue que se $g_j(x^*) < 0$, então $\mu_j^k = 0$ para todo k suficientemente grande. Logo,

$$\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) + (x^k - x^*) = 0.$$

Como $x^k \rightarrow x^*$, passando ao limite concluímos que x^* satisfaz AKKT. $\square$

O próximo Teorema mostra que a condição Aproximadamente KKT implica em Fritz John, ou seja, do ponto de vista de otimização, pontos AKKT são melhores candidatos à soluções do que pontos Fritz John.

Teorema 2.42. *Se $x \in \Omega$ é um ponto AKKT, então x satisfaz a condição necessária de Fritz John.*

Demonstração. Suponha que x satisfaz AKKT. Logo, existem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}_+^p$ com $\mu_j^k = 0$, $j \notin A(x)$ e $x^k \rightarrow x$ tais que

$$\varepsilon^k := \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x)} \mu_j^k \nabla g_j(x^k) \rightarrow 0. \quad (2.12)$$

Se a sequência $\{(\lambda^k, \mu^k)\}$ for limitada, tomando o limite em uma subsequência convergente e usando a continuidade dos gradientes das funções, concluímos que x satisfaz KKT. Caso contrário, definimos, para todo $k \in \mathbb{N}$,

$$M_k = \max\{|\lambda_i^k|, \mu_j^k : i = 1, \dots, m; j \in A(x)\}.$$

Neste caso, temos que $\lim_{k \rightarrow \infty} M_k = \infty$. A sequência $\{(\frac{\lambda^k}{M_k}, \frac{\mu^k}{M_k})\}$ é limitada, pois $\|(\frac{\lambda^k}{M_k}, \frac{\mu^k}{M_k})\|_\infty = 1$. Logo, existe uma subsequência convergente, digamos

$$\left(\frac{\lambda^k}{M_k}, \frac{\mu^k}{M_k} \right) \xrightarrow{\mathbb{N}'} (\hat{\lambda}, \hat{\mu}).$$

Assim, dividindo ε^k por M_k e passando ao limite quando $k \rightarrow \infty$, obtemos

$$\sum_{i=1}^m \hat{\lambda}_i \nabla h_i(x) + \sum_{j \in A(x)} \hat{\mu}_j \nabla g_j(x) = 0,$$

com $\|(\hat{\lambda}, \hat{\mu})\|_\infty = 1$. Portanto, os gradientes $\{\nabla h_i(x), \nabla g_j(x) : i = 1, \dots, m, j \in A(x)\}$ são PLD, contradizendo a definição de MFCQ. Assim, se x é um ponto AKKT, então x satisfaz “KKT ou não-MFCQ”. Portanto, x é um ponto Fritz John. $\square$

Este resultado mostra que um ponto AKKT que satisfaz MFCQ deve ser um ponto KKT. Isto segue do fato de que MFCQ garante a compacidade do conjunto dos multiplicadores de Lagrange. Devido ao fato de AKKT ser uma condição de otimalidade, podemos concluir que MFCQ é de fato uma condição de qualificação.

Proposição 2.43. *O ponto x satisfaz AKKT se, e somente se, existem sequências $x^k \rightarrow x$, $\{(\lambda^k, \mu^k)\} \subset \mathbb{R}^m \times \mathbb{R}_+^p$ e $\{\varepsilon_k\} \subset \mathbb{R}_+$, $\varepsilon_k \rightarrow 0$ tais que:*

$$(i) \quad \|\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k)\| \leq \varepsilon_k;$$

$$(ii) \quad \|h(x^k)\| \leq \varepsilon_k, \quad \|g^+(x^k)\| \leq \varepsilon_k;$$

$$(iii) \quad g_j(x^k) < -\varepsilon_k \Rightarrow \mu_j^k = 0.$$

Demonstração. $(\Rightarrow)$ Suponha que x satisfaz a condição AKKT. Para todo $k \in \mathbb{N}$, definimos

$$\begin{aligned} \varepsilon_k &:= \max \left\{ \|(h_i(x^k))_{i=1}^m\|, \|(g_j^+(x^k))_{j=1}^p\|, \right. \\ &\quad \left. \|\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k)\|, -g_j(x^k) : j \in A(x) \right\}. \end{aligned}$$

Pela definição de AKKT, segue que $\varepsilon_k \rightarrow 0^+$. Dado $r \in \{1, \dots, p\}$ tal que $g_r(x^k) < -\varepsilon_k$, pela definição de ε_k , segue que $-g_r(x^k) > \varepsilon_k \geq -g_j(x^k)$ para todo $j \in A(x)$. Logo $r \notin A(x)$. Definindo $\mu_r^k = 0$, o resultado segue.

($\Leftarrow$) Pela continuidade das funções h_i e g_j , segue que x é viável. Agora note que se $j \notin A(x)$, então $g_j(x) < 0$ e, portanto, $g_j(x^k) < -\varepsilon_k$ para k suficientemente grande. Logo, pela hipótese, $\mu_j^k = 0$. Portanto vale AKKT. $\square$

A estrutura geral do método Lagrangiano aumentado é bem conhecida na literatura [21, 22, 55], além de ser um dos métodos mais populares para resolver problemas de otimização. A seguir, apresentamos a descrição geral do Algoritmo Lagrangiano aumentado. Dado $\rho > 0$ e $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$, considere a função Lagrangiana aumentada

$$\mathcal{L}_\rho(x, \lambda, \mu) := f(x) + \frac{\rho}{2} \left[\sum_{i=1}^m \left(h_i(x) + \frac{\lambda_i}{\rho} \right)^2 + \sum_{j=1}^p \left(\max \left\{ 0, g_j(x) + \frac{\mu_j}{\rho} \right\} \right)^2 \right].$$

Algoritmo 2.44 (Método Lagrangiano Aumentado). *Sejam $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, $\gamma > 1$, $\rho_1 > 0$ e $\tau \in (0, 1)$. Tome $\{\varepsilon_k\} \subset \mathbb{R}_+$ com $\varepsilon_k \rightarrow 0$. Escolha $x^0 \in \mathbb{R}^n$ como ponto inicial e $\lambda^1 \in [\lambda_{\min}, \lambda_{\max}]^m$, $\mu^1 \in [0, \mu_{\max}]^p$. Defina $V_j^0 = g_j^+(x)$, $j = 1, \dots, p$. Inicialize $k \leftarrow 1$.*

(1) (Minimização). Encontre $x^k \in \mathbb{R}^n$ tal que

$$\|\nabla \mathcal{L}_{\rho_k}(x^k, \lambda^k, \mu^k)\| \leq \varepsilon_k.$$

Caso não seja possível, interrompa a execução do algoritmo declarando falha.

(2) (Atualização do parâmetro de penalidade) Defina para todo $j = 1, \dots, p$

$$V_j^k := \max\{g_j(x^k), -\mu_j^k/\rho_k\}.$$

Se $\max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\} \leq \tau \max\{\|h(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\}$,

defina $\rho_{k+1} = \rho_k$. Caso contrário, coloque $\rho_{k+1} = \gamma \rho_k$;

(3) (Atualização dos multiplicadores de Lagrange). Calcule

$$\bar{\lambda}_i^k := \lambda_i^k + \rho_k h_i(x^k), \quad i = 1, \dots, m \quad \text{e} \quad \bar{\mu}_j^k := \max\{0, \mu_j^k + \rho_k g_j(x^k), \quad j = 1, \dots, p.$$

Defina $\lambda_i^{k+1} := \text{proj}_{[\lambda_{\min}, \lambda_{\max}]}(\bar{\lambda}_i^k) \in [\lambda_{\min}, \lambda_{\max}]$, para todo $i = 1, \dots, m$ e $\mu_j^{k+1} := \text{proj}_{[0, \mu_{\max}]}(\bar{\mu}_j^k) \in [0, \mu_{\max}]$, para todo $j = 1, \dots, p$. Faça $k \leftarrow k+1$ e vá para o passo (1).

O Teorema a seguir mostra que o método Lagrangiano aumentado gera sequências cujos pontos limites são pontos AKKT.

Teorema 2.45. *Suponha que a sequência $\{x^k\}$ gerada pelo algoritmo Lagrangiano aumentado admite um ponto limite $x^* \in \Omega$. Então x^* é um ponto AKKT.*

Demonstração. Usando a definição de $\mathcal{L}_\rho$ e o passo (1) do Algoritmo 2.44, temos

$$\|\nabla f(x^k) + \sum_{i=1}^m \hat{\lambda}_i^k \nabla h_i(x^k) + \sum_{j=1}^p \hat{\mu}_j^k \nabla g_j(x^k)\| \leq \varepsilon_k,$$

onde $\hat{\lambda}^k := \lambda^k + \rho_k h(x^k)$ e $\hat{\mu}^k := \max\{0, \mu^k + \rho_k g(x^k)\}$. Vamos mostrar que $\hat{\mu}_j^k = 0$ para $j \notin A(x^*)$ e k suficientemente grande. Suponha que $g_{j_0}(x^*) < 0$ para algum $j_0 \in \{1, \dots, p\}$. Se $\rho_k \rightarrow +\infty$ e já que a sequência $\{\mu_{j_0}^k\}$ é limitada, segue que $\hat{\mu}_{j_0}^k = 0$. Se $\{\rho_k\}$ é limitada, temos que $V_{j_0}^k = \max\{g_{j_0}(x^k), -\mu_{j_0}^k/\rho_k\} \rightarrow 0$ e, portanto, $-\mu_{j_0}^k/\rho_k \rightarrow 0$. Logo $\hat{\mu}_{j_0}^k = \rho_k \max\{0, g_{j_0}(x^k) + \mu_{j_0}^k/\rho_k\} = 0$. Portanto x^* satisfaz AKKT. $\square$

2.2.2 Condição Gradiente Projetado Aproximado

Outra condição sequencial de otimalidade bastante utilizada em métodos de restauração inexata e mais forte que AKKT é a condição Gradiente Projetado Aproximado (AGP), introduzida por Martínez e Svaiter em [51].

Definição 2.46 (AGP, [51]). Dado um número não positivo $\gamma \in [-\infty, 0]$ e um ponto viável $x^* \in \Omega$. Dizemos que x^* satisfaz a condição $AGP(\gamma)$ (*Approximate Gradient Projection*) se existe uma sequência $x^k \rightarrow x^*$ tal que

$$P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k)) - x^k \rightarrow 0,$$

onde $P_{\Omega(x^k, \gamma)}$ é a projeção ortogonal sobre o conjunto convexo fechado $\Omega(x^k, \gamma)$ definido como

$$\Omega(x^k, \gamma) := \left\{ z \in \mathbb{R}^n : \begin{array}{ll} \langle \nabla h_i(x^k), z - x^k \rangle = 0, & \text{para todo } i = 1, \dots, m \\ \langle \nabla g_j(x^k), z - x^k \rangle \leq 0, & \text{se } 0 \leq g_j(x^k), j = 1, \dots, p \\ g_j(x^k) + \langle \nabla g_j(x^k), z - x^k \rangle \leq 0, & \text{se } \gamma < g_j(x^k) < 0, j = 1, \dots, p \end{array} \right\}.$$

Neste caso, dizemos que a sequência $\{x^k\}$ é uma sequência $AGP(\gamma)$ associada a x^* .

O conjunto $\Omega(x^k, \gamma)$ pode ser interpretado como uma aproximação linear do conjunto

$$\left\{ z \in \mathbb{R}^n : \begin{array}{ll} h_i(z) = h_i(x^k), & i = 1, \dots, m \\ g_j(z) \leq g_j(x^k), & \text{se } g_j(x^k) \geq 0 \\ g_j(z) \leq 0, & \text{se } \gamma < g_j(x^k) < 0 \end{array} \right\}.$$

Proposição 2.47 ([51], PROPRIEDADE 2.1). *Seja $x^* \in \Omega$ e suponha que x^* satisfaz $AGP(\gamma)$ para algum $\gamma \in [-\infty, 0)$. Então x^* satisfaz $AGP(\gamma')$ para todo $\gamma' < 0$.*

Proposição 2.48 ([8], PROPOSIÇÃO 3). *Sejam x, ε em $\mathbb{R}^n$ tal que $x + \varepsilon \in \Omega(x, 0)$. Então, todo elemento do cone $N_{\Omega(x, 0)}(x + \varepsilon)$ é da forma*

$$v = \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j: g_j(x) \geq 0} \mu_j \nabla g_j(x),$$

onde $\lambda \in \mathbb{R}^m$, $\mu_j \geq 0$ e $\mu_j(\langle \nabla g_j(x), \varepsilon \rangle) = 0$ se $g_j(x) \geq 0$. Além disso, $N_{\Omega(x, 0)}(x + \varepsilon) \subset N_{\Omega(x, 0)}(x)$.

Note que isso significa que $N_{\Omega(x, \gamma)}(x) = L_{\Omega}(x)^\circ$ para todo $x \in \Omega$.

O Teorema a seguir estabelece uma relação entre a projeção euclidiana e as condições de KKT.

Teorema 2.49. *Dado $x \in \Omega$, então $P_{\Omega(x,\gamma)}(x - \nabla f(x)) = x$ se, e somente se, as condições KKT valem em x .*

Demonstração. ($\Rightarrow$) Suponha que $P_{\Omega(x,\gamma)}(x - \nabla f(x)) = x$. Logo x é a única solução do problema de otimização:

$$\text{minimizar } \frac{1}{2} \|z - x + \nabla f(x)\|^2, \text{ sujeito a } z \in \Omega(x, \gamma).$$

Pela linearidade das restrições, as condições KKT valem para esse problema, ou seja, existem $\lambda \in \mathbb{R}^m$, $\mu_j \geq 0$ e $\tilde{\mu}_j \geq 0$ tais que

$$\begin{aligned} z - x + \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x)} \mu_j \nabla g_j(x) + \sum_{\gamma < g_j(x) < 0} \tilde{\mu}_j \nabla g_j(x) &= 0, \\ \mu_j \langle \nabla g_j(x), z - x \rangle &= 0, \\ \tilde{\mu}_j (g_j(x) + \langle \nabla g_j(x), z - x \rangle) &= 0. \end{aligned} \quad (2.13)$$

Substituindo $z = x$ em (2.13), obtemos $\tilde{\mu}_j = 0$. Portanto x satisfaz KKT.

($\Leftarrow$) Agora suponha que x satisfaz KKT, isto é, existem $\lambda \in \mathbb{R}^m$ e $\mu_j \geq 0$, $j \notin A(x)$ tais que

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

Pela Proposição 2.48, temos que $-\nabla f(x) \in N_{\Omega(x,\gamma)}(x)$. Como $\Omega(x, \gamma)$ é convexo e fechado, a Proposição A.40 garante que $P_{\Omega(x,\gamma)}(x - \nabla f(x)) = x$. $\square$

O Teorema a seguir mostra que a condição AGP é uma condição sequencial de otimalidade.

Teorema 2.50. *Se x^* é um minimizador local e $\gamma \in [-\infty, 0]$, então x^* satisfaz a condição AGP(γ).*

Demonstração. Seja x^* um minimizador local. Considere uma sequência $\{\rho_k\} \subset \mathbb{R}_+$ tal que $\rho_k \rightarrow +\infty$ e $\gamma \in [-\infty, 0]$. Seguindo os mesmos passos da demonstração do Teorema 2.41, segue-se que existe uma sequência $x^k \rightarrow x^*$ tal que

$$\nabla f(x^k) + \sum_{i=1}^m \rho_k h_i(x^k) \nabla h_i(x^k) + \sum_{j=1}^p \rho_k g_j^+(x^k) \nabla g_j(x^k) + (x^k - x^*) = 0.$$

Definindo $\lambda_i^k = \rho_k h_i(x^k)$ para todo $i = 1, \dots, m$ e $\mu_j^k = \rho_k g_j^+(x^k)$ para todo $j = 1, \dots, p$, temos que se $g_j(x^*) < 0$, então $\mu_j^k = 0$ para todo k suficientemente grande. Logo,

$$\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) + (x^k - x^*) = 0,$$

no qual equivale a

$$x^k - \nabla f(x^k) - \left(x^k + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) \right) = (x^k - x^*). \quad (2.14)$$

Tomando a projeção sobre $\Omega(x^k, \gamma)$ e usando o fato que a projeção é uma aplicação não expansiva, segue que

$$\left\| P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k)) - P_{\Omega(x^k, \gamma)}\left(x^k + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k)\right) \right\| \leq \|x^k - x^*\|. \quad (2.15)$$

Em virtude da Proposição 2.48, temos que

$$\sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) \in N_{\Omega(x^k, \gamma)}(x^k).$$

Como $\Omega(x^k, \gamma)$ é convexo e fechado, a Proposição A.40 garante que

$$P_{\Omega(x^k, \gamma)}\left(x^k + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k)\right) = x^k. \quad (2.16)$$

Substituindo (2.16) em (2.15), obtemos

$$\|P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k)) - x^k\| \leq \|x^k - x^*\|. \quad (2.17)$$

Como $\|x^k - x^*\| \rightarrow 0$, passando ao limite em (2.17), concluímos que x^* satisfaz AGP(γ). $\square$

O próximo Teorema mostra que a condição AGP é mais forte do que a condição AKKT.

Teorema 2.51. *Seja $x^* \in \Omega$. Se x^* satisfaz a condição AGP, então x^* satisfaz AKKT.*

Demonstração. Como x^* satisfaz a condição AGP, então pela definição existe uma sequência $x^k \rightarrow x^*$ tal que

$$y^k := P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k)) - x^k \rightarrow 0.$$

para algum $\gamma \in [-\infty, 0)$.

A projeção $P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k)) = y^k + x^k$ é a única solução do problema de otimização:

$$\text{minimizar } \frac{1}{2} \|z - x^k + \nabla f(x^k)\|^2, \text{ sujeito a } z \in \Omega(x^k, \gamma).$$

Pela linearidade das restrições, as condições KKT valem em $y^k + x^k$, ou seja, existe $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ tal que

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

$$\mu_j^k (\langle \nabla g_j(x^k), y^k \rangle) = 0, \quad \text{se } 0 \leq g_j(x^k), \quad (2.19)$$

$$\mu_j^k (g_j(x^k) + \langle \nabla g_j(x^k), y^k \rangle) = 0, \quad \text{se } \gamma < g_j(x^k) < 0, \quad (2.20)$$

$$\mu_j^k = 0, \quad \text{caso contrário} \quad (2.21)$$

Como $y^k \rightarrow 0$, então por (2.18), obtemos

$$\lim_{k \rightarrow \infty} \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) = 0.$$

Se $j \notin A(x^*)$, então $g_j(x^*) < 0$. Pela continuidade de g , temos que $g_j(x^k) + \langle \nabla g_j(x^k), y^k \rangle < 0$ para cada k suficientemente grande. Logo, usando a relação (2.20), obtemos que $\mu_j^k = 0$ para todo $j \notin A(x^*)$. Portanto vale AKKT. $\square$

Contra-Exemplo 2.52 (AKKT não implica AGP). Considere o problema

$$\begin{aligned} &\text{minimizar } f(x) = -x_2, \\ &\text{sujeito a } h_1(x) = x_1x_2 = 0, \\ &\quad g_1(x) = -x_1 \leq 0. \end{aligned}$$

O ponto $x^* = (0, 1)$ satisfaz a condição AKKT. De fato, considere as sequências $x^k = (1/k, 1)$, $\lambda_1^k = \mu_1^k = k$. Neste caso, temos $A(x^*) = \{1\}$. Logo, para todo $k \in \mathbb{N}$, temos

$$\nabla f(x^k) + \lambda_1^k \nabla h_1(x^k) + \mu_1^k \nabla g_1(x^k) = \begin{pmatrix} 0 \\ -1 \end{pmatrix} + k \begin{pmatrix} 1 \\ 1/k \end{pmatrix} + k \begin{pmatrix} -1 \\ 0 \end{pmatrix} = 0.$$

Portanto, x^* satisfaz AKKT. Entretanto, x^* não satisfaz AGP. Com efeito, considere uma sequência $x^k \rightarrow x^*$ tal que $x_1^k \geq 0$ para todo k . Dado $x = (x_1, x_2) \in \Omega(x^k, \gamma)$. Pela definição de $\Omega(x^k, \gamma)$, temos

$$\nabla h_1(x^k)^T(x - x^k) = 0 \quad \text{e} \quad \nabla g_1(x^k)^T(x - x^k) \leq 0.$$

Assim,

$$\nabla h_1(x^k)^T(x - x^k) = \begin{pmatrix} x_2^k & x_1^k \end{pmatrix} \begin{pmatrix} x_1 - x_1^k \\ x_2 - x_2^k \end{pmatrix} = x_2^k(x_1 - x_1^k) + x_1^k(x_2 - x_2^k) = 0,$$

o que implica $(x_2 - x_2^k) = -\frac{x_2^k}{x_1^k}(x_1 - x_1^k)$. Observe que esta equação é a equação da reta tangente a $h_1(x) = h_1(x^k)$ no ponto x^k . Por outro lado, temos

$$\nabla g_1(x^k)^T(x - x^k) = \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} x_1 - x_1^k \\ x_2 - x_2^k \end{pmatrix} \leq 0 \Leftrightarrow x_1 \geq x_1^k \geq 0.$$

Logo, $\Omega(x^k, \gamma)$ é a interseção entre o semi-plano $x_1 \geq 0$ e a reta tangente a $h_1(x) = h_1(x^k)$ no ponto x^k . Quando $x^k \rightarrow x^*$, a reta tangente tende para a reta vertical. Portanto, $P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k)) - x^k$ tende para $(0, 1)$. Analogamente, se $x_1^k < 0$, o conjunto $\Omega(x^k, \gamma)$ é a interseção do semi-plano $x_1 \geq x_1^k$ com a reta tangente a $h_1(x) = x_1^k x_2^k$ no ponto x^k . Logo, $P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k)) - x^k$ tende para $(0, 1)$. Portanto, para qualquer sequência $x^k \rightarrow x^*$, temos que $P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k)) - x^k$ não tende a zero. Como consequência, x^* não satisfaz AGP.

Teorema 2.53 ([51], TEOREMA 3.1). *Considere $\gamma \in [-\infty, 0]$ e suponha que as funções f , g_j são convexas para todo $j = 1, \dots, p$ e h é afim. Sejam $x^* \in \Omega$ e $\{x^k\}$ uma sequência AGP(γ) satisfazendo $h(x^k) = 0$. Então, x^* é um minimizador global de (1.2).*

Demonstração. Vejamos primeiro que $\Omega \subset \Omega(x^k, \gamma)$ para qualquer $x^k \in \mathbb{R}^n$. Considere $y \in \Omega$. Se $g_j(x^k) < 0$, pela convexidade de g_j , segue que

$$0 \geq g_j(y) \geq g_j(x^k) + \nabla g_j(x^k)^T(y - x^k). \quad (2.22)$$

Por outro lado, se $g_j(x^k) \geq 0$, temos

$$0 \geq g_j(y) \geq g_j(x^k) + \nabla g_j(x^k)^T(y - x^k) \geq \nabla g_j(x^k)^T(y - x^k). \quad (2.23)$$

Como h é afim, então $h(x) = Ax + b$ com $A \in \mathbb{R}^{m \times n}$ e $b \in \mathbb{R}^m$. Além disso, $h(x^k) = h(y) = 0$, logo

$$\nabla h(x^k)^T(y - x^k) = A(y - x^k) = (Ay + b) - (Ax^k + b) = 0$$

e, portanto, $\nabla h_i(x^k)^T(y - x^k) = 0$ para todo $i = 1, \dots, m$.

Logo, pelas relações (2.22) e (2.23), segue que $y \in \Omega(x^k, \gamma)$.

Defina $y^k := P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k))$. Logo, y^k é a única solução do problema

$$\text{minimizar } \frac{1}{2}\|y - x^k + \nabla f(x^k)\|^2 \text{ sujeito a } y \in \Omega(x^k, \gamma).$$

Assim, temos que

$$\begin{aligned} \frac{1}{2}\|y - x^k + \nabla f(x^k)\|^2 &= \frac{1}{2}(y - x^k + \nabla f(x^k))^T(y - x^k + \nabla f(x^k)) \\ &= \frac{1}{2}\|y - x^k\|^2 + \nabla f(x^k)^T(y - x^k) + \frac{1}{2}\|\nabla f(x^k)\|^2. \end{aligned}$$

Como y^k é solução, obtemos

$$\nabla f(x^k)^T(y - x^k) + \frac{1}{2}\|y - x^k\|^2 \geq \nabla f(x^k)^T(y^k - x^k) + \frac{1}{2}\|y^k - x^k\|^2.$$

Passando ao limite na desigualdade acima e usando o fato que $\|y^k - x^k\| \rightarrow 0$, obtemos

$$\nabla f(x^*)^T(y - x^*) + \frac{1}{2}\|y - x^*\|^2 \geq 0.$$

Como Ω é convexo, tomando $y = x^* + t(y - x^*)$ na desigualdade acima, segue-se que para todo $t \in [0, 1]$, tem-se

$$t\nabla f(x^*)^T(y - x^*) + \frac{t^2}{2}\|y - x^*\|^2 \geq 0.$$

Dividindo por t e tomando o limite quando $t \rightarrow 0$, obtemos

$$\nabla f(x^*)^T(y - x^*) \geq 0.$$

Como consequência do Corolário A.36 segue que x^* é minimizador global de f . □

Proposição 2.54. *A condição AGP implica em “KKT ou não-CPLD”.*

Demonstração. Seja $x^* \in \Omega$ satisfazendo a condição AGP. Logo existe uma sequência $x^k \rightarrow x^*$ tal que

$$P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k)) - x^k \rightarrow 0,$$

para algum $\gamma \in [-\infty, 0)$. Defina $y^k := P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k))$. Logo y^k resolve o problema

$$\text{minimizar } \frac{1}{2}\|z - x^k + \nabla f(x^k)\|^2 \text{ sujeito a } z \in \Omega(x^k, \gamma).$$

Como $\Omega(x^k, \gamma)$ é definido por restrições lineares, as condições KKT valem em y^k , ou seja, existem $\lambda^k \in \mathbb{R}^m$ e $\mu^k \in \mathbb{R}_+^p$ tais que

$$(y^k - x^k) + \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) = 0, \quad (2.24)$$

$$\mu_j^k (\langle \nabla g_j(x^k), y^k - x^k \rangle) = 0, \quad \text{se } 0 \leq g_j(x^k), \quad (2.25)$$

$$\mu_j^k (g_j(x^k) + \langle \nabla g_j(x^k), y^k - x^k \rangle) = 0, \quad \text{se } \gamma < g_j(x^k) < 0, \quad (2.26)$$

$$\mu_j^k = 0, \quad \text{caso contrário} \quad (2.27)$$

Se $g_j(x^*) < 0$, pela continuidade de g_j e usando o fato de que $y^k - x^k \rightarrow 0$, obtemos $g_j(x^k) + \langle \nabla g_j(x^k), y^k - x^k \rangle < 0$ para todo k suficientemente grande. Assim, segue de (2.26) que $\mu_j^k = 0$ para k suficientemente grande. Logo, usando a relação (2.24), obtemos

$$(y^k - x^k) + \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) = 0.$$

Pelo Lema de Carathéodory, existem $\mathcal{I}_k \subset \{1, \dots, m\}$, $\mathcal{J}_k \subset A(x^*)$, $\tilde{\lambda}^k \in \mathbb{R}^m$, $\tilde{\mu}^k \geq 0$ tais que

$$(y^k - x^k) + \nabla f(x^k) + \sum_{i \in \mathcal{I}_k} \tilde{\lambda}_i^k \nabla h_i(x^k) + \sum_{j \in \mathcal{J}_k} \tilde{\mu}_j^k \nabla g_j(x^k) = 0 \quad \text{e} \quad (2.28)$$

$$\{\nabla h_i(x^k)\}_{i \in \mathcal{I}_k} \cup \{\nabla g_j(x^k)\}_{j \in \mathcal{J}_k} \quad \text{é PLI.}$$

Podemos tomar uma subsequência tal que $\mathcal{I}_k = \{1, \dots, m\}$ e $\mathcal{J}_k = A(x^*)$. Se $\{(\tilde{\lambda}_i^k, \tilde{\mu}_j^k)\}$ é limitada, passando ao limite em subsequência convergente, concluímos que x^* satisfaz KKT. Caso contrário, defina

$$M_k = \max\{\|\tilde{\lambda}^k\|_\infty, \|\tilde{\mu}^k\|_\infty\}.$$

Assim, temos que

$$M_k \rightarrow \infty \quad \text{e} \quad \left\| \left(\frac{\tilde{\lambda}^k}{M_k}, \frac{\tilde{\mu}^k}{M_k} \right) \right\|_\infty = 1.$$

Logo, passando a uma subsequência, podemos supor que $(\frac{\tilde{\lambda}^k}{M_k}, \frac{\tilde{\mu}^k}{M_k}) \rightarrow (\hat{\lambda}, \hat{\mu})$ com $\|(\hat{\lambda}, \hat{\mu})\|_\infty = 1$ e $\hat{\mu} \geq 0$. Assim, dividindo (2.28) por M_k e tomando o limite, obtemos $\sum_{i=1}^m \hat{\lambda}_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \hat{\mu}_j \nabla g_j(x^*) = 0$ com $(\hat{\lambda}, \hat{\mu}) \neq 0$, uma contradição com CPLD. $\square$

2.2.3 Condição sequencial SAKKT

Definição 2.55 (SAKKT, [39]). Dizemos que $x^* \in \Omega$ satisfaz a condição *Strong Approximate KKT (SAKKT)* se existem sequências $\{x^k\} \subset \mathbb{R}^n$ e $\{(\lambda^k, \mu^k)\} \subset \mathbb{R}^m \times \mathbb{R}_+^p$ tais que:

(i) $x^k \rightarrow x^*$;

(ii) $\lim_{k \rightarrow \infty} \left(\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \right) = 0$;

(iii) $g_j(x^k) < 0 \Rightarrow \mu_j^k = 0$ para todo $j = 1, \dots, p$ e para todo $k \in \mathbb{N}$.

Teorema 2.56. *Se $x^* \in \Omega$ satisfaz a condição SAKKT, então x^* satisfaz $AGP(\gamma)$ para qualquer $\gamma \in [-\infty, 0]$.*

Demonstração. Suponha que x^* satisfaz a condição SAKKT e considere qualquer $\gamma \in [-\infty, 0]$. Logo, pela definição, existem sequências $\{x^k\} \subset \mathbb{R}^n$ com $x^k \rightarrow x^*$ e $\{(\lambda^k, \mu^k)\} \subset \mathbb{R}^m \times \mathbb{R}_+^p$ tais que

$$\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j:0 \leq g_j(x^k)} \mu_j^k \nabla g_j(x^k) \rightarrow 0.$$

Assim,

$$x^k - \nabla f(x^k) - \left(x^k + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j:0 \leq g_j(x^k)} \mu_j^k \nabla g_j(x^k) \right) \rightarrow 0.$$

Tomando a projeção sobre o conjunto $\Omega(x^k, \gamma)$ e usando a não-expansividade da projeção, obtemos

$$P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k)) - P_{\Omega(x^k, \gamma)} \left(x^k + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j:0 \leq g_j(x^k)} \mu_j^k \nabla g_j(x^k) \right) \rightarrow 0. \quad (2.29)$$

Pelo mesmo argumento usado na demonstração do Teorema 2.50, temos que

$$P_{\Omega(x^k, \gamma)} \left(x^k + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j:0 \leq g_j(x^k)} \mu_j^k \nabla g_j(x^k) \right) = x^k. \quad (2.30)$$

Substituindo (2.30) em (2.29), obtemos

$$P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k)) - x^k \rightarrow 0.$$

Portanto vale $AGP(\gamma)$. □

Teorema 2.57. *A condição SAKKT é equivalente à condição $AGP(0)$.*

Demonstração. Pelo Teorema anterior, basta mostrar que $AGP(0)$ implica em SAKKT. Pela definição de $AGP(0)$, temos que existe uma sequência $\{x^k\}$ com $x^k \rightarrow 0$ tal que

$$y^k := P_{\Omega(x^k, 0)}(x^k - \nabla f(x^k)) - x^k \rightarrow 0.$$

Logo, $y^k + x^k$ é a única solução do problema

$$\text{minimizar } \frac{1}{2} \|z - x^k + \nabla f(x^k)\|^2 \text{ sujeito a } z \in \Omega(x^k, 0).$$

Como as restrições deste problema são lineares, as condições de KKT valem em $y^k + x^k$, isto é, existem $\lambda^k \in \mathbb{R}^m$ e $\mu^k \in \mathbb{R}_+^p$ tais que

$$y^k + \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) = 0, \quad (2.31)$$

$$\mu_j^k (\langle \nabla g_j(x^k), y^k \rangle) = 0, \quad \text{se } 0 \leq g_j(x^k), \quad (2.32)$$

$$\mu_j^k = 0, \quad \text{caso contrário.} \quad (2.33)$$

Por (2.31), temos que

$$\lim_{k \rightarrow \infty} \left(\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \right) = 0.$$

Pelas relações (2.32) e (2.33), concluímos que $g_j(x^k) < 0 \Rightarrow \mu_j^k = 0$, completando a prova. $\square$

Corolário 2.58. *A condição SAKKT é uma condição de otimalidade.*

2.2.4 Condições sequenciais C-AGP e L-AGP

Algumas variações da condição AGP tem sido utilizada na literatura no contexto de métodos de restauração inexata e programação matemática. Essas variações consistem em incluir algumas restrições lineares na definição de Ω , impondo que a sequência $\{x^k\}$ associada satisfaça essas restrições.

Definição 2.59 (C-AGP, [11]). Suponha que as funções h_i são afins para todo $i \in \{r+1, \dots, m\}$ e g_j são convexas para todo $j \in \{q+1, \dots, p\}$. Considere o conjunto convexo e fechado, Ω_C , definido por $h_i(x) = 0$ e $g_j(x) \leq 0$ para todo $i \in \{r+1, \dots, m\}$ e $j \in \{q+1, \dots, p\}$. Dizemos que x^* satisfaz a *condição convex-AGP (C-AGP)* se existe uma sequência $\{x^k\} \subset \Omega_C$ com $x^k \rightarrow x^*$ tal que

$$P_{\hat{\Omega}(x^k, \gamma) \cap \Omega_C}(x^k - \nabla f(x^k)) - x^k \rightarrow 0,$$

onde

$$\hat{\Omega}(x^k, \gamma) := \left\{ z \in \mathbb{R}^n : \begin{array}{ll} \langle \nabla h_i(x^k), z - x^k \rangle = 0, & i = 1, \dots, r \\ \langle \nabla g_j(x^k), z - x^k \rangle \leq 0, & \text{se } 0 \leq g_j(x^k), j = 1, \dots, q \\ g_j(x^k) + \langle \nabla g_j(x^k), z - x^k \rangle \leq 0, & \text{se } \gamma < g_j(x^k) < 0, j = 1, \dots, q \end{array} \right\}.$$

O Teorema a seguir mostra que se as restrições em Ω_C cumprem alguma condição de qualificação, então C-AGP é uma condição de otimalidade.

Teorema 2.60. *Seja x^* um minimizador local do problema (1.2) e, suponha que, para todo $x \in \Omega_C$, as restrições h_i , $i = r+1, \dots, m$, e g_j , $j = q+1, \dots, p$, cumprem alguma condição de qualificação. Então, x^* satisfaz C-AGP.*

Demonstração. Seja x^* um minimizador local. Logo, existe $\delta > 0$ tal que x^* é a única solução global do problema

$$\begin{array}{ll} \text{minimizar} & f(x) + \frac{1}{2} \|x - x^*\|^2, \\ \text{sujeito a} & h(x) = 0, \\ & g(x) \leq 0, \\ & \|x - x^*\| \leq \delta. \end{array}$$

Considere uma sequência $\{x^k\}$ gerada pelo algoritmo de penalidade externa aplicado ao problema acima. Pelo mesmo argumento usado na demonstração do Teorema 2.41,

temos que $x^k \rightarrow x^*$. Logo, $\|x^k - x^*\| < \delta$ para k suficientemente grande. Suponha que $\rho_k \rightarrow +\infty$ e, para cada k , seja x^k um minimizador global de

$$\begin{aligned} &\text{minimizar} && f(x) + \frac{1}{2}\|x - x^*\|^2 + \frac{\rho_k}{2} \left(\sum_{i=1}^r h_i(x)^2 + \sum_{j=1}^q g_j^+(x)^2 \right), \\ &\text{sujeito a} && \|x - x^*\| \leq \delta. \end{aligned}$$

O gradiente da função objetivo deve ser nulo em x^k , isto é,

$$\nabla f(x^k) + \sum_{i=1}^r \rho_k h_i(x^k) \nabla h_i(x^k) + \sum_{j=1}^q \rho_k g_j^+(x^k) \nabla g_j(x^k) + (x^k - x^*) = 0.$$

Além disso, uma vez que as restrições h_i e g_j para $i = r+1, \dots, m$ e $j = q+1, \dots, p$ cumprem alguma condição de qualificação, as condições KKT para o subproblema devem ser mantidas. Defina

$$A_k = \{j \in \{q+1, \dots, p\} : g_j(x^k) = 0\}.$$

Portanto, para k suficientemente grande, existem λ_i^k , $i = r+1, \dots, m$ e μ_j^k , $j = q+1, \dots, p$ tais que

$$\begin{aligned} \nabla f(x^k) + \sum_{i=1}^r \rho_k h_i(x^k) \nabla h_i(x^k) + \sum_{j=1}^q \rho_k g_j^+(x^k) \nabla g_j(x^k) + (x^k - x^*) + \\ + \sum_{i=r+1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A_k} \mu_j^k \nabla g_j(x^k) = 0. \end{aligned}$$

Logo, podemos reescrever como

$$\begin{aligned} x^k - \nabla f(x^k) - \left[x^k + \sum_{i=1}^r \rho_k h_i(x^k) \nabla h_i(x^k) + \sum_{j=1}^q \rho_k g_j^+(x^k) \nabla g_j(x^k) + \right. \\ \left. + \sum_{i=r+1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A_k} \mu_j^k \nabla g_j(x^k) \right] = x^k - x^*. \end{aligned}$$

Pela propriedade da não-expansividade da projeção, obtemos

$$\begin{aligned} \left\| P_{\hat{\Omega}(x^k, \gamma) \cap \Omega_C}(x^k - \nabla f(x^k)) - P_{\hat{\Omega}(x^k, \gamma) \cap \Omega_C} \left[x^k + \sum_{i=1}^r \rho_k h_i(x^k) \nabla h_i(x^k) + \sum_{j=1}^q \rho_k g_j^+(x^k) \nabla g_j(x^k) + \right. \right. \\ \left. \left. + \sum_{i=r+1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A_k} \mu_j^k \nabla g_j(x^k) \right] \right\| \leq \|x^k - x^*\|. \end{aligned} \quad (2.34)$$

Além disso, temos

$$\begin{aligned} v(x^k) := \left[\sum_{i=1}^r \rho_k h_i(x^k) \nabla h_i(x^k) + \sum_{j=1}^q \rho_k g_j^+(x^k) \nabla g_j(x^k) + \sum_{i=r+1}^m \lambda_i^k \nabla h_i(x^k) + \right. \\ \left. + \sum_{j \in A_k} \mu_j^k \nabla g_j(x^k) \right] \in N_{\hat{\Omega}(x^k, \gamma) \cap \Omega_C}(x^k). \end{aligned}$$

De fato, seja $y \in \hat{\Omega}(x^k, \gamma) \cap \Omega_C$. Pela convexidade de g_j , para todo $j \in A_k$, temos

$$0 \geq g_j(y) \geq g_j(x^k) + \nabla g_j(x^k)^T (y - x^k) = \nabla g_j(x^k)^T (y - x^k). \quad (2.35)$$

Como h_i é afim, digamos $h_i(x) = a_i^T x + b_i$, para todo $i = r+1, \dots, m$, temos

$$\begin{aligned} \nabla h_i(x^k)^T(y - x^k) &= a_i^T(y - x^k) = (a_i^T y + b_i) - (a_i^T x^k + b_i) \\ &= h_i(y) - h_i(x^k) = 0. \end{aligned} \quad (2.36)$$

Logo, pela definição de $\widehat{\Omega}(x^k, \gamma)$ e usando as relações (2.35) e (2.36), obtemos

$$\begin{aligned} \langle v(x^k), y - x^k \rangle &= \left\langle \sum_{i=1}^r \rho_k h_i(x^k) \nabla h_i(x^k), y - x^k \right\rangle + \left\langle \sum_{j=1}^q \rho_k g_j^+(x^k) \nabla g_j(x^k), y - x^k \right\rangle + \\ &\quad + \left\langle \sum_{i=r+1}^m \lambda_i^k \nabla h_i(x^k), y - x^k \right\rangle + \left\langle \sum_{j \in A_k} \mu_j^k \nabla g_j(x^k), y - x^k \right\rangle \leq 0. \end{aligned}$$

Portanto, sendo $\widehat{\Omega}(x^k, \gamma) \cap \Omega_C$ convexo fechado, obtemos

$$\begin{aligned} P_{\widehat{\Omega}(x^k, \gamma) \cap \Omega_C} \left[x^k + \sum_{i=1}^r \rho_k h_i(x^k) \nabla h_i(x^k) + \sum_{j=1}^q \rho_k g_j^+(x^k) \nabla g_j(x^k) + \right. \\ \left. + \sum_{i=r+1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A_k} \mu_j^k \nabla g_j(x^k) \right] = x^k. \end{aligned}$$

Substituindo esta última equação em (2.34), obtemos

$$\|P_{\widehat{\Omega}(x^k, \gamma) \cap \Omega_C}(x^k - \nabla f(x^k)) - x^k\| \leq \|x^k - x^*\| \rightarrow 0,$$

completando a prova. $\square$

Em [11], os autores mostram que C-AGP é uma condição de otimalidade forte no sentido que implica “KKT ou não-MFCQ”.

Contra-Exemplo 2.61 (AGP não implica C-AGP). Considere o seguinte problema

$$\text{minimizar } f(x) = x_2 \quad \text{sujeito a } h_1(x) = x_1 x_2 = 0, \quad g_1(x) = -x_1 \leq 0.$$

O ponto $x^* = (0, 1)$ satisfaz a condição AGP. De fato, considere a sequência $x^k = (-1/k, 1)$. Neste caso, temos que $-\nabla f(x^k) \in N_{\Omega(x^k, \gamma)}(x^k)$. De fato, seja $x = (x_1, x_2) \in \Omega(x^k, \gamma)$. Assim, temos

$$\nabla h_1(x^k)^T(x - x^k) = \begin{pmatrix} 1 & -1/k \end{pmatrix} \begin{pmatrix} x_1 + 1/k \\ x_2 - 1 \end{pmatrix} = 0 \Leftrightarrow x_1 + 1/k = (1/k)(x_2 - 1).$$

$$\nabla g_1(x^k)^T(x - x^k) = \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} x_1 + 1/k \\ x_2 - 1 \end{pmatrix} \leq 0 \Leftrightarrow x_1 + 1/k \geq 0.$$

Das relações acima, segue que $x_1 + \frac{1}{k} = \frac{1}{k}(x_2 - 1) \geq 0$, ou seja, $1 - x_2 \leq 0$. Assim, dado qualquer $x = (x_1, x_2) \in \Omega(x^k, \gamma)$, temos

$$-\nabla f(x^k)^T(x - x^k) = \begin{pmatrix} 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 + 1/k \\ x_2 - 1 \end{pmatrix} = 1 - x_2 \leq 0.$$

Logo, $-\nabla f(x^k) \in N_{\Omega(x^k, \gamma)}(x^k)$ e, como consequência, $P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k)) = x^k$. Portanto, x^* satisfaz a condição AGP. Entretanto, o mesmo argumento usado no contra-exemplo 2.52 pode ser usado para mostrar que x^* não satisfaz C-AGP.

Dado um problema de otimização não linear com restrições lineares, podemos considerar outra variação da condição AGP, conhecida como AGP linear [11]. Essa condição consiste em separar as restrições lineares das restrições não lineares. Suponha que as funções h_i e g_j são lineares para todo $i \in \{r+1, \dots, m\}$ e para todo $j \in \{q+1, \dots, p\}$. Denotemos por Ω_L o conjunto dos pontos que satisfazem todas as restrições lineares.

Definição 2.62. Dizemos que um ponto viável x^* satisfaz a *condição AGP linear (L-AGP)*, se existe uma sequência $\{x^k\} \subset \Omega_L$ com $x^k \rightarrow x^*$ tal que

$$P_{\hat{\Omega}(x^k, \gamma) \cap \Omega_L}(x^k - \nabla f(x^k)) - x^k \rightarrow 0.$$

Em outras palavras, um ponto satisfaz L-AGP se satisfaz C-AGP no caso em que as funções g_j são afins para todo $j = q+1, \dots, p$.

Assim como AGP, a condição L-AGP também é uma condição sequencial de otimalidade.

Teorema 2.63. Se $x^* \in \Omega$ é um minimizador local, então x^* satisfaz L-AGP.

Demonstração. Basta seguir os mesmos passos da demonstração do Teorema 2.50. $\square$

No contra-exemplo a seguir, veremos que AGP não implica L-AGP. Por outro lado, se x^* satisfaz L-AGP, a sequência correspondente $\{x^k\}$ pode ser usada para mostrar que AGP se verifica em x^* . Em outras palavras, L-AGP é estritamente mais forte do que AGP.

Contra-Exemplo 2.64 (AGP não implica L-AGP). Considerando o mesmo problema do contra-exemplo 2.61, temos que o ponto $x^* = (0, 1)$ cumpre a condição AGP. Entretanto, o mesmo argumento feito no contra-exemplo 2.52 pode ser usado para mostrar que x^* não satisfaz L-AGP.

2.2.5 Condição sequencial CAKKT

A seguir, apresentamos uma condição sequencial mais forte do que as apresentadas anteriormente, denominada *Complementary Approximate KKT (CAKKT)*, introduzida por Andreani, Martínez e Svaiter em [15].

Definição 2.65 (CAKKT, [15]). Dizemos que um ponto viável $x \in \Omega$ cumpre a condição *Complementary Approximate KKT (CAKKT)*, se existem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$ e $\{\mu^k\} \subset \mathbb{R}_+^p$ com $x^k \rightarrow x$ tais que

- $\lim_{k \rightarrow \infty} \left(\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \right) = 0,$
- $\lim_{k \rightarrow \infty} \lambda_i^k h_i(x^k) = 0$ para todo $i = 1, \dots, m,$
- $\lim_{k \rightarrow \infty} \mu_j^k g_j(x^k) = 0$ para todo $j = 1, \dots, p.$

Claramente, CAKKT é mais forte do que SAKKT, uma vez que o último não requer qualquer exigência com respeito à complementaridade relacionada às restrições de igualdade. As duas últimas condições de complementaridade podem ser substituídas por

$$\lim_{k \rightarrow \infty} \left(\sum_{i=1}^m |\lambda_i^k h_i(x^k)| + \sum_{j=1}^p |\mu_j^k g_j(x^k)| \right) = 0. \quad (2.37)$$

A condição CAKKT é uma condição de otimalidade como mostra o próximo teorema.

Teorema 2.66. *Se $x^* \in \Omega$ é um minimizador local, então x^* satisfaz CAKKT.*

Demonstração. Seja x^* um minimizador local e considere uma sequência $\{\rho_k\} \subset \mathbb{R}_+$ tal que $\rho_k \rightarrow +\infty$. Seguindo os mesmos passos da demonstração do Teorema 2.41, segue que existe uma sequência $x^k \rightarrow x^*$ tal que

$$\nabla f(x^k) + \sum_{i=1}^m \rho_k h_i(x^k) \nabla h_i(x^k) + \sum_{j=1}^p \rho_k g_j^+(x^k) \nabla g_j(x^k) + (x^k - x^*) = 0. \quad (2.38)$$

Como x^* é viável e x^k é minimizador do problema

$$\begin{aligned} &\text{minimizar} \quad f(x) + \frac{1}{2} \|x - x^*\|^2 + \frac{\rho_k}{2} \left(\sum_{i=1}^m h_i(x)^2 + \sum_{j=1}^p g_j^+(x)^2 \right) \\ &\text{sujeito a} \quad \|x - x^*\| \leq \delta, \end{aligned}$$

temos que

$$\sum_{i=1}^m \rho_k h_i(x^k)^2 + \sum_{j=1}^p \rho_k g_j^+(x^k)^2 \leq 2 \left(f(x^*) - f(x^k) - \frac{1}{2} \|x^k - x^*\|^2 \right). \quad (2.39)$$

Defina $\lambda_i^k = \rho_k h_i(x^k)$ para todo $i = 1, \dots, m$ e $\mu_j^k = \rho_k g_j^+(x^k)$ para todo $j = 1, \dots, p$. Como $x^k \rightarrow x^*$, pela continuidade de f , o lado direito da desigualdade acima converge a zero, logo de (2.38) e (2.39), segue

$$\begin{aligned} \lim_{k \rightarrow \infty} \left(\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \right) &= 0, \\ \lim_{k \rightarrow \infty} \left(\sum_{i=1}^m |\lambda_i^k h_i(x^k)| + \sum_{j=1}^p |\mu_j^k g_j(x^k)| \right) &= 0, \end{aligned}$$

e o resultado segue. $\square$

Sejam ε_{feas} , ε_{opt} e ε_{mult} as tolerâncias referentes à viabilidade, à otimalidade e à condição (2.37), respectivamente. O critério de parada natural associado à condição CAKKT é declarar convergência quando

$$\begin{aligned} \|\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k)\| &\leq \varepsilon_{opt}, \\ \|h(x^k)\| &\leq \varepsilon_{feas}, \quad \|g^+(x^k)\| \leq \varepsilon_{feas}, \\ |\lambda_i^k h_i(x^k)| &\leq \varepsilon_{mult}, \quad |\mu_j^k g_j(x^k)| \leq \varepsilon_{mult}, \quad i = 1, \dots, m; \quad j = 1, \dots, p. \end{aligned}$$

Teorema 2.67. *Suponha que $x^* \in \Omega$ cumpre a condição CAKKT. Então, x^* satisfaz a condição AGP(γ) para todo $\gamma \in [-\infty, 0]$. Além disso, se a sequência $\{x^k\}$ associada à definição de CAKKT satisfaz todas as restrições lineares, então x^* satisfaz L-AGP.*

Demonstração. Sabemos que CAKKT implica em SAKKT e que SAKKT implica AGP, logo a primeira parte segue. Para provar a segunda parte, basta fazer uma adaptação da prova do Teorema 2.56, usando o fato de que $\{x^k\}$ satisfaz todas as restrições lineares. $\square$

O próximo Teorema mostra que, no caso convexo, a condição CAKKT é uma condição suficiente para a otimalidade. Como consequência, em problemas convexos, CAKKT é uma condição necessária e suficiente para soluções globais.

Teorema 2.68 ([15], TEOREMA 4.2). *Suponha que as funções f , g_j , $j = 1, \dots, p$, são convexas e h é afim. Se $x^* \in \Omega$ satisfaz a condição CAKKT, então x^* é um minimizador global de (1.2).*

Demonstração. Seja $x^* \in \Omega$ e suponha que x^* satisfaz CAKKT. Logo, existe $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+^p$ com $x^k \rightarrow x^*$ tal que

$$\lim_{k \rightarrow \infty} \left(\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \right) = 0 \quad \text{e} \quad (2.40)$$

$$\lim_{k \rightarrow \infty} \left(\sum_{i=1}^m |\lambda_i^k h_i(x^k)| + \sum_{j=1}^p |\mu_j^k g_j(x^k)| \right) = 0. \quad (2.41)$$

Seja $z \in \Omega$. Pela convexidade das funções f , g_j e pela linearidade de h , temos que, para todo $k \in \mathbb{N}$,

$$\begin{aligned} f(z) &\geq f(x^k) + \nabla f(x^k)^T (z - x^k), \\ g_j(z) &\geq g_j(x^k) + \nabla g_j(x^k)^T (z - x^k), \quad j = 1, \dots, p, \\ h_i(z) &= h_i(x^k) + \nabla h_i(x^k)^T (z - x^k) = 0, \quad i = 1, \dots, m. \end{aligned}$$

Assim, como $h(z) = 0$ e $g(z) \leq 0$, segue que

$$\begin{aligned} f(z) &\geq f(x^k) + \nabla f(x^k)^T (z - x^k) + \sum_{i=1}^m \lambda_i^k h_i(z) + \sum_{j=1}^p \mu_j^k g_j(z) \\ &\geq f(x^k) + \nabla f(x^k)^T (z - x^k) + \sum_{i=1}^m \lambda_i^k [h_i(x^k) + \nabla h_i(x^k)^T (z - x^k)] + \\ &\quad + \sum_{j=1}^p \mu_j^k [g_j(x^k) + \nabla g_j(x^k)^T (z - x^k)] \\ &= f(x^k) + \left(\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \right)^T (z - x^k) + \\ &\quad + \sum_{i=1}^m \lambda_i^k h_i(x^k) + \sum_{j=1}^p \mu_j^k g_j(x^k). \end{aligned}$$

Passando ao limite na desigualdade acima e usando o fato de que x^* satisfaz CAKKT, obtemos $f(z) \geq f(x^*)$, completando a prova. $\square$

Lema 2.69 ([15], LEMA 3.2). *Suponha que $x^* \in \Omega$ satisfaz a condição CAKKT. Então existem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{(\lambda^k, \mu^k)\} \subset \mathbb{R}^m \times \mathbb{R}_+^p$ com $\mu_j^k = 0$ para $j \notin A(x^*)$ e $x^k \rightarrow x^*$ tais que*

- $\lim_{k \rightarrow \infty} \left(\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) \right) = 0,$
- $\lim_{k \rightarrow \infty} \left(\sum_{i=1}^m |\lambda_i^k h_i(x^k)| + \sum_{j \in A(x^*)} |\mu_j^k g_j(x^k)| \right) = 0.$

2.2.6 Condições sequenciais mais recentes

Mais recentemente, outras condições sequenciais para otimização não linear surgiram na literatura. Nesta subseção apresentamos as condições *positive approximate KKT* (PAKKT) [5] e *positive complementary approximate KKT* (PCAKKT) [7]. Do ponto de vista de otimização, pontos PAKKT são melhores candidatos do que pontos obtidos pela condição AKKT. Outro fato importante é que PAKKT e CAKKT são independentes entre si e, portanto, fornecem características diferentes para os minimizadores.² A condição PCAKKT unifica essas condições, no sentido de que, todo ponto PCAKKT é também um ponto CAKKT e PAKKT.

Definição 2.70 (PAKKT, [5]). Dizemos que um ponto $x^* \in \Omega$ satisfaz a condição *Positive Approximate KKT* (PAKKT) se existem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{(\lambda^k, \mu^k)\} \subset \mathbb{R}^m \times \mathbb{R}_+^p$ com $x^k \rightarrow x^*$ tais que

$$\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \rightarrow 0, \quad (2.42)$$

$$\min\{-g_j(x^k), \mu_j^k\} \rightarrow 0, \quad \forall j = 1, \dots, p, \quad (2.43)$$

$$\lambda_i^k h_i(x^k) > 0 \text{ se } \lim_{k \rightarrow \infty} \frac{|\lambda_i^k|}{\delta^k} > 0, \quad \mu_j^k g_j(x^k) > 0 \text{ se } \lim_{k \rightarrow \infty} \frac{\mu_j^k}{\delta^k} > 0, \quad (2.44)$$

onde $\delta^k := \|(1, \lambda^k, \mu^k)\|_\infty$. Dizemos que $\{x^k\}$ é uma sequência PAKKT associada a x^* .

PAKKT é uma condição necessária de otimalidade, como mostra o próximo Teorema.

Teorema 2.71. *Se x^* é um minimizador local, então x^* satisfaz PAKKT.*

Demonstração. Seja x^* um minimizador local. Considere a aplicação do algoritmo de penalidade externa para o problema

$$\text{minimizar } f(x) + \frac{1}{2} \|x - x^*\|^2 \quad \text{sujeito a } h(x) = 0, \quad g(x) \leq 0, \quad \|x - x^*\| \leq \delta,$$

onde $\delta > 0$ é dado pela definição de minimizador local. Usando os mesmos argumentos utilizados na demonstração do Teorema 2.41, juntamente com o Teorema 2.38, temos que a sequência $\{x^k\}$ gerada pelo algoritmo de penalidade externa converge para x^* . Logo, $\|x^k - x^*\| < \delta$ para todo k suficientemente grande. Das condições de otimalidade do problema penalizado, obtemos

$$\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) + (x^k - x^*) = 0, \quad (2.45)$$

onde $\lambda_i^k := \rho_k h_i(x^k)$, $i = 1, \dots, m$ e $\mu_j^k := \rho_k g_j^+(x^k) \geq 0$ para todo $j = 1, \dots, p$. Como $x^k \rightarrow x^*$, tomando o limite quando $k \rightarrow \infty$ na expressão (2.45), segue que a condição (2.42) se cumpre. Para provar a condição (2.43), note que, se $j \notin A(x^*)$, então pela continuidade de g , temos que $g_j(x^k) < 0$ para todo k suficientemente grande e, portanto, $\min\{-g_j(x^k), \mu_j^k\} = \mu_j^k = 0$ para todo $j \notin A(x^*)$ e para todo k . Caso contrário, para todo k suficientemente grande, temos $\min\{-g_j(x^k), \mu_j^k\} \rightarrow 0$ para todo $j \in A(x^*)$. Finalmente, se $\mu_j^k > 0$, então $g_j(x^k) > 0$ e logo $\mu_j^k g_j(x^k) = \rho_k g_j(x^k)^2 > 0$. Analogamente, se $\lambda_i^k \neq 0$ então $h_i(x^k) \neq 0$ e portanto $\lambda_i^k h_i(x^k) = \rho_k h_i(x^k)^2 > 0$, completando a prova. $\square$

²Andreani et al. [5]

Teorema 2.72. *Toda sequência PAKKT é também uma sequência AKKT. Em particular, todo ponto PAKKT é um ponto AKKT.*

Demonstração. Seja $\{x^k\}$ uma sequência PAKKT tal que $x^k \rightarrow x^*$. Observe que vale a condição $\min\{-g_j(x^k), \mu_j^k\} \rightarrow 0$ para todo j se, e somente se, $\mu_j^k = 0$, para todo $j \notin A(x^*)$. Logo a sequência $\{x^k\}$ pode ser usada para mostrar que x^* satisfaz AKKT. $\square$

Contra-Exemplo 2.73 (CAKKT não implica PAKKT). Considere o problema

$$\text{minimizar } f(x) = \frac{(x_1 - 1)^2}{2} + \frac{(x_2 + 1)^2}{2} \quad \text{sujeito a } g(x) = x_1 x_2 \leq 0.$$

Temos

$$\nabla f(x) + \mu \nabla g(x) = \begin{pmatrix} x_1 - 1 \\ x_2 + 1 \end{pmatrix} + \mu \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}.$$

A origem $x^* = (0, 0)$ satisfaz CAKKT para as sequências $x^k = (-1/k, 1/k)$ e $\mu^k = k$ para todo $k \geq 1$. Além disso, qualquer sequência CAKKT $\{x^k\}$ com $x^k \rightarrow x^*$ satisfaz $\mu^k x_1^k \rightarrow -1$ e $\mu^k x_2^k \rightarrow 1$. Logo, para todo k suficientemente grande, temos $\text{sgn } \mu^k = 1$, $\text{sgn } x_1^k = -1$ e $\text{sgn } x_2^k = 1$ e, portanto, $\text{sgn}(\mu^k x_1^k x_2^k) = -1$. Segue que x^* não é um ponto PAKKT.

Contra-Exemplo 2.74 (PAKKT não implica AGP). Considere o problema de minimizar $f(x) = x_2$ sujeito a $h(x) = x_1^2 x_2 = 0$ e $g(x) = -x_1 \leq 0$. O ponto $x^* = (0, -1)$ satisfaz PAKKT para as sequências $x^k = (1/k, -1)$, $\lambda^k = -k^2$ e $\mu^k = 2k$. De fato, temos

$$\nabla_x \mathcal{L}(x^k, \lambda^k, \mu^k) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + (-k^2) \begin{pmatrix} -2/k \\ 1/k^2 \end{pmatrix} + 2k \begin{pmatrix} -1 \\ 0 \end{pmatrix} = 0.$$

Além disso, $\min\{-g(x^k), \mu^k\} = 1/k \rightarrow 0$, $\delta^k = \|(1, \lambda^k, \mu^k)\|_\infty = k^2$ e portanto $|\lambda^k|/\delta^k = 1 > 0$, $\mu^k/\delta^k = 2/k \rightarrow 0$ com $\lambda^k h(x^k) = 1 > 0$. Portanto x^* satisfaz PAKKT. Para mostrar que x^* não satisfaz AGP, considere qualquer sequência $x^k \rightarrow x^*$ com $x_1^k \neq 0$ e $\gamma \in [-\infty, 0)$. Dado $x = (x_1, x_2) \in \Omega(x^k, \gamma)$, temos

$$\nabla h(x^k)^T (x - x^k) = \begin{pmatrix} 2x_1^k x_2^k & (x_1^k)^2 \end{pmatrix} \begin{pmatrix} x_1 - x_1^k \\ x_2 - x_2^k \end{pmatrix} = 0 \Leftrightarrow (x_2 - x_2^k) = -\frac{2x_2^k}{x_1^k} (x_1 - x_1^k),$$

$$g(x^k) + \nabla g(x^k)^T (x - x^k) = -x_1^k + \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} x_1 - x_1^k \\ x_2 - x_2^k \end{pmatrix} \leq 0 \Leftrightarrow x_1 \geq 0.$$

Assim, $\Omega(x^k, \gamma)$ é a interseção do semi-plano $x_1 \geq 0$ com a reta tangente a $x_1^2 x_2 = (x_1^k)^2 x_2^k$ no ponto x^k . Como $x^k - \nabla f(x^k) = (x_1^k, x_2^k - 1) \rightarrow (0, -2)$, segue que

$$\|P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k)) - x^k\|_\infty \rightarrow 1.$$

Portanto x^* não satisfaz AGP.

Em geral, não existe uma relação de implicação entre as condições CAKKT e PAKKT, ou seja, elas são independentes uma da outra (veja Figura 2.4). De fato, se PAKKT implicasse CAKKT, devido ao fato de CAKKT implicar AGP, teríamos que PAKKT implicaria AGP, o que não acontece pelo contra-exemplo 2.74. Logo, como consequência do contra-exemplo 2.73, segue que CAKKT e PAKKT são independentes. Suponha que um ponto x^* satisfaça ambas as condições CAKKT e PAKKT, ou seja, existe uma sequência PAKKT e uma sequência CAKKT ambas convergindo para x^* . A pergunta é se existe uma sequência comum que caracteriza x^* como um ponto CAKKT e um ponto PAKKT. Um exemplo apresentado em [7] mostra que um ponto pode ser caracterizado por duas sequências distintas, uma CAKKT e outra PAKKT, sem a existência de uma sequência comum. Problemas dessa forma motivaram outra condição sequencial de otimalidade que consiste exatamente na existência de uma sequência $x^k \rightarrow x^*$ que seja simultaneamente CAKKT e PAKKT. Essa condição é chamada de *Positive Complementary Approximate KKT* (PCAKKT) [7].

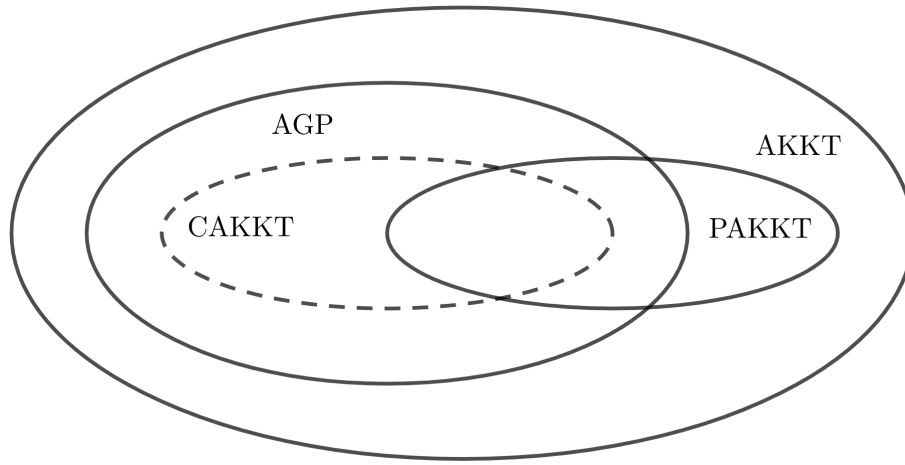


Figura 2.4: As condições CAKKT e PAKKT são independentes uma da outra.

Definição 2.75 (PCAKKT, [7]). Dizemos que um ponto $x^* \in \Omega$ satisfaz a condição *Positive Complementary Approximate KKT* (PCAKKT) se existem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$ e $\{\mu^k\} \subset \mathbb{R}_+^p$ com $x^k \rightarrow x^*$ tais que valem as condições (2.42), (2.44) e

$$\sum_{i=1}^m |\lambda_i^k h_i(x^k)| + \sum_{j=1}^p |\mu_j^k g_j(x^k)| \rightarrow 0. \quad (2.46)$$

Neste caso, dizemos que $\{x^k\}$ é uma sequência PCAKKT associada a x^* .

Observe que todo ponto PCAKKT satisfaz CAKKT. Além disso, todo ponto PCAKKT é também um ponto PAKKT, pois (2.46) implica (2.43). Portanto, pontos PCAKKT herdam todas as propriedades de PAKKT e CAKKT. A condição PCAKKT também é uma condição necessária de otimalidade.

Teorema 2.76. *Se x^* é um minimizador local, então x^* satisfaz PCAKKT.*

Demonstração. Segue os mesmos passos da prova dos Teoremas 2.66 e 2.71. $\square$

Um resultado importante é que todo ponto KKT é também um ponto PCAKKT.

Teorema 2.77 ([7], TEOREMA 2.5). *Se $x^* \in \Omega$ satisfaz KKT, então x^* é um ponto PCAKKT.*

Estudos recentes foram realizados utilizando o método Lagrangiano aumentado de Powell-Hestenes-Rockafellar (PHR) [40, 56, 60] por meio das condições PAKKT e CAKKT (ver [5]). A primeira trata do sinal entre restrições e variáveis duais, uma condição que está relacionada com a condição de otimalidade de Fritz John, enquanto o segundo trata da complementaridade. A condição PCAKKT [7] unifica essas características, melhorando a convergência do método sob essas condições.

Definição 2.78. Dizemos que uma função continuamente diferenciável $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfaz a *desigualdade generalizada de Lojasiewicz* em x^* se existir uma vizinhança V de x^* e uma função contínua $\varphi : V \rightarrow \mathbb{R}$ com $\lim_{x \rightarrow x^*} \varphi(x) = 0$ e $|\Phi(x) - \Phi(x^*)| \leq \varphi(x) \|\nabla \Phi(x)\|$ para todo $x \in V$.

Para introduzir o Algoritmo Lagrangiano aumentado de Powell-Hestenes-Rockafellar, considere a função

$$\mathcal{L}_\rho^{\text{PHR}}(x, \lambda, \mu) := f(x) + \frac{\rho}{2} \left[\sum_{i=1}^m \left(h_i(x) + \frac{\lambda_i}{\rho} \right)^2 + \sum_{j=1}^p \left(\max \left\{ 0, g_j(x) + \frac{\mu_j}{\rho} \right\} \right)^2 \right]$$

com $\rho > 0$ e $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$.

Algoritmo 2.79 (Método Lagrangiano Aumentado PHR). *Sejam $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, $\gamma > 1$, $\rho_1 > 0$ e $\tau \in (0, 1)$. Tome $\{\varepsilon_k\} \subset \mathbb{R}_+$ com $\varepsilon_k \rightarrow 0$. Escolha*

$\lambda^1 \in [\lambda_{\min}, \lambda_{\max}]^m$, $\mu^1 \in [0, \mu_{\max}]^p$. Inicialize com $k \leftarrow 1$.

(1) (Minimização). *Encontre um minimizador aproximado x^k de $\mathcal{L}_{\rho_k}^{\text{PHR}}(x, \lambda^k, \mu^k)$, isto é, determine x^k que satisfaça*

$$\|\nabla \mathcal{L}_{\rho_k}^{\text{PHR}}(x^k, \lambda^k, \mu^k)\| \leq \varepsilon_k.$$

(2) (Atualização do parâmetro de penalidade). *Defina para todo $j = 1, \dots, p$*

$$V_j^k := \max\{\|h(x^k)\|_\infty, \|\min\{-g_j(x^k), \mu_j^k/\rho_k\}\|_\infty\}.$$

Se $k > 1$ e $V^k \leq \tau V^{k-1}$, defina $\rho_{k+1} := \rho_k$. Caso contrário, coloque $\rho_{k+1} \geq \gamma \rho_k$.

(3) (Atualização dos multiplicadores de Lagrange). *Escolha $\lambda^{k+1} \in [\lambda_{\min}, \lambda_{\max}]^m$ e $\mu^{k+1} \in [0, \mu_{\max}]^p$. Faça $k \leftarrow k + 1$ e vá para o passo (1).*

Observe que definindo os multiplicadores como $\lambda_i^{k+1} := \lambda_i^k + \rho_k h_i(x^k)$ para $i = 1, \dots, m$ e $\mu_j^{k+1} := \max\{0, \mu_j^k + \rho_k g_j(x^k)\}$ para $j = 1, \dots, p$, dados pelo gradiente de $\mathcal{L}_\rho^{\text{PHR}}$, o Algoritmo 2.79 se comporta semelhante ao método Lagrangiano aumentado clássico dado pelo Algoritmo 2.44. Em [7] provou-se que o Algoritmo 2.79 gera sequências com convergência a pontos PCAKKT sob a hipótese adicional de que a medida de inviabilidade $\Phi(x) = \|h(x)\|^2 + \|g^+(x)\|^2$ cumpra a desigualdade generalizada de Lojasiewicz. Como consequência, esse método atinge pontos KKT sob a condição de qualificação associada à condição PCAKKT, a condição conhecida como PCAKKT-regular que definiremos na próxima seção. Em [7] os autores também introduziram um novo algoritmo Lagrangiano aumentado primal-dual com convergência a pontos PCAKKT sob uma hipótese adicional.

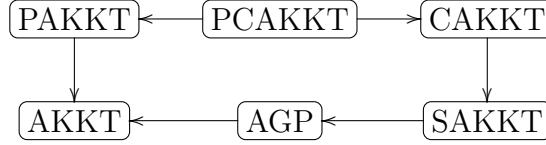


Figura 2.5: Relações entre as principais condições sequenciais para otimização restrita.

2.3 Condições de qualificação fracas e relações

Nesta seção, estamos interessados em estudar condições de qualificação mais fracas que possam ser usadas na análise de convergência de algoritmos práticos. Lembre-se que condições de qualificação fracas geram condições de otimalidade fortes da forma “KKT ou não-CQ”. Dada uma condição sequencial de otimalidade $\mathcal{S}$, podemos associar uma condição de qualificação mínima para garantir a validade das condições KKT. Dizemos que SCQ é uma condição de qualificação estrita para a condição sequencial $\mathcal{S}$, se $\mathcal{S} + \text{SCQ}$ implica KKT. Em outras palavras, se temos algum algoritmo prático que gera sequências cujos pontos limites satisfazem uma condição sequencial de otimalidade $\mathcal{S}$, temos que, sob uma SCQ, o algoritmo gera sequências cujos pontos limites são pontos KKT. O objetivo é encontrar a SCQ mais fraca possível que garante a implicação “ $\mathcal{S} + \text{SCQ} \Rightarrow \text{KKT}$ ”. Assim, a condição de otimalidade “KKT ou não-SCQ” fornece uma medida de força para cada condição sequencial $\mathcal{S}$. Uma vez que toda condição sequencial é uma condição de otimalidade, então uma SCQ é de fato sempre uma condição de qualificação. A recíproca nem sempre é verdadeira, por exemplo, a condição de Abadie é uma condição de qualificação mas não é uma SCQ para a condição sequencial AKKT. Nesta seção, apresentamos as condições de qualificação mínimas associadas às condições sequenciais.

No Teorema a seguir, mostramos que CPG é uma condição de qualificação estrita para a condição sequencial AKKT, ou seja, sob CPG, todo ponto limite de uma sequência AKKT satisfaz as condições KKT. Em outras palavras, AKKT implica “KKT ou não-CPG”.

Teorema 2.80. *Seja $x \in \Omega$ um ponto AKKT. Se x satisfaz CPG, então x é um ponto KKT.*

Demonstração. Pela definição de AKKT, temos que existem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}_+^p$ com $\mu_j^k = 0$, $j \notin A(x)$ e $x^k \rightarrow x$ tais que

$$\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x)} \mu_j^k \nabla g_j(x^k) \rightarrow 0.$$

Como vale a condição CPG, então existe uma base $(\mathcal{I}', \mathcal{J}')$ para $\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}(x))$, onde $\mathcal{I} = \{1, \dots, m\}$ e $\mathcal{J} = A(x)$, tal que

$$\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}(y)) \subset \text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}(y))$$

para todo y em uma vizinhança de x . Logo, para k suficientemente grande, existem $\tilde{\lambda}_i^k, i \in \mathcal{I}'$ e $\tilde{\mu}_j^k \geq 0, j \in \mathcal{J}'$, tais que

$$\nabla f(x^k) + \sum_{i \in \mathcal{I}'} \tilde{\lambda}_i^k \nabla h_i(x^k) + \sum_{j \in \mathcal{J}'} \tilde{\mu}_j^k \nabla g_j(x^k) \rightarrow 0. \quad (2.47)$$

Defina $M_k := \max\{|\tilde{\lambda}_i^k|, \tilde{\mu}_j^k : i \in \mathcal{I}', j \in \mathcal{J}'\}$. Se $\{M_k\}$ for limitada, podemos supor que $\tilde{\lambda}_i^k \rightarrow \tilde{\lambda}_i$ e $\tilde{\mu}_j^k \rightarrow \tilde{\mu}_j$ em uma subsequência. Logo, passando ao limite em (2.47), obtemos

$$\nabla f(x) + \sum_{i \in \mathcal{I}'} \tilde{\lambda}_i \nabla h_i(x) + \sum_{j \in \mathcal{J}'} \tilde{\mu}_j \nabla g_j(x) = 0.$$

Como

$$\sum_{i \in \mathcal{I}'} \tilde{\lambda}_i \nabla h_i(x) + \sum_{j \in \mathcal{J}'} \tilde{\mu}_j \nabla g_j(x) \in \text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}(x)) = L_\Omega(x)^\circ,$$

concluimos que x satisfaz KKT. Por outro lado, se $M_k \rightarrow \infty$, a sequência $\{(\frac{\tilde{\lambda}_i^k}{M_k}, \frac{\tilde{\mu}_j^k}{M_k})\}$ é limitada, pois $\|(\frac{\tilde{\lambda}_i^k}{M_k}, \frac{\tilde{\mu}_j^k}{M_k})\|_\infty = 1$. Logo, existe uma subsequência convergente, digamos

$$\left(\frac{\tilde{\lambda}_i^k}{M_k}, \frac{\tilde{\mu}_j^k}{M_k}\right) \xrightarrow{\mathbb{N}'} (\hat{\lambda}_i, \hat{\mu}_j).$$

Dividindo a expressão (2.47) por M_k e tomando o limite, obtemos

$$\sum_{i \in \mathcal{I}'} \hat{\lambda}_i \nabla h_i(x) + \sum_{j \in \mathcal{J}'} \hat{\mu}_j \nabla g_j(x) = 0,$$

com $\|(\hat{\lambda}_i, \hat{\mu}_j)\|_\infty = 1$, contradizendo o fato de $(\mathcal{I}', \mathcal{J}')$ ser PLI. $\square$

2.3.1 Propriedade de continuidade do cone

Acreditava-se que CPG era a condição de qualificação estrita mais fraca possível para a condição AKKT. Entretanto, em [2] os autores introduzem uma condição conhecida como Propriedade de Continuidade do Cone (CCP), onde prova-se que CCP é estritamente mais fraca que CPG. Além disso, os autores provam que CCP é a condição de qualificação mínima associada a AKKT, ou seja, CCP é a mínima condição tal que AKKT implica as condições KKT. Em outras palavras, a condição de otimalidade oriunda de CCP, ou seja, “KKT ou não-CCP” fornece a maior medida de força da condição AKKT. A condição CCP afirma a continuidade da multifunção $x \in \mathbb{R}^n \rightrightarrows K(x)$ no ponto $x = x^*$, onde $K(x)$ é o conjunto definido abaixo.

Definição 2.81. Considere um ponto viável $x^* \in \Omega$. Definimos o conjunto $K(x)$ como sendo

$$K(x) := \left\{ v \in \mathbb{R}^n : v = \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x), \mu_j \geq 0, \lambda_i \in \mathbb{R} \right\}. \quad (2.48)$$

Observe que $K(x)$ é um cone convexo fechado que coincide com o polar do cone linearizado em $x = x^*$. Da expressão (2.48), temos que $K(x)$ pode ser considerado como uma perturbação do polar do cone linearizado ao redor de x^* . Note também que as condições KKT podem ser escritas como $-\nabla f(x^*) \in K(x^*)$.

Observação 2.82. Note que, pela definição de $K(x)$ e pela continuidade dos gradientes, a multifunção $x \in \mathbb{R}^n \rightrightarrows K(x)$ é sempre semi-contínua interiormente em x^* . Assim, a semi-continuidade exterior é suficiente para garantir a continuidade de $K(x)$ em x^* .

Definição 2.83 (CCP). Dizemos que a *Propriedade de Continuidade do Cone (CCP)* vale no ponto $x^* \in \Omega$ se a multifunção $x \in \mathbb{R}^n \rightrightarrows K(x)$ é semi-continua exteriormente em x^* .

Observação 2.84. A condição AKKT está naturalmente associada à condição CCP. De fato, pela definição de AKKT, existe $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+^p$ com $\mu_j^k = 0$ para $j \notin A(x^*)$, tais que $x^k \rightarrow x^*$ e

$$\lim_{k \rightarrow \infty} \left(\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) \right) = 0.$$

Logo, se vale AKKT, os gradientes $-\nabla f(x^k)$ se aproximam de $K(x^k)$ quando $k \rightarrow \infty$.

Teorema 2.85. A condição CCP é a propriedade mínima tal que $AKKT + CCP \Rightarrow KKT$, para toda função objetivo f .

Demonstração. Primeiro vamos mostrar que se vale CCP, então a condição AKKT implica em KKT independentemente da função objetivo. Seja f uma função objetivo tal que a condição AKKT vale em x^* . Pela definição de AKKT, existem sequências $x^k \rightarrow x^*$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}_+^p$ com $\mu_j^k = 0$, $j \notin A(x)$ e $\{y^k\} \subset \mathbb{R}^n$ tais que

$$y^k := \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) \rightarrow 0. \quad (2.49)$$

Defina

$$w^k := \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k).$$

Pela definição de $K(x)$ e pela relação (2.49), temos que

$$w^k \in K(x^k) \text{ e } w^k = y^k - \nabla f(x^k). \quad (2.50)$$

Passando ao limite em (2.50) e usando a continuidade do gradiente de f , juntamente com a semi-continuidade exterior de $K(x)$, obtemos

$$\lim_{k \rightarrow \infty} w^k = -\nabla f(x^*) \in \limsup_{x \rightarrow x^*} K(x) \subset K(x^*).$$

Portanto, $-\nabla f(x^*) \in K(x^*)$ e, conseqüentemente, x^* satisfaz KKT.

Agora, vamos mostrar que se AKKT implica KKT, para toda função objetivo f , então vale CCP. Seja $y \in \limsup_{x \rightarrow x^*} K(x)$. Pela definição de limite exterior, existe $(x^k, y^k) \rightarrow (x^*, y)$ tal que $y^k \in K(x^k)$. Defina $f(x) := -y^T x$ para todo $x \in \mathbb{R}^n$. Assim,

$$\begin{aligned} \nabla f(x^k) + y^k &= \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) \\ &= -y + y^k \rightarrow 0. \end{aligned}$$

Logo, x^* satisfaz AKKT e, pela hipótese, temos que as condições KKT valem em x^* . Portanto, $-\nabla f(x^*) = y \in K(x^*)$, completando a prova. $\square$

Como AKKT é uma condição de otimalidade, este resultado nos permite concluir que CCP é uma condição de qualificação.

2.3.2 Condição de qualificação mais fraca associada a AGP

Nesta subseção, apresentamos a condição de qualificação mínima associada a AGP, conhecida como AGP-regular [8]. Ou seja, a condição AGP-regular é condição mais fraca para garantir que AGP implica as condições KKT, independentemente da função objetivo.

Definição 2.86. Dizemos que $x^* \in \Omega$ cumpre a condição *AGP-regular* se a multifunção $(x, \varepsilon) \in \mathbb{R}^n \times \mathbb{R}^n \Rightarrow N_{\Omega(x, -\infty)}(x + \varepsilon)$ é semi-contínua exteriormente em $(x^*, 0)$, isto é,

$$\limsup_{(x, \varepsilon) \rightarrow (x^*, 0)} N_{\Omega(x, -\infty)}(x + \varepsilon) \subset N_{\Omega(x^*, -\infty)}(x^*) = K(x^*).$$

Teorema 2.87 ([8], TEOREMA 1). *A condição AGP-regular é a condição mínima para a qual a condição AGP implica em KKT, independentemente da função objetivo.*

Demonstração. Primeiramente, vamos supor que vale a condição AGP-regular e mostrar que AGP implica em KKT, seja qual for a função objetivo. Seja f uma função objetivo tal que AGP(γ) vale em $x^* \in \Omega$ para algum $\gamma \in [-\infty, 0)$. Logo, pela definição de AGP, existe uma sequência $\{x^k\} \subset \mathbb{R}^n$ tal que $x^k \rightarrow x^*$ e $P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k)) - x^k \rightarrow 0$. Defina $y^k := P_{\Omega(x^k, \gamma)}(x^k - \nabla f(x^k))$ e $\varepsilon^k := y^k - x^k$. Logo, $y^k = x^k + \varepsilon^k$ e $\varepsilon^k \rightarrow 0$. Pela Proposição A.40, temos que

$$w^k := x^k - \nabla f(x^k) - y^k \in N_{\Omega(x^k, \gamma)}(y^k).$$

Como $\Omega(x^k, -\infty) \subset \Omega(x^k, \gamma)$, então $N_{\Omega(x^k, \gamma)}(y^k) \subset N_{\Omega(x^k, -\infty)}(y^k)$. Logo, temos que

$$w^k \in N_{\Omega(x^k, -\infty)}(x^k + \varepsilon^k) \quad \text{e} \quad w^k = -\nabla f(x^k) - \varepsilon^k. \quad (2.51)$$

Tomando o limite em (2.51) e usando o fato de que x^* satisfaz AGP-regular, obtemos

$$\lim_{k \rightarrow \infty} w^k = -\nabla f(x^*) \in \limsup_{(x, \varepsilon) \rightarrow (x^*, 0)} N_{\Omega(x, -\infty)}(x + \varepsilon) \subset N_{\Omega(x^*, -\infty)}(x^*) = K(x^*).$$

Assim, $-\nabla f(x^*) \in K(x^*)$ e, portanto, x^* cumpre as condições KKT.

Agora, vamos mostrar que se AGP implica KKT, então vale AGP-regular. Seja $y \in \limsup_{(x, \varepsilon) \rightarrow (x^*, 0)} N_{\Omega(x, -\infty)}(x + \varepsilon)$. Pela definição de limite exterior, existem sequências $\{(x^k, \varepsilon^k)\}$ com $(x^k, \varepsilon^k) \rightarrow (x^*, 0)$ e $y^k \rightarrow y$ tal que $y^k \in N_{\Omega(x^k, -\infty)}(x^k + \varepsilon^k)$. Defina a função objetivo por $f(x) = -y^T x$ para todo $x \in \mathbb{R}^n$. Mostraremos que para essa função, a condição AGP($-\infty$) vale. Defina $w^k := x^k + \varepsilon^k$ e $z^k := P_{\Omega(x^k, -\infty)}(x^k - \nabla f(x^k)) = P_{\Omega(x^k, -\infty)}(x^k + y)$. Como $y^k \in N_{\Omega(x^k, -\infty)}(w^k)$, então da Proposição A.40, segue que $P_{\Omega(x^k, -\infty)}(w^k + y^k) = w^k$. Usando a desigualdade triangular e a não expansividade da projeção ortogonal, obtemos

$$\|z^k - w^k\| = \|P_{\Omega(x^k, -\infty)}(x^k + y) - P_{\Omega(x^k, -\infty)}(w^k + y^k)\| \leq \|y - y^k\| + \|w^k - x^k\|. \quad (2.52)$$

Como $y^k \rightarrow y$ e $w^k - x^k = \varepsilon^k \rightarrow 0$, tomando o limite na desigualdade (2.52), temos que $\|z^k - w^k\| \rightarrow 0$. Logo,

$$\|P_{\Omega(x^k, -\infty)}(x^k - \nabla f(x^k)) - x^k\| = \|z^k - x^k\| \leq \|z^k - w^k\| + \|w^k - x^k\| \rightarrow 0.$$

Portanto, x^* cumpre a condição AGP e, pela hipótese, temos que x^* satisfaz as condições KKT, ou seja, $-\nabla f(x^*) = y \in L_{\Omega}(x^*)^\circ = N_{\Omega(x, -\infty)}(x^*)$. Assim,

$$\limsup_{(x, \varepsilon) \rightarrow (x^*, 0)} N_{\Omega(x, -\infty)}(x + \varepsilon) \subset N_{\Omega(x^*, -\infty)}(x^*),$$

concluindo que vale AGP-regular. □

2.3.3 Condições de qualificação associadas a CAKKT e SAKKT

Outras condições de qualificação mínimas associadas às condições sequenciais foram estudadas em [8, 58]. Nesta subseção, apresentamos as condições mínimas associadas às condições sequenciais CAKKT e SAKKT, baseando-se nos resultados e demonstrações feitas em [8].

Condição mínima associada a CAKKT

Definição 2.88. Dizemos que a condição *CAKKT-regular* vale no ponto $x^* \in \Omega$ se a multifunção $(x, r) \in \mathbb{R}^n \times \mathbb{R}_+ \Rightarrow K_C(x, r)$ é semi-contínua exteriormente em $(x^*, 0)$, ou seja,

$$\limsup_{(x,r) \rightarrow (x^*,0)} K_C(x, r) \subset K_C(x^*, 0) = K(x^*),$$

onde $K_C(x, r)$ é definido como

$$K_C(x, r) := \left\{ \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x) : (\lambda, \mu) \in M(x, r) \right\}$$

e $M(x, r)$ é conjunto de multiplicadores $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ com $\mu_j = 0$ para $j \notin A(x^*)$ tais que

$$\sum_{i=1}^m |\lambda_i h_i(x)| + \sum_{j \in A(x^*)} |\mu_j g_j(x)| \leq r.$$

Teorema 2.89. A condição *CAKKT-regular* é a propriedade mínima para a qual a condição *CAKKT* implica em *KKT*, independentemente da função objetivo.

Demonstração. Primeiro, mostremos que se vale *CAKKT-regular*, então *CAKKT* implica *KKT*, independentemente da função objetivo. Seja f uma função objetivo tal que *CAKKT* vale em $x^* \in \Omega$. Pelo Lema 2.69, existem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{(\lambda^k, \mu^k)\} \subset \mathbb{R}^m \times \mathbb{R}_+^p$ com $\mu_j^k = 0$ para $j \notin A(x^*)$ tais que $x^k \rightarrow x^*$,

$$\zeta^k := \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) \rightarrow 0,$$

e

$$r^k := \sum_{i=1}^m |\lambda_i^k h_i(x^k)| + \sum_{j \in A(x^*)} |\mu_j^k g_j(x^k)| \rightarrow 0.$$

Denote $w^k := \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k)$. Claramente a sequência $\{w^k\}$ satisfaz

$$w^k \in K_C(x, r) \quad \text{e} \quad w^k = \zeta^k - \nabla f(x^k).$$

Como $\zeta^k \rightarrow 0$ e $\nabla f(x^k) \rightarrow \nabla f(x^*)$, segue que $w^k \rightarrow -\nabla f(x^*)$. Logo, pela definição de limite exterior e usando a hipótese, temos

$$\lim_{k \rightarrow \infty} w^k = -\nabla f(x^*) \in \limsup_{(x,r) \rightarrow (x^*,0)} K_C(x, r) \subset K_C(x^*, 0) = K(x^*),$$

o que significa que vale *KKT*. Agora, mostremos que se *CAKKT* implica *KKT* para qualquer função objetivo f , então *CAKKT-regular* vale. Seja $w \in \limsup_{(x,r) \rightarrow (x^*,0)} K_C(x, r)$.

Pela definição de limite exterior, existem sequências $x^k \rightarrow x^*$, $w^k \rightarrow w$ e $r^k \rightarrow 0$ tal que $w^k \in K_C(x^k, r^k)$. Defina a função $f(x) := -w^T x$ para todo $x \in \mathbb{R}^n$. Mostremos que CAKKT vale em x^* para a função f . Como $w^k \in K_C(x^k, r^k)$, então existem sequências de multiplicadores $\{\lambda^k\} \subset \mathbb{R}^m$ e $\{\mu^k\} \subset \mathbb{R}_+^p$ com $\mu_j^k = 0$, $j \notin A(x^*)$ tais que

$$w^k = \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) \quad \text{e} \quad \sum_{i=1}^m |\lambda_i^k h_i(x^k)| + \sum_{j \in A(x^*)} |\mu_j^k g_j(x^k)| \leq r^k.$$

Como $r^k \rightarrow 0$, $w^k \rightarrow w$ e $\zeta^k := \nabla f(x^k) + w^k = -w + w^k \rightarrow 0$, segue que CAKKT vale. Logo, pela hipótese, segue que $-\nabla f(x^*) = w \in K(x^*) = K_C(x^*, 0)$, terminando a prova. $\square$

Corolário 2.90. *A condição CAKKT-regular é uma condição de qualificação.*

Condição mínima associada a SAKKT

Definição 2.91. Dizemos que a condição SAKKT-regular vale no ponto $x^* \in \Omega$ se a multifunção $x \in \mathbb{R}^n \Rightarrow N_{\Omega(x,0)}(x)$ é semi-contínua exteriormente em x^* , ou seja,

$$\limsup_{x \rightarrow x^*} N_{\Omega(x,0)}(x) \subset N_{\Omega(x^*,0)}(x^*) = K(x^*).$$

Teorema 2.92. *A condição SAKKT-regular é a propriedade mínima para o qual a condição SAKKT implica em KKT, independentemente da função objetivo. Ou em outras palavras, a condição mínima para garantir que SAKKT implique em KKT é a semi-continuidade exterior de $N_{\Omega(x,0)}(x)$ em x^* .*

Demonstração. Primeiro, mostraremos que se vale SAKKT-regular, então a condição SAKKT implica em KKT, seja qual for a função objetivo. Seja f uma função objetivo tal que SAKKT vale em $x^* \in \Omega$. Pelo Teorema 2.57, temos que SAKKT é equivalente à condição AGP(0). Logo, existe uma sequência $x^k \rightarrow x^*$ tal que $P_{\Omega(x^k,0)}(x^k - \nabla f(x^k)) - x^k \rightarrow 0$. Defina $y^k := P_{\Omega(x^k,0)}(x^k - \nabla f(x^k))$ e $\varepsilon^k := y^k - x^k$. Pelas Proposições A.40 e 2.48, temos que

$$w^k := x^k - \nabla f(x^k) - y^k \in N_{\Omega(x^k,0)}(y^k) \subset N_{\Omega(x^k,0)}(x^k). \quad (2.53)$$

Portanto, usando a continuidade do gradiente de f , temos que

$$w^k \in N_{\Omega(x^k,0)}(x^k) \quad \text{e} \quad w^k = x^k - \nabla f(x^k) - y^k \rightarrow -\nabla f(x^*).$$

Assim, pela definição de limite exterior e usando a semi-continuidade exterior de $N_{\Omega(x,0)}(x)$ em x^* , obtemos

$$-\nabla f(x^*) \in \limsup_{x \rightarrow x^*} N_{\Omega(x,0)}(x) \subset N_{\Omega(x^*,0)}(x^*) = K(x^*).$$

Portanto x^* cumpre as condições KKT.

Agora, vamos mostrar que se SAKKT implica em KKT, então a condição SAKKT-regular vale, para toda função objetivo f . Seja $y \in \limsup_{x \rightarrow x^*} N_{\Omega(x,0)}(x)$. Logo existem sequências $x^k \rightarrow x^*$, $y^k \rightarrow y$ tal que $y^k \in N_{\Omega(x^k,0)}(x^k)$. Defina a função objetivo por $f(x) := -y^T x$ para todo $x \in \mathbb{R}^n$. Mostraremos que, para essa função, a condição AGP(0)

vale em x^* . Denote $z^k := P_{\Omega(x^k,0)}(x^k - \nabla f(x^k)) = P_{\Omega(x^k,0)}(x^k + y)$. Como $y^k \in N_{\Omega(x^k,0)}(x^k)$, então $P_{\Omega(x^k,0)}(x^k + y^k) = x^k$. Logo, pela não expansividade da projeção ortogonal, obtemos

$$\|z^k - x^k\| = \|P_{\Omega(x^k,0)}(x^k + y) - P_{\Omega(x^k,0)}(x^k + y^k)\| \leq \|y - y^k\| \rightarrow 0.$$

Portanto, $P_{\Omega(x^k,0)}(x^k - \nabla f(x^k)) - x^k \rightarrow 0$ e vale AGP(0) em x^* . Como consequência, a condição SAKKT vale em x^* . Assim, pela hipótese, as condições KKT valem em x^* . Logo, $-\nabla f(x^*) = y \in K(x^*) = N_{\Omega(x,0)}(x^*)$, completando a prova. $\square$

2.3.4 Condição de qualificação mais fraca associada a PAKKT

Aqui apresentamos a condição mínima associada à condição PAKKT, chamada de PAKKT-regular [5], mais fraca que CCP. Para isso, considere $x^* \in \Omega$ e $\alpha, \beta \geq 0$. Defina o seguinte o conjunto

$$K_+(x, \alpha, \beta) := \left\{ \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x) \left| \begin{array}{l} \lambda_i h_i(x) \geq \alpha \text{ se } |\lambda_i| > \beta \|(1, \lambda, \mu)\|_\infty, \\ \mu_j g_j(x) \geq \alpha \text{ se } \mu_j > \beta \|(1, \lambda, \mu)\|_\infty, \\ \lambda \in \mathbb{R}^m, \mu_j \geq 0, j \in A(x^*) \end{array} \right. \right\}.$$

Observe que as condições KKT podem ser escritas como $-\nabla f(x^*) \in K_+(x^*, 0, 0)$.

Definição 2.93. Dizemos que $x^* \in \Omega$ satisfaz a condição *PAKKT-regular* se a multifunção $(x, \alpha, \beta) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}_+ \Rightarrow K_+(x, \alpha, \beta)$ é semi-contínua exteriormente em $(x^*, 0, 0)$, ou seja, quando

$$\limsup_{x \rightarrow x^*, \alpha \downarrow 0, \beta \downarrow 0} K_+(x, \alpha, \beta) \subset K_+(x^*, 0, 0).$$

O próximo Teorema mostra que PAKKT-regular é a SCQ mais fraca para a condição PAKKT.

Teorema 2.94. A condição *PAKKT-regular* é a propriedade mínima para a qual a condição *PAKKT* implica as condições *KKT*, independentemente da função objetivo.

Demonstração. Seja $x^* \in \Omega$. Primeiro vamos mostrar que se x^* é um ponto PAKKT que cumpre PAKKT-regular, então x^* satisfaz as condições KKT. Pela definição de PAKKT existem sequências $\{x^k\} \subset \mathbb{R}^m$, $\{\lambda^k\} \subset \mathbb{R}^m$ e $\{\mu^k\} \subset \mathbb{R}_+^p$ com $x^k \rightarrow x^*$ tais que

$$\nabla f(x^k) + w^k \rightarrow 0, \quad (2.54)$$

$$\min\{-g_j(x^k), \mu_j^k\} \rightarrow 0, \quad \forall j = 1, \dots, p \quad (2.55)$$

$$\lambda_i^k h_i(x^k) > 0 \text{ se } \lim_{k \rightarrow \infty} \frac{|\lambda_i^k|}{\delta^k} > 0, \quad \mu_j^k g_j(x^k) > 0 \text{ se } \lim_{k \rightarrow \infty} \frac{\mu_j^k}{\delta^k} > 0, \quad (2.56)$$

onde

$$w^k = \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \quad \text{e} \quad \delta^k = \|(1, \lambda^k, \mu^k)\|_\infty.$$

Por (2.55), podemos supor que $\mu_j^k = 0$ para $j \notin A(x^*)$. Defina $I_+ = \{i \in \{1, \dots, m\} : \lim |\lambda_i^k|/\delta^k > 0\}$ e $J_+ = \{j \in A(x^*) : \lim \mu_j^k/\delta^k > 0\}$. Para cada k , definimos as sequências

$$\alpha^k := \min \left\{ \frac{1}{k}, \min_{i \in I_+} \{\lambda_i^k h_i(x^k)\}, \min_{j \in J_+} \{\mu_j^k g_j(x^k)\} \right\} \quad \text{e} \quad \beta^k := \max \left\{ \frac{1}{k}, \max_{i \notin I_+} \frac{|\lambda_i^k|}{\delta^k}, \max_{j \notin J_+} \frac{\mu_j^k}{\delta^k} \right\} + \frac{1}{k}.$$

Assim, temos que $\alpha^k \downarrow 0$, $\beta^k \downarrow 0$ e $w^k \in K_+(x^k, \alpha^k, \beta^k)$. Usando a hipótese de que x^* cumpre PAKKT-regular, segue que

$$\lim_{k \rightarrow \infty} w^k = -\nabla f(x^*) \in \limsup_{x \rightarrow x^*, \alpha \downarrow 0, \beta \downarrow 0} K_+(x, \alpha, \beta) \subset K_+(x^*, 0, 0).$$

Portanto, $-\nabla f(x^*) \in K_+(x^*, 0, 0) = L_\Omega(x^*)^\circ$, o que significa que x^* cumpre KKT.

Agora vamos mostrar que se PAKKT implica KKT, então a condição PAKKT-regular vale para toda função objetivo f . Seja $w \in \limsup_{x \rightarrow x^*, \alpha \downarrow 0, \beta \downarrow 0} K_+(x, \alpha, \beta)$. Logo existem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{w^k\} \subset \mathbb{R}^n$, $\{\alpha^k\} \subset \mathbb{R}$ e $\{\beta^k\} \subset \mathbb{R}$ tais que $x^k \rightarrow x^*$, $w^k \rightarrow w$, $\alpha^k \downarrow 0$, $\beta^k \downarrow 0$ e $w^k \in K_+(x^k, \alpha^k, \beta^k)$. Assim, para cada k , existem $\lambda^k \in \mathbb{R}^m$ e $\mu^k \in \mathbb{R}_+^{|A(x^*)|}$ tais que

$$w^k = \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k).$$

Defina $f(x) = -w^T x$ para todo $x \in \mathbb{R}^n$. Se $\lim_{k \rightarrow \infty} \frac{|\lambda_i^k|}{\delta^k} > 0$, então $|\lambda_i^k| > \beta^k \delta^k$ para todo k suficientemente grande (analogamente para μ). Além disso, temos que $\nabla f(x^k) + w^k = -w + w^k \rightarrow 0$ e, portanto, x^* satisfaz PAKKT. Logo, usando a hipótese, segue que x^* é um ponto KKT, isto é, $-\nabla f(x^*) = w \in K_+(x^*, 0, 0)$, concluindo a prova. $\square$

Devido ao fato de PAKKT ser uma condição de otimalidade, segue que PAKKT-regular é uma condição de qualificação.

Observação 2.95. Note que $K_+(x, \alpha, \beta) \subset K_+(x, 0, \infty)$ para todo $x \in \mathbb{R}^n$ e $\alpha, \beta > 0$. Além disso, $K_+(x, 0, \infty) = K_+(x, 0, 0) = K(x)$ para todo $x \in \Omega$.

Portanto $x^* \in \Omega$ satisfaz a condição CCP quando

$$\limsup_{x \rightarrow x^*} K_+(x, 0, \infty) \subset K_+(x^*, 0, \infty).$$

Teorema 2.96. A condição CCP implica a condição PAKKT-regular.

Demonstração. Seja $x^* \in \Omega$ um ponto que cumpre CCP. Logo, temos

$$\limsup_{x \rightarrow x^*, \alpha \downarrow 0, \beta \downarrow 0} K_+(x, \alpha, \beta) \subset \limsup_{x \rightarrow x^*} K_+(x, 0, \infty) \subset K_+(x^*, 0, \infty) = K_+(x^*, 0, 0).$$

Portanto x^* satisfaz PAKKT-regular. $\square$

2.3.5 Condição de qualificação mais fraca associada a PCAKKT

Procedemos a definir a condição mínima associada à condição sequencial PCAKKT, chamada PCAKKT-regular [7]. Para x^* fixado, sejam $x \in \mathbb{R}^n$ e $\alpha, \beta, \sigma \geq 0$. Definimos o seguinte conjunto:

$$K_{PC}(x, \alpha, \beta, \sigma) := \left\{ \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j=1}^p \mu_j \nabla g_j(x) : (\lambda, \mu) \in M(x, \alpha, \beta, \sigma) \right\},$$

onde $M(x, \alpha, \beta, \sigma)$ é conjunto dos multiplicadores $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ com $\mu_j = 0$ para $j \notin A(x^*)$ tais que

$$\lambda_i h_i(x) \geq \alpha \text{ se } |\lambda_i| > \beta \|(1, \lambda, \mu)\|_\infty, \quad (2.57)$$

$$\mu_j g_j(x) \geq \alpha \text{ se } \mu_j > \beta \|(1, \lambda, \mu)\|_\infty, \quad j \in A(x^*), \quad (2.58)$$

$$\sum_{i=1}^m |\lambda_i h_i(x)| + \sum_{j \in A(x^*)} |\mu_j g_j(x)| \leq \sigma. \quad (2.59)$$

Note que $K_{PC}(x^*, 0, 0, 0) = K(x^*)$. Para definir a condição PCAKKT-regular, considere:

$$\limsup_{x \rightarrow x^*, \alpha \downarrow 0, \beta \downarrow 0, \sigma \downarrow 0} K_{PC}(x, \alpha, \beta, \sigma) = \left\{ w \in \mathbb{R}^n : \begin{array}{l} \exists (x^k, w^k) \rightarrow (x^*, w), \alpha^k \downarrow 0, \beta^k \downarrow 0, \sigma^k \downarrow 0 \\ \text{tal que } w^k \in K_{PC}(x^k, \alpha^k, \beta^k, \sigma^k) \end{array} \right\}.$$

Definição 2.97. Dizemos que $x^* \in \Omega$ satisfaz a condição PCAKKT-regular quando

$$\limsup_{x \rightarrow x^*, \alpha \downarrow 0, \beta \downarrow 0, \sigma \downarrow 0} K_{PC}(x, \alpha, \beta, \sigma) \subset K_{PC}(x^*, 0, 0, 0).$$

Teorema 2.98. *Todo ponto PCAKKT que satisfaz PCAKKT-regular é um ponto KKT. Reciprocamente, se PCAKKT implica as condições KKT, para toda função objetivo f , então PCAKKT-regular vale.*

Demonstração. Seja $x^* \in \Omega$ e suponha que x^* satisfaz PCAKKT. Da definição de PCAKKT e do Lema 2.69 segue que existem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{(\lambda^k, \mu^k)\} \subset \mathbb{R}^m \times \mathbb{R}_+^p$ com $\mu_j^k = 0$, $j \notin A(x^*)$ e $x^k \rightarrow x^*$ tais que valem as condições (2.54), (2.56) e

$$\lim_{k \rightarrow \infty} \left(\sum_{i=1}^m |\lambda_i^k h_i(x^k)| + \sum_{j \in A(x^*)} |\mu_j^k g_j(x^k)| \right) = 0, \quad (2.60)$$

onde

$$w^k = \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) \quad \text{e} \quad \delta^k = \|(1, \lambda^k, \mu^k)\|_\infty.$$

Defina os conjuntos de índices I_+ e J_+ como na demonstração do Teorema 2.94 e as seguintes sequências:

$$\alpha^k := \min \left\{ \frac{1}{k}, \min_{i \in I_+} \{\lambda_i^k h_i(x^k)\}, \min_{j \in J_+} \{\mu_j^k g_j(x^k)\} \right\}, \quad (2.61)$$

$$\beta^k := \max \left\{ \frac{1}{k}, \max_{i \notin I_+} \frac{|\lambda_i^k|}{\delta^k}, \max_{j \notin J_+} \frac{\mu_j^k}{\delta^k} \right\} \quad \text{e} \quad \sigma^k := c^k, \quad (2.62)$$

onde

$$c^k := \sum_{i=1}^m |\lambda_i^k h_i(x^k)| + \sum_{j=1}^p |\mu_j^k g_j(x^k)|, \quad \forall k \in \mathbb{N}.$$

Pelas condições (2.56) e (2.60) segue que $\alpha^k \downarrow 0$ e $\sigma^k \downarrow 0$. Além disso, pela definição de I_+ e J_+ , temos que $\beta^k \downarrow 0$. Dado $k \in \mathbb{N}$, considere um índice i tal que $|\lambda_i^k| > \beta^k \delta^k$. Como $\delta^k > 0$, segue de (2.62) que $|\lambda_i^k|/\delta^k > \beta^k \geq \max_{t \notin I_+} |\lambda_t^k|/\delta^k$ e portanto $i \in I_+$. Por

(2.61), temos que $\lambda_i^k h_i(x^k) \geq \alpha^k$, o que significa que essas sequências cumprem a condição (2.57). Analogamente, podemos mostrar que tais sequências também satisfazem (2.58). Claramente a condição (2.59) é satisfeita. Logo, $w^k \in K_{PC}(x^k, \alpha^k, \beta^k, \sigma^k)$. Como x^* satisfaz PCAKKT-regular, segue que

$$-\nabla f(x^*) = \lim_{k \rightarrow \infty} w^k \in \limsup_{x \rightarrow x^*, \alpha \downarrow 0, \beta \downarrow 0, \sigma \downarrow 0} K_{PC}(x, \alpha, \beta, \sigma) \subset K_{PC}(x^*, 0, 0, 0) = K(x^*).$$

Portanto as condições de KKT valem em x^* . Agora vamos provar que se PCAKKT implica as condições KKT para toda função objetivo f , então PCAKKT-regular vale. Seja x^* um ponto PCAKKT. Para provar que x^* satisfaz PCAKKT-regular considere $w \in \limsup_{x \rightarrow x^*, \alpha \downarrow 0, \beta \downarrow 0, \sigma \downarrow 0} K_{PC}(x, \alpha, \beta, \sigma)$. Pela definição de limite exterior, existem sequências $x^k \rightarrow x^*$, $w^k \rightarrow w$, $\alpha^k \downarrow 0$, $\beta^k \downarrow 0$ e $\sigma^k \downarrow 0$ tal que $w^k \in K_{PC}(x^k, \alpha^k, \beta^k, \sigma^k)$. Logo existem sequências $\{\lambda^k\} \subset \mathbb{R}^m$ e $\{\mu^k\} \subset \mathbb{R}_+^p$ com $\mu_j^k = 0$ para $j \notin A(x^*)$ tais que

$$w^k = \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k)$$

e valem as condições (2.57), (2.58) e (2.59). Defina a função $f(x) = -w^T x$. Então x^* satisfaz PCAKKT para a função f . De fato, claramente a condição (2.54) se cumpre. Além disso, segue de (2.59) que (2.60) é satisfeita. Se $\lim_{k \rightarrow \infty} |\lambda_i^k|/\delta^k > 0$, temos que $|\lambda_i^k| > \beta^k \delta^k$ para todo k suficientemente grande. Assim, por (2.57) segue que $\lambda_i^k h_i(x^k) \geq \alpha^k > 0$. Pelo mesmo raciocínio, temos $\mu_j g_j(x^k) > 0$ se $\lim_{k \rightarrow \infty} \mu_j^k/\delta^k > 0$ o que significa que (2.56) vale. Portanto x^* é um ponto PCAKKT. Assim, pela hipótese, segue que x^* satisfaz as condições KKT, isto é, $-\nabla f(x^*) = w \in K(x^*) = K_{PC}(x^*, 0, 0, 0)$, provando que x^* satisfaz PCAKKT-regular. $\square$

No artigo [7], os autores provam que PCAKKT-regular é mais forte que a condição de qualificação de Abadie. Porém, o contra-exemplo apresentado a seguir mostra que essas condições não são equivalentes.

Contra-Exemplo 2.99 (Abadie não implica PCAKKT-regular). Considere as restrições de desigualdade $g_1(x) = x_2 - x_1^2 \leq 0$, $g_2(x) = -x_2 - x_1^2 \leq 0$, $g_3(x) = x_2 - x_1^5 \leq 0$, $g_4(x) = -x_2 - x_1^5 \leq 0$ e $g_5(x) = -x_1 \leq 0$. Todas as restrições são ativas no ponto $x^* = (0, 0)$. Um cálculo simples mostra que $T_\Omega(x^*) = L_\Omega(x^*) = \{(x_1, 0) : x_1 \geq 0\}$ e vale Abadie em x^* . Entretanto, x^* não satisfaz PCAKKT-regular. Com efeito, seja $w = (1, 0)$ e considere as sequências definidas por $x^k = (-1/k, 0)$, $\mu^k = (k/4, k/4, k^3, k^3, 0)$, $\alpha^k = 1/k^2$, $\beta^k = 1/k$, $\sigma^k = \sum_{j \in A(x^*)} |\mu_j^k g_j(x^k)|$ e $w^k = \sum_{j=1}^5 \mu_j^k \nabla g_j(x^k)$. Logo, temos $\alpha^k \downarrow 0$, $\beta^k \downarrow 0$, $\sigma^k \downarrow 0$ e

$$w^k = \sum_{j=1}^5 \mu_j^k \nabla g_j(x^k) = \begin{pmatrix} 1/2 \\ k/4 \end{pmatrix} + \begin{pmatrix} 1/2 \\ -k/4 \end{pmatrix} + \begin{pmatrix} -5/k \\ k^3 \end{pmatrix} + \begin{pmatrix} -5/k \\ -k^3 \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} = w.$$

Ainda, $w^k \in K_{PC}(x^k, \alpha^k, \beta^k, \sigma^k)$ para todo k . Logo, $w \in \limsup_{x \rightarrow x^*, \alpha \downarrow 0, \beta \downarrow 0, \sigma \downarrow 0} K_{PC}(x, \alpha, \beta, \sigma)$.

Mas um cálculo simples mostra que $w \notin K(x^*) = K_{PC}(x^*, 0, 0, 0)$. Portanto x^* não satisfaz PCAKKT-regular.

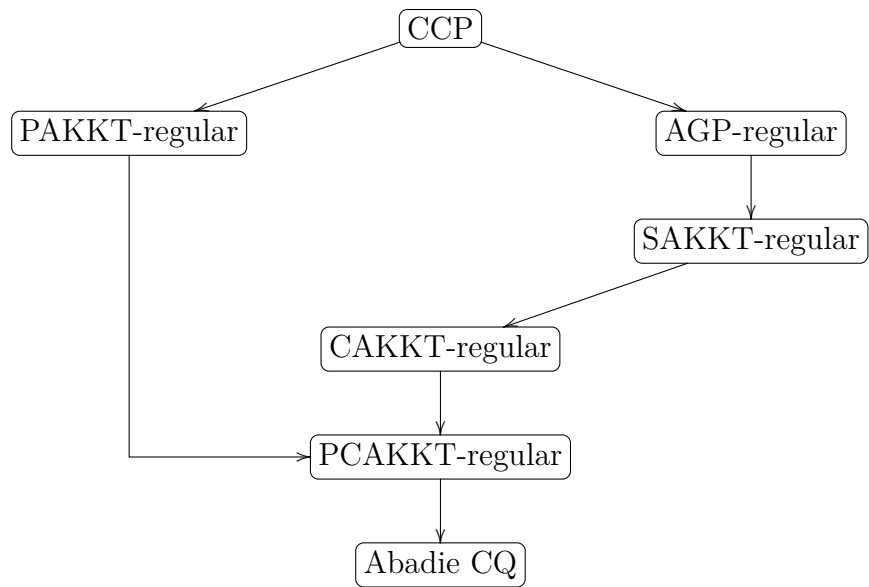


Figura 2.6: Relações entre as condições de qualificação mínimas

3 Condições de Otimalidade de Segunda Ordem

Para a realização deste capítulo nos baseamos nas definições e demonstrações apresentadas em [4, 6, 10, 19, 38, 58].

As condições de otimalidade de segunda ordem desempenham um papel importante em otimização. Tais condições levam em consideração a curvatura da função Lagrangiana sobre o conjunto de direções viáveis. As informações de primeira ordem, tais como os gradientes ou direções do cone tangente que estudamos no capítulo anterior, não são suficientes para caracterizar as soluções. Para entender melhor essa discussão, vejamos o exemplo a seguir.

Exemplo 3.1. Considere o problema de otimização:

$$\begin{aligned} \text{minimizar} \quad & f(x) = -(x_1 - 2)^2 - (x_2 - 3)^2, \\ \text{sujeito a} \quad & g_1(x) = -3x_1 - 2x_2 + 6 \leq 0, \\ & g_2(x) = -x_1 + x_2 - 3 \leq 0, \\ & g_3(x) = x_1 - 2 \leq 0. \end{aligned}$$

No ponto $x^* = (1, 4)$, temos $A(x^*) = \{2\}$. Logo,

$$\nabla f(x^*) + \mu_2 \nabla g_2(x^*) = 0 \Rightarrow \mu_2 = 2.$$

Portanto, as condições de KKT valem em x^* . Entretanto, x^* não é uma solução para este problema. Conforme a Figura 3.1, podemos ver que $(2, 0)$ é a solução global deste problema, enquanto que $(0, 3)$ e $(2, 5)$ são soluções locais. Assim, informações de primeira ordem não são suficientes para a otimalidade.

Usando informações de segunda ordem, podemos pensar em formular condições de otimalidade de segunda ordem. Essas condições são geralmente muito mais fortes do que as condições de primeira ordem e, portanto, são mais interessantes. Tais condições foram amplamente estudadas em [4, 6, 10, 13, 19, 21, 23, 31, 38] com aplicações importantes em programação matemática, otimização composta, controle ótimo, entre outros.

Definição 3.2. Definimos o *cone crítico fraco* como sendo

$$C^W(x) = \{d \in \mathbb{R}^n : \nabla h_i(x)^T d = 0, i = 1, \dots, m; \nabla g_j(x)^T d = 0, j \in A(x)\}$$

e o *cone crítico forte* por

$$C^S(x) = \left\{ d \in \mathbb{R}^n : \begin{array}{l} \nabla h_i(x)^T d = 0, i = 1, \dots, m; \\ \nabla g_j(x)^T d = 0, j \in A^+(x); \\ \nabla g_j(x)^T d \leq 0, j \in A^0(x) \end{array} \right\},$$

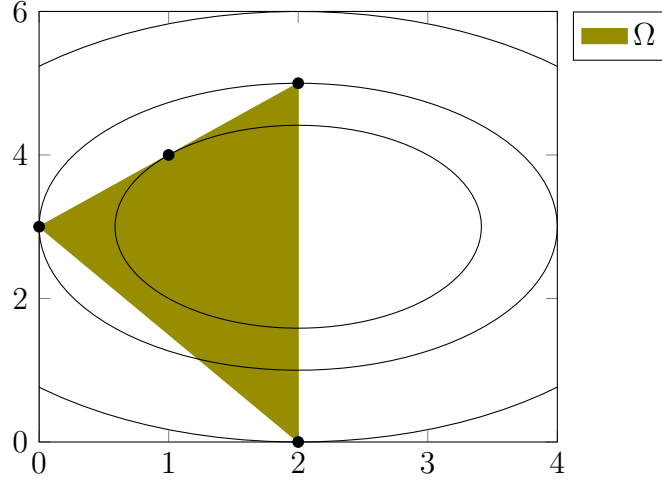


Figura 3.1: A condição KKT não é suficiente para a otimalidade

onde

$$A^+(x) := \{j \in A(x) : \mu_j > 0\}, \quad A^0(x) := \{j \in A(x) : \mu_j = 0\}.$$

Definição 3.3 (WSOnc). Dizemos que o ponto $x \in \Omega$ cumpre a *condição necessária (fraca) de segunda ordem*, *Weak Second-Order Necessary Optimality Condition (WSOnc)*, se existe um multiplicador de Lagrange $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ com $\mu_j = 0$ para $j \notin A(x)$ satisfazendo as condições

$$\nabla f(x) + \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x)} \mu_j \nabla g_j(x) = 0 \quad (3.1)$$

e

$$d^T \left(\nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(x) + \sum_{j \in A(x)} \mu_j \nabla^2 g_j(x) \right) d \geq 0, \quad \forall d \in C^W(x). \quad (3.2)$$

Quando o cone $C^W(x)$ é substituído por $C^S(x)$, dizemos que a *condição necessária (forte) de segunda ordem (SSOnc)* é satisfeita no ponto x . Observe que $C^W(x) \subset C^S(x)$ e que $C^W(x)$ não depende da função objetivo. Note também que ambas as condições dependem apenas da existência de um multiplicador de Lagrange adequado, enquanto outras condições de segunda ordem consideram todo o conjunto de Multiplicadores de Lagrange.

Observação 3.4. Observe que se $x \in \Omega$ cumpre KKT, então o cone crítico forte pode ser descrito da seguinte maneira:

$$C^S(x) = \{d \in \mathbb{R}^n : \nabla h_i(x)^T d = 0, i = 1, \dots, m; \nabla g_j(x)^T d \leq 0, j \in A(x); \nabla f(x)^T d \leq 0\}.$$

Observe que o ponto x^* do Exemplo 3.1, não cumpre a condição (3.2). De fato, calculando o cone crítico fraco, temos

$$C^W(x^*) = \{d \in \mathbb{R}^2 : d_1 = d_2\}.$$

Logo, dado $d \in C^W(x^*)$, temos

$$d^T \left(\nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla^2 g_j(x^*) \right) d = \begin{pmatrix} d_1 & d_1 \end{pmatrix} \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} d_1 \\ d_1 \end{pmatrix} = -4d_1^2.$$

No Teorema a seguir, temos uma condição de otimalidade de segunda ordem, conhecida como *Condição de Fritz John de segunda ordem*, que estende a condição de Fritz John (Teorema 2.4) no caso em que as funções possuem segunda derivada contínua.

Teorema 3.5 ([23], PROPOSIÇÃO 5.48). *Seja x^* um minimizador local do problema (1.2). Então, para cada $d \in C^S(x^*)$, existe $0 \neq (\lambda_0, \lambda, \mu) \in \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}_+^p$ com $\lambda_0 \geq 0$ e $\mu_j = 0$ para $j \notin A(x^*)$ tal que*

$$\lambda_0 \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*) = 0, \quad (3.3)$$

$$d^T \left(\lambda_0 \nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla^2 g_j(x^*) \right) d \geq 0. \quad (3.4)$$

O próximo Teorema fornece uma condição suficiente para a otimalidade local.

Teorema 3.6. *Seja $x^* \in \Omega$. Se para cada $d \in C^S(x^*) \setminus \{0\}$ existe um multiplicador de Lagrange $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ associado a x^* tal que $d^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda, \mu) d > 0$, então x^* é um minimizador local de (1.2).*

Demonstração. Suponha que x^* não seja um minimizador local. Logo existe uma sequência $\{x^k\} \subset \Omega$ com $x^k \rightarrow x^*$ tal que $f(x^k) \leq f(x^*) + o(\|x^k - x^*\|^2)$. Assim,

$$\lim_{k \rightarrow \infty} \frac{f(x^k) - f(x^*)}{\|x^k - x^*\|^2} \leq 0. \quad (3.5)$$

Defina $d^k := (x^k - x^*)/\|x^k - x^*\|$. Extrairdo uma subsequência se necessário, podemos supor que $d^k \rightarrow \hat{d}$. Como $x^k, x^* \in \Omega$, usando expansão de Taylor, obtemos

$$\begin{aligned} h_i(x^k) &= h_i(x^*) + \nabla h_i(x^*)^T (x^k - x^*) + o(\|x^k - x^*\|) = 0, \quad i = 1, \dots, m, \\ g_j(x^k) &= g_j(x^*) + \nabla g_j(x^*)^T (x^k - x^*) + o(\|x^k - x^*\|) \leq 0, \quad j \in A(x^*), \end{aligned}$$

onde $o(\|x^k - x^*\|)/\|x^k - x^*\| \rightarrow 0$ para $k \rightarrow \infty$. Dividindo ambas as expressões por $\|x^k - x^*\|$ e tomando o limite, obtemos

$$\nabla h_i(x^*)^T \hat{d} = 0, i = 1, \dots, m \quad \text{e} \quad \nabla g_j(x^*)^T \hat{d} \leq 0, j \in A(x^*).$$

Segue que $\hat{d} \in C^S(x^*)$. Logo, usando a hipótese, existe um multiplicador de Lagrange (λ', μ') associado a x^* tal que

$$(\hat{d})^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda', \mu') \hat{d} > 0.$$

Além disso, temos

$$\begin{aligned} f(x^k) - f(x^*) &\geq \mathcal{L}(x^k, \lambda', \mu') - \mathcal{L}(x^*, \lambda', \mu') \\ &= \nabla \mathcal{L}(x^*, \lambda', \mu')^T (x^k - x^*) + \frac{1}{2} (x^k - x^*)^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda', \mu') (x^k - x^*) + o(\|x^k - x^*\|^2) \end{aligned}$$

Assim,

$$f(x^k) - f(x^*) \geq \frac{1}{2} (x^k - x^*)^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda', \mu') (x^k - x^*) + o(\|x^k - x^*\|^2).$$

Dividindo a desigualdade acima por $\|x^k - x^*\|^2$ e tomando o limite, obtemos

$$\lim_{k \rightarrow \infty} \frac{f(x^k) - f(x^*)}{\|x^k - x^*\|^2} \geq \frac{1}{2} (\hat{d})^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda', \mu') \hat{d} > 0,$$

contradizendo (3.5). □

3.1 Condições de qualificação de segunda ordem

No capítulo anterior, vimos que, para garantir a validade das condições KKT em um minimizador local é necessário que alguma condição de qualificação seja satisfeita. No caso de segunda ordem, para garantir a validade de SSONC ou WSONC em um minimizador local, é necessário que uma condição de qualificação de segunda ordem (CQ2) seja satisfeita. Na prática, a maioria das condições de otimalidade de segunda ordem usuais são da forma: “Se um minimizador local satisfaz alguma condição de qualificação, então a condição WSONC é satisfeita”. Ou seja, essas condições de otimalidade são da forma “WSONC ou não-CQ2”. Nesta seção, apresentamos condições mais fracas que LICQ e que garantem a validade das condições de otimalidade fraca e forte de segunda ordem.

Teorema 3.7. *Seja $x^* \in \Omega$ um minimizador local que cumpre MFCQ. Então, para cada $d \in C^S(x^*)$ existe um multiplicador de Lagrange $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ associado a x^* tal que*

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

Demonstração. Basta fazer uma adaptação do Teorema 3.5 observando que sob MFCQ os multiplicadores de Fritz John ($\lambda_0 > 0$) são multiplicadores de Lagrange. $\square$

Entretanto, esta condição não é muito interessante, pois ela depende do conjunto de multiplicadores de Lagrange, $\Lambda(x^*)$. O exemplo a seguir mostra que existem problemas que cumprem MFCQ na solução sem que a condição necessária de otimalidade de segunda ordem seja satisfeita.

Exemplo 3.8. Considere o problema de otimização

$$\begin{aligned} &\text{minimizar} && x_3 \\ &\text{sujeito a} && g_1(x) = 2\sqrt{3}x_1x_2 - 2x_2^2 - x_3 \leq 0, \\ &&& g_2(x) = x_2^2 - 3x_1^2 - x_3 \leq 0, \\ &&& g_3(x) = -2\sqrt{3}x_1x_2 - 2x_2^2 - x_3 \leq 0. \end{aligned}$$

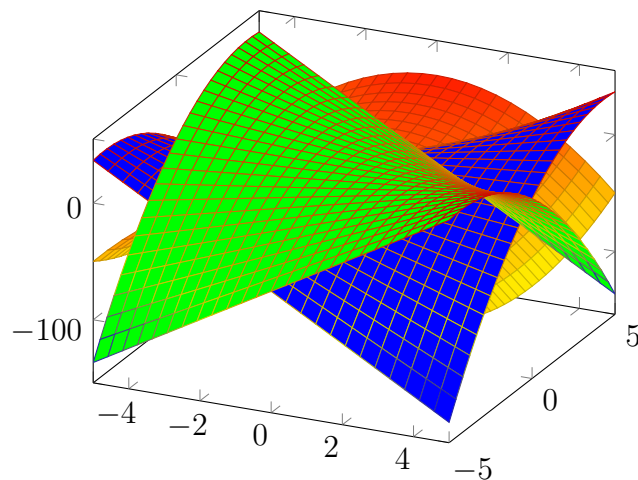


Figura 3.2: MFCQ não garante a condição de segunda ordem

A Figura 3.2 mostra a representação gráfica das superfícies $x_3 = x_2^2 - 3x_1^2$, $x_3 = 2\sqrt{3}x_1x_2 - 2x_2^2$ e $x_3 = -2\sqrt{3}x_1x_2 - 2x_2^2$. O conjunto viável é constituído dos pontos que estão acima das três superfícies. A origem $x^* = 0$ é uma solução ótima para este problema e cumpre MFCQ. As condições KKT ficam

$$\nabla f(x^*) + \mu_1 \nabla g_1(x^*) + \mu_2 \nabla g_2(x^*) + \mu_3 \nabla g_3(x^*) = 0, \quad \mu_1 \geq 0, \mu_2 \geq 0, \mu_3 \geq 0,$$

onde

$$\nabla f(x^*) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \nabla g_1(x^*) = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \quad \nabla g_2(x^*) = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \quad \nabla g_3(x^*) = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}.$$

Logo, o conjunto de multiplicadores de Lagrange fica

$$\Lambda(x^*) = \{\mu \in \mathbb{R}^3 : \mu_1 + \mu_2 + \mu_3 = 1, \mu_1 \geq 0, \mu_2 \geq 0, \mu_3 \geq 0\}.$$

Claramente podemos ver que o conjunto de multiplicadores de Lagrange é compacto e não vazio, o que nos permite concluir que MFCQ se cumpre em x^* . Entretanto, calculando o cone de direções críticas e a hessiana da Lagrangiana, obtemos

$$C^W(x^*) = \{d \in \mathbb{R}^3 : \nabla g_j(x^*)^T d = 0, j \in A(x^*)\} = \{(d_1, d_2, d_3) : d_3 = 0\},$$

$$\nabla_{xx}^2 \mathcal{L}(x^*, \mu) = \mu_1 \begin{pmatrix} 0 & 2\sqrt{3} & 0 \\ 2\sqrt{3} & -4 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \mu_2 \begin{pmatrix} -6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \mu_3 \begin{pmatrix} 0 & -2\sqrt{3} & 0 \\ -2\sqrt{3} & -4 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Portanto, para as direções críticas $\tilde{d} = (1, 0, 0)$ e $\hat{d} = (0, 1, 0)$, segue que não existe $\mu \in \Lambda(x^*)$ tal que

$$(\tilde{d})^T \nabla_{xx}^2 \mathcal{L}(x^*, \mu) \tilde{d} \geq 0 \quad \text{e} \quad (\hat{d})^T \nabla_{xx}^2 \mathcal{L}(x^*, \mu) \hat{d} \geq 0.$$

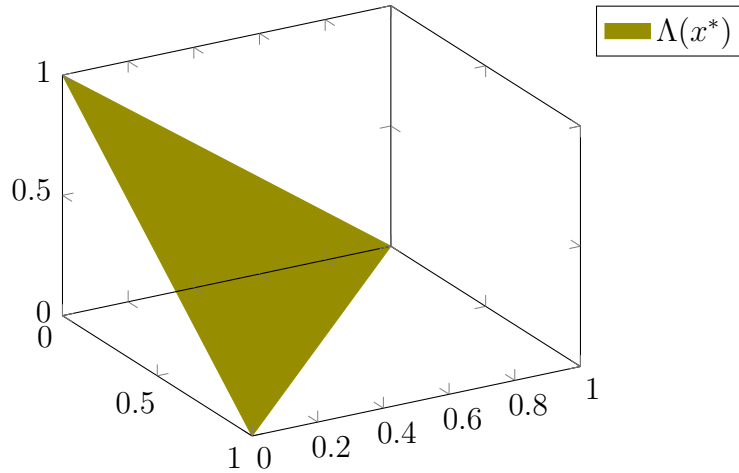


Figura 3.3: Representação do conjunto de multiplicadores de Lagrange

Proposição 3.9 ([41], PROPOSIÇÃO 2.2). *Seja $\{f_i : \mathbb{R}^n \rightarrow \mathbb{R}^m, i \in K\}$ uma família de funções diferenciáveis em $\mathbb{R}^n$ tal que o posto de $\{\nabla f_i(y) : i \in K\}$ é constante para todo y em uma vizinhança de x . Então existe um difeomorfismo local $\phi : U \rightarrow V$, com U e V vizinhanças de x tal que $\phi(x) = x$, $\nabla \phi(x)$ é a matriz identidade e $f_i \circ \phi^{-1}$ ($i \in K$) é constante para todo $y \in E$, onde E é o subespaço definido por*

$$E := \{y \in \mathbb{R}^n : \nabla f_i(x)^T y = 0, i \in K\}.$$

Proposição 3.10. *Suponha que a condição RCRCQ seja válida em $x^* \in \Omega$. Então, para cada $d \in L_\Omega(x^*)$, existe uma curva $\xi : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}^n$, $\varepsilon > 0$ com $\xi(0) = x^*$, $\lim_{t \rightarrow 0^+} \frac{\xi(t) - x^*}{t} = d$ e $\xi(t) \in \Omega$ para todo $t \in [0, \varepsilon)$. Além disso, para todo $j \in A(x^*)$ tal que $\nabla g_j(x^*)^T d = 0$, então $g_j(\xi(t)) = 0$ para todo $t \in (-\varepsilon, \varepsilon)$.*

Demonstração. Suponha que $A(x^*) = \{1, \dots, r\}$. Defina $f_i(x) = h_i(x)$ para todo $i = 1, \dots, m$, $f_{m+j}(x) = g_j(x)$ para todo $j = 1, \dots, r$. Considere o conjunto de índices $K = \{i \in \{1, \dots, m+r\} : \nabla f_i(x^*)^T d = 0\}$. Como vale a condição RCRCQ em x^* , podemos aplicar a Proposição 3.9 para a família de funções $\{f_i : i \in K\}$. Logo, existe um difeomorfismo local ϕ com as propriedades mencionadas. Defina a curva $\xi(t)$ por $\xi(t) = \phi^{-1}(x^* + td)$, para $t > 0$ suficientemente pequeno. Logo, pela continuidade e pela Proposição 3.9, obtemos

$$\lim_{t \rightarrow 0^+} \xi(t) = \xi(0) = x^*, \quad \lim_{t \rightarrow 0^+} \frac{\xi(t) - x^*}{t} = d.$$

Agora, usando o fato de que $f_i \circ \phi^{-1}$ é constante, para cada $i \in K$ e $t > 0$ suficientemente pequeno, temos

$$f_i(\xi(t)) = f_i(\phi^{-1}(x^* + td)) = f_i(\phi^{-1}(x^*)) = f_i(x^*) = 0.$$

Portanto, para todo $j \in A(x^*)$ tal que $\nabla g_j(x^*)^T d = 0$, temos que $g_j(\xi(t)) = 0$ para $t > 0$ suficientemente pequeno. Falta mostrar que $\xi(t) \in \Omega$. Se $j \in A(x^*) \setminus K$, temos $\nabla f_j(x^*)^T d < 0$. Logo, por Taylor, $f_j(\xi(t)) = t \nabla f_j(x^*)^T d + o(t)$, onde $o(t)/t \rightarrow 0$. Assim, $f_j(\xi(t)) < 0$ para $t > 0$ suficientemente pequeno. Por outro lado, se $j \notin A(x^*)$, usando a continuidade de f_j e reduzindo possivelmente o intervalo, obtemos $f_j(\xi(t)) < 0$. $\square$

Andreani, Echagüe e Schuverdt provam em [10] que sob CRCQ todo minimizador local cumpre a condição necessária de segunda ordem forte. Em outras palavras, os autores mostram que CRCQ é uma forte condição de qualificação de segunda ordem. Porém, é possível mostrar que sob uma condição mais fraca que CRCQ, a condição necessária de segunda ordem forte é mantida. A seguir mostramos que a condição RCRCQ é uma condição de qualificação de segunda ordem no sentido que implica SSONC.

Teorema 3.11. *Seja $x^* \in \Omega$ um minimizador local que cumpre RCRCQ. Então, para qualquer multiplicador de Lagrange $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ associado a x^* , a condição necessária forte de segunda ordem se mantém em x^* .*

Demonstração. Pelo fato de RCRCQ ser uma condição de qualificação, a existência dos multiplicadores de Lagrange está garantida. Seja $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ um multiplicador de Lagrange associado a x^* . Dado $d \in C^S(x^*)$, temos

$$\begin{aligned} \nabla h_i(x^*)^T d &= 0, & i &= 1, \dots, m, \\ \nabla g_j(x^*)^T d &= 0, & j &\in A^+(x^*), \\ \nabla g_j(x^*)^T d &\leq 0, & j &\in A^0(x^*). \end{aligned}$$

Logo, temos

$$\nabla f(x^*)^T d + \sum_{i=1}^m \lambda_i \nabla h_i(x^*)^T d - \sum_{j \in A^+(x^*)} \mu_j \nabla g_j(x^*)^T d = 0, \quad \nabla f(x^*)^T d = 0.$$

Pela Proposição 3.10, segue que existe uma curva $\xi(t) \in \Omega$, $t \in [0, \varepsilon)$ tal que

$$\xi(0) = x^*, \quad \xi'(0) = \lim_{t \rightarrow 0^+} \frac{\xi(t) - x^*}{t} = d, \quad g_j(\xi(t)) = 0, \quad \forall j \in A^+(x^*).$$

Defina $\phi(t) = f(\xi(t))$. Como x^* é minimizador local de f em Ω e $\xi(0) = x^*$, segue que $t^* = 0$ é minimizador local de ϕ . Logo, $\phi''(0) \geq 0$ e, como $\phi''(t) = \xi'(t)^T \nabla^2 f(\xi(t)) \xi'(t) + \nabla f(\xi(t))^T \xi''(t)$, temos

$$\phi''(0) = d^T \nabla^2 f(x^*) d + \nabla f(x^*)^T \xi''(0) \geq 0. \quad (3.6)$$

Além disso, como $\xi(t) \in \Omega$ e usando que $g_j(\xi(t)) = 0$ para todo $j \in A^+(x^*)$, temos que

$$R(t) = \sum_{i=1}^m \lambda_i h_i(\xi(t)) + \sum_{j \in A(x^*)} \mu_j g_j(\xi(t)) = 0, \quad t \in [0, \varepsilon).$$

Logo, derivando duas vezes com relação a t , obtemos

$$\begin{aligned} R''(0) &= d^T \left(\sum_{i=1}^m \lambda_i \nabla^2 h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla^2 g_j(x^*) \right) d + \\ &\quad + \left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*) \right)^T \xi''(0) = 0. \end{aligned} \quad (3.7)$$

Somando (3.6) com (3.7) e usando a hipótese de x^* ser um ponto KKT, obtemos

$$\begin{aligned} \phi''(0) + R''(0) &= d^T \left(\nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla^2 g_j(x^*) \right) d + \\ &\quad + \left(\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*) \right)^T \xi''(0) \\ &= d^T \left(\nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla^2 g_j(x^*) \right) d \geq 0. \end{aligned}$$

Portanto, a condição necessária forte de segunda ordem vale em x^* como queríamos provar. $\square$

Podemos encontrar uma versão mais fraca da condição RCRCQ na literatura, exigindo posto constante apenas para o conjunto de todos os gradientes de restrições ativas.

Definição 3.12. Dizemos que a condição *Weak Constant Rank (WCR)* vale no ponto viável $x \in \Omega$ se existe uma vizinhança V de x tal que o posto de $\{\nabla h_i(y), \nabla g_j(y) : i = 1, \dots, m; j \in A(x)\}$ é constante para todo $y \in V$.

Contra-Exemplo 3.13 (WCR não implica KKT). Considere o seguinte problema

$$\text{minimizar } f(x) = -x_1 \text{ s.a. } h_1(x) = x_2 - x_1^2 = 0, g_1(x) = -x_1 \leq 0, g_2(x) = x_2 \leq 0.$$

O ponto $x^* = (0, 0)$ é a solução deste problema e não cumpre as condições KKT, pois

$$\nabla f(x^*) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \nabla h_1(x^*) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \nabla g_1(x^*) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \nabla g_2(x^*) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Entretanto, existe uma vizinhança V da origem tal que o posto de $\{\nabla h_1(x), \nabla g_j(x) : j = 1, 2\}$ é constante para todo $x \in V$, e portanto, x^* cumpre WCR. Em outras palavras, WCR não é uma condição de qualificação. Porém, um ponto WCR que cumpre KKT, deve satisfazer WSONC. No Teorema a seguir, provamos que, sob WCR a condição necessária de segunda ordem fraca vale em um minimizador, desde que as restrições cumprem alguma condição de qualificação de primeira ordem.

Teorema 3.14. *Seja $x^* \in \Omega$ um minimizador local que cumpre KKT. Se x^* satisfaz WCR, então para qualquer multiplicador de Lagrange $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ associado a x^* , a condição necessária fraca de segunda ordem se mantém em x^* .*

Demonstração. Dado $d \in C^W(x^*)$, temos que $\nabla h_i(x^*)^T d = 0$ para todo $i = 1, \dots, m$ e $\nabla g_j(x^*)^T d = 0$ para todo $j \in A(x^*)$ e, portanto, $d \in L_\Omega(x^*)$. Como vale WCR, a Proposição 3.10 pode ser aplicada para garantir a existência de uma curva viável satisfazendo as propriedades mencionadas. Assim, a demonstração segue como no Teorema 3.11. $\square$

Em [10], os autores provam que WCR é uma condição de qualificação de segunda ordem sempre que o problema possuir apenas restrições de igualdade.

Teorema 3.15 ([10], TEOREMA 3.3). *Seja x^* um minimizador local do problema de minimizar $f(x)$ sujeito a $h(x) = 0$. Se x^* satisfaz WCR, então x^* satisfaz as condições KKT e a condição necessária forte de segunda ordem, para qualquer multiplicador de Lagrange $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ associado a x^* .*

Observe que, quando o problema têm apenas restrições de igualdade, o cone crítico fraco coincide com o cone crítico forte e, portanto, a condição necessária de segunda ordem fraca é equivalente à condição forte. A condição WCR para problemas de igualdade é mais fraca do que a condição MFCQ (que neste caso é equivalente a condição de independência linear dos gradientes) e mais fraca do que a condição de posto constante. Portanto, a condição de posto constante pode ser vista como uma generalização dessas condições.¹

Considere $x^* \in \Omega$. Definimos o conjunto viável auxiliar

$$\Omega' := \{x \in \mathbb{R}^n : h_i(x) = 0, i = 1, \dots, m; g_j(x) = 0, j \in A(x^*)\}.$$

Observe que, se $A(x^*) = \emptyset$, então $\Omega' = \{x \in \mathbb{R}^n : h_i(x) = 0, i = 1, \dots, m\}$, mas se $A(x^*) \neq \emptyset$, sempre temos $x^* \in \Omega'$. Note também que a condição WCR independe se considerarmos $x^* \in \Omega$ ou $x^* \in \Omega'$. O próximo Teorema mostra que WCR implica a condição Abadie para o conjunto Ω' .

Teorema 3.16. *Se $x^* \in \Omega$ satisfaz WCR, então x^* cumpre a condição de Abadie para o conjunto Ω' .*

Demonstração. Devemos mostrar que $L_{\Omega'}(x^*) \subset T_{\Omega'}(x^*)$. Seja $0 \neq d \in L_{\Omega'}(x^*)$. Como vale WCR em x^* , podemos aplicar a Proposição 3.9 para construir uma curva $\xi : (-\varepsilon, \varepsilon) \rightarrow \Omega'$ com $\varepsilon > 0$, $\xi(0) = x^*$ e $\xi'(0) = d$. Logo, $d = \lim_{t \rightarrow 0^+} (\xi(t) - x^*)/t$. Como $d = \xi'(0) \neq 0$, podemos reduzir ε se necessário de modo que $\xi(t) \neq x^*$. Assim,

$$\frac{\xi(t) - x^*}{\|\xi(t) - x^*\|} = \frac{\xi(t) - x^*}{t} \frac{t}{\|\xi(t) - x^*\|} \rightarrow \frac{d}{\|d\|} \text{ quando } t \rightarrow 0^+,$$

o que mostra que $d \in T_{\Omega'}(x^*)$. $\square$

¹Andreani, Echagüe e Schuverdt [10]

Na sequência prosseguimos com um Lema que será útil para demonstrar outros resultados.

Lema 3.17. *Sejam $x^* \in \Omega$ um minimizador local e $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ um multiplicador de Lagrange associado a x^* . Considere uma direção não nula $d \in \mathbb{R}^n$ e uma sequência $\{x^k\} \subset \Omega$ com $x^k \rightarrow x^*$ satisfazendo as seguintes propriedades*

- $\frac{x^k - x^*}{\|x^k - x^*\|} \rightarrow \frac{d}{\|d\|},$
- $h_i(x^k) = o(\|x^k - x^*\|^2),$ para todo $i = 1, \dots, m$ com $\lambda_i \neq 0,$
- $g_j(x^k) = o(\|x^k - x^*\|^2),$ para todo $j = 1, \dots, p$ com $\mu_j > 0.$

Então,

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

Demonstração. Primeiro observe que

$$\mathcal{L}(x^k, \lambda, \mu) = f(x^k) + \sum_{i=1}^m \lambda_i h_i(x^k) + \sum_{j \in A(x^*)} \mu_j g_j(x^k) = f(x^k) + o(\|x^k - x^*\|^2).$$

Como x^* é minimizador local e $\{x^k\}$ é uma sequência de pontos viáveis, temos

$$\begin{aligned} 0 &\leq f(x^k) - f(x^*) \\ &= \mathcal{L}(x^k, \lambda, \mu) - \mathcal{L}(x^*, \lambda, \mu) + o(\|x^k - x^*\|^2) \\ &= \nabla_x \mathcal{L}(x^*, \lambda, \mu)^T (x^k - x^*) + \frac{1}{2} (x^k - x^*)^T \nabla_{xx}^2 \mathcal{L}(\bar{x}^k, \lambda, \mu) (x^k - x^*) + o(\|x^k - x^*\|^2) \\ &= \frac{1}{2} (x^k - x^*)^T \nabla_{xx}^2 \mathcal{L}(\bar{x}^k, \lambda, \mu) (x^k - x^*) + o(\|x^k - x^*\|^2), \end{aligned}$$

onde $\bar{x}^k$ pertence ao segmento de reta que liga x^* a x^k . Dividindo a desigualdade acima por $\|x^k - x^*\|^2$ e tomando o limite em k , obtemos

$$d^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda, \mu) d \geq 0,$$

completando a prova. □

Note que o resultado acima estabelece uma relação entre o cone de direções tangentes com a condição de otimalidade de segunda ordem. De fato, o conjunto de direções que satisfazem as hipóteses do Lema 3.17 é um cone contido no cone tangente.

Teorema 3.18. *Sejam $x^* \in \Omega$ um minimizador local e $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ um multiplicador de Lagrange associado a x^* . Suponha que x^* satisfaz a condição de qualificação de Abadie para o conjunto $\Omega' = \{x \in \mathbb{R}^n : h_i(x) = 0, i = 1, \dots, m; g_j(x) = 0, j \in A(x^*)\}$. Então, a condição de otimalidade fraca de segunda ordem é satisfeita em x^* para o multiplicador (λ, μ) .*

Demonstração. Primeiramente, observe que $C^W(x^*) = L_{\Omega'}(x^*)$. Como vale Abadie para Ω' , então $L_{\Omega'}(x^*) = T_{\Omega'}(x^*)$. Logo, para qualquer $0 \neq d \in C^W(x^*)$, existe uma sequência $\{x^k\}$ de pontos viáveis com $x^k \rightarrow x^*$ tal que

$$\frac{x^k - x^*}{\|x^k - x^*\|} \rightarrow \frac{d}{\|d\|}.$$

Portanto, pelo Lema 3.17, concluímos que $d^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda, \mu) d \geq 0$ e o resultado segue. $\square$

Note que, quando o problema têm apenas restrições de igualdade, ou seja, quando $\Omega = \Omega'$ com $A(x^*) = \emptyset$, o Teorema 3.18 mostra que a condição de Abadie garante a validade da condição de otimalidade de segunda ordem fraca para qualquer multiplicador de Lagrange. Resumindo, temos o seguinte Corolário:

Corolário 3.19. *Seja $x^* \in \Omega$ um minimizador local que satisfaz a condição de Abadie e suponha que Ω só possui restrições de igualdade. Então, x^* é um ponto KKT e x^* cumpre a condição de otimalidade (forte) de segunda ordem para qualquer multiplicador de Lagrange associado a x^* .*

Demonstração. Como Ω só possui restrições de igualdade, temos que $C^W(x^*) = C^S(x^*)$. Além disso, temos $A(x^*) = \emptyset$ e a suposição de Abadie feita no Teorema 3.18 pode ser aplicada para o conjunto Ω e o resultado segue. $\square$

Lema 3.20 ([6], COROLÁRIO 3.2). *Se $x \in \Omega$ satisfaz as condições KKT, então existe $d \in C^S(x)$ tal que*

$$\nabla h_i(x)^T d = 0, i = 1, \dots, m, \quad \nabla g_j(x)^T d = 0, j \in A^+(x), \quad \nabla g_j(x)^T d < 0, j \in A^0(x).$$

Teorema 3.21. *Seja $x^* \in \Omega$ um minimizador local e suponha que x^* cumpre a condição de qualificação de Abadie para o conjunto de restrições*

$$\Omega^+ := \{x \in \mathbb{R}^n : h_i(x) = 0, i = 1, \dots, m; g_j(x) = 0, j \in A^+(x^*); g_j(x) \leq 0, j \in A^0(x^*)\}.$$

Então, a condição de otimalidade forte de segunda é satisfeita para qualquer multiplicador de Lagrange $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ associado a x^ .*

Demonstração. Dado $d \in C^S(x^*)$ com $d \neq 0$, temos que

$$\nabla h_i(x^*)^T d = 0, i = 1, \dots, m, \quad \nabla g_j(x^*)^T d = 0, j \in A^+(x^*) \text{ e } \nabla g_j(x^*)^T d \leq 0, j \in A^0(x^*).$$

Logo, usando a hipótese, segue que $d \in L_{\Omega^+}(x^*) = T_{\Omega^+}(x^*)$. Assim, existe uma sequência $\{x^k\} \subset \Omega^+$ tal que

$$\frac{x^k - x^*}{\|x^k - x^*\|} \rightarrow \frac{d}{\|d\|}.$$

Pelo Lema 3.17, segue que $d^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda, \mu) d \geq 0$. $\square$

Teorema 3.22. *Se x^* é um minimizador local, então*

$$C^S(x^*) \subset T_{\Omega^+}(x^*) \Leftrightarrow C^S(x^*) \subset T_{F^+}(x^*).$$

Demonstração. A implicação direta segue de $\Omega^+ \subset F^+$. Para provar a recíproca, seja $d \in C^S(x^*)$ com $d \neq 0$ e suponha sem perda de generalidade que $\|d\| = 1$. Considere $s \in C^S(x^*)$ dado pelo Lema 3.20, logo temos que

$$\nabla h_i(x^*)^T s = 0, \quad i = 1, \dots, m, \quad \nabla g_j(x^*)^T s = 0, \quad j \in A^+(x^*) \quad \text{e} \quad \nabla g_j(x^*)^T s < 0, \quad j \in A^0(x^*).$$

Defina a sequência $\{d^k\}$ por

$$d^k := \frac{d + (1/k)s}{\|d + (1/k)s\|}.$$

Logo, segue que

$$\begin{aligned} \nabla h_i(x^*)^T d^k &= 0, \quad i = 1, \dots, m, \quad \nabla g_j(x^*)^T d^k = 0, \quad j \in A^+(x^*), \\ \nabla g_j(x^*)^T d^k &< 0, \quad j \in A^0(x^*), \quad \|d^k\| = 1 \quad \text{e} \quad d^k \rightarrow d. \end{aligned}$$

Ou seja, $d^k \in C^S(x^*)$ para todo $k \in \mathbb{N}$. Usando a hipótese, temos que $d^k \in T_{F^+}(x^*)$ para todo k . Logo existe uma sequência $\{x^\ell\} \subset F^+$ com $x^\ell \rightarrow x^*$ tal que

$$\frac{x^\ell - x^*}{\|x^\ell - x^*\|} \rightarrow d^k.$$

Mostraremos que $g_j(x^\ell) < 0$, $j \in A^0(x^*)$ para ℓ suficientemente grande. Para $j \in A^0(x^*)$, temos que

$$g_j(x^\ell) = g_j(x^*) + \nabla g_j(\bar{x}^\ell)^T (x^\ell - x^*) = \nabla g_j(\bar{x}^\ell)^T (x^\ell - x^*)$$

para algum $\bar{x}^\ell$ entre x^* e x^ℓ . Assim, como $\nabla g_j(\bar{x}^\ell) \rightarrow \nabla g_j(x^*)$, $(x^\ell - x^*)/\|x^\ell - x^*\| \rightarrow d^k$ e $\nabla g_j(x^*)^T d^k < 0$, segue que

$$\frac{g_j(x^\ell)}{\|x^\ell - x^*\|} = \nabla g_j(\bar{x}^\ell)^T \frac{(x^\ell - x^*)}{\|x^\ell - x^*\|} \rightarrow \nabla g_j(x^*)^T d^k < 0.$$

Portanto $g_j(x^\ell) < 0$ para ℓ suficientemente grande. Logo, $\{x^\ell\} \subset \Omega^+$, $x^\ell \rightarrow x^*$ e $\frac{x^\ell - x^*}{\|x^\ell - x^*\|} \rightarrow d^k$ e, portanto, $d^k \in T_{\Omega^+}(x^*)$. Além disso, como $d^k \rightarrow d$ e $T_{\Omega^+}(x^*)$ é fechado, obtemos $d \in T_{\Omega^+}(x^*)$. $\square$

Teorema 3.23. *Seja x^* um minimizador local e $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ um multiplicador de Lagrange associado a x^* . Defina $A_\mu^+ := \{j \in A(x^*) : \mu_j > 0\}$ e $A_\mu^0 := A(x^*) \setminus A_\mu^+$. Se x^* cumpre a condição de qualificação de Abadie para o conjunto*

$$\Omega_\mu := \{x \in \mathbb{R}^n : h_i(x) = 0, i = 1, \dots, m; g_j(x) = 0, j \in A_\mu^+; g_j(x) \leq 0, j \in A_\mu^0\},$$

então, a condição de otimalidade forte de segunda é satisfeita em x^ para o multiplicador (λ, μ) .*

Para qualquer multiplicador de Lagrange (λ, μ) temos, $C^S(x^*) = L_{\Omega_\mu}(x^*)$. Observe que o Teorema anterior tem uma hipótese diferente para cada multiplicador, ele afirma que se vale Abadie para o conjunto Ω_μ , então vale a condição de otimalidade de segunda ordem para o multiplicador (λ, μ) . Portanto, para garantir a validade da condição de otimalidade segunda ordem para todos os multiplicadores, a condição de Abadie deve ser mantida em cada conjunto viável associado. Em outras palavras, a condição de segunda ordem pode ser satisfeita para algum multiplicador, mas não para todos. O exemplo a seguir mostra que o Teorema 3.23 não garante a validade da condição de segunda ordem para todos os multiplicadores de Lagrange.

Exemplo 3.24. Considere o seguinte problema

$$\min. \quad f(x) = x_2 \quad \text{s.a.} \quad g_1(x) = -x_1^2 - x_2 \leq 0, g_2(x) = -x_2 \leq 0, g_3(x) = x_1 \leq 0.$$

Claramente o ponto $x^* = (0, 0)$ é uma solução, onde

$$\Lambda(x^*) = \{(\mu_1, \mu_2, \mu_3) : \mu_1 + \mu_2 = 1, \mu_3 = 0\}.$$

Se considerarmos um multiplicador $0 \neq \mu \in \Lambda(x^*)$, por exemplo $\mu = (1/2, 1/2, 0)$, temos que $A_\mu^+ = \{1, 2\}$, $A_\mu^0 = \{3\}$ e $\Omega_\mu = \Omega^+ = \{(0, 0)\}$. Calculando o cone crítico forte, obtemos $C^S(x^*) = \{(d_1, d_2) : d_1 \leq 0, d_2 = 0\}$. Claramente $T_{\Omega_\mu}(x^*)$ não contém $C^S(x^*) = L_{\Omega_\mu}(x^*)$. Entretanto, tomando $\tilde{\mu} = (0, 1, 0)$, temos $\Omega_{\tilde{\mu}} = \{(x_1, x_2) : x_1 \leq 0, x_2 = 0\}$ e, portanto, $\Omega_{\tilde{\mu}} = C^S(x^*)$. Assim, a validade da condição de otimalidade forte de segunda é garantida apenas para $\tilde{\mu}$.

3.2 Condição sequencial AKKT de segunda ordem

Vejamos agora uma condição sequencial de otimalidade de segunda ordem, chamada AKKT de segunda ordem. Introduzida por Ramos [58] em seu doutoramento, essa condição é uma genuína condição de otimalidade apropriada para a análise de algoritmos com convergência a pontos estacionários de segunda ordem. Muitos desses algoritmos geram sequências cujos pontos limites satisfazem esta condição.

Definição 3.25 ([58]). Dizemos que um ponto viável $x^* \in \Omega$ cumpre a condição *Aproximadamente KKT de segunda ordem (AKKT2)* se existem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\eta^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}_+^p$, $\{\theta^k\} \subset \mathbb{R}_+^p$, $\{\delta_k\} \subset \mathbb{R}_+$ com $\mu_j^k = 0$ para $j \notin A(x^*)$, $\theta_j^k = 0$ para $j \notin A(x^*)$ tais que $x^k \rightarrow x^*$, $\delta_k \rightarrow 0$,

$$\lim_{k \rightarrow \infty} \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) = 0 \quad (3.8)$$

e

$$\nabla^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^m \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j \in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succeq -\delta_k \mathbb{I} \quad (3.9)$$

para todo k suficientemente grande, onde $A \succeq B$ significa que as matrizes simétricas A e B são tais que $A - B$ é semi-definida positiva e $\mathbb{I}$ denota a matriz identidade. Dizemos que $\{x^k\}$ é uma sequência AKKT2.

A demonstração de que AKKT2 é uma condição de otimalidade, pode ser encontrada em [4, 58]

Teorema 3.26 ([4], TEOREMA 3.4). *Se x^* é um minimizador local, então x^* satisfaz a condição AKKT2.*

Por (3.8), podemos ver claramente que a condição AKKT2 implica a condição sequencial AKKT. Porém, essas condições não são equivalentes, como mostra o exemplo a seguir.

Exemplo 3.27. Considere o problema de minimizar $f(x) = -x_1 - x_2$ sujeito a $g(x) = x_1^2 x_2^2 - 1 \leq 0$. No ponto $x^* = (1, 1)$, as condições KKT ficam

$$\nabla f(x^*) + \mu \nabla g(x^*) = 0 \Rightarrow \mu = \frac{1}{2}.$$

Portanto vale AKKT. Agora suponha que x^* satisfaz AKKT2. Defina $d^k = (x_1^k, -x_2^k)$ onde $x^k = (x_1^k, x_2^k)$ é a sequência dada pela definição de AKKT2. Como $\nabla g(x^k)^T d^k = 0$ para todo k , por (3.9), segue que

$$\mu^k (d^k)^T \nabla^2 g(x^k) d^k + \delta_k \|d^k\|^2 = -4\mu^k (x_1^k x_2^k)^2 + \delta_k \|d^k\|^2 \geq 0, \quad (3.10)$$

para algum $\mu^k \geq 0$, $\delta_k \geq 0$ com $\delta_k \rightarrow 0$. Por outro lado, por (3.8), temos

$$\nabla f(x^k) + \mu^k \nabla g(x^k) \rightarrow 0 \Leftrightarrow 2\mu^k x_1^k (x_2^k)^2 \rightarrow 1$$

e como $x_1^k \rightarrow 1$, temos $2\mu^k (x_1^k x_2^k)^2 \rightarrow 1$. Logo $-4\mu^k (x_1^k x_2^k)^2 + \delta_k \|d^k\|^2 \rightarrow -2$, contradizendo (3.10). Portanto não vale AKKT2 em x^* .

Sabe-se que a condição AKKT2 é uma condição de otimalidade de segunda ordem no sentido que implica “WSOnc ou não-CQ2” para condições de qualificação fracas. Em seguida, mostramos que a condição conjunta MFCQ + WCR garante a validade da condição WSONC em um minimizador. Para provar isso, considere o seguinte Lema:

Lema 3.28 ([13], LEMA 3.1). *Se $d \in C^W(x^*)$, então para toda sequência $\{x^k\} \subset \mathbb{R}^n$ com $x^k \rightarrow x^*$, existe uma sequência $\{d^k\} \subset \mathbb{R}^n$ tal que $d^k \in C^W(x^k)$ para todo k e $d^k \rightarrow d$.*

Proposição 3.29. *Seja $x^* \in \Omega$ tal que AKKT2 vale. Se a condição de qualificação MFCQ + WCR se cumpre em x^* , então vale WSONC em x^* .*

Demonstração. Pela definição de AKKT2, existem sequências $\{x^k\}$, $\{\lambda^k\}$, $\{\mu^k\}$, $\{\eta^k\}$, $\{\theta^k\}$ com $\mu_j^k = 0$ e $\theta_j^k = 0$ para $j \notin A(x^*)$ tais que $x^k \rightarrow x^*$, $\delta_k \rightarrow 0$ e

$$\lim_{k \rightarrow \infty} \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) = 0; \quad (3.11)$$

$$\nabla_x^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^m \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j \in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succeq -\delta_k \mathbb{I}. \quad (3.12)$$

Afirmamos que a sequência $\{(\lambda^k, \mu^k)\}$ é limitada. De fato, suponha por absurdo que a sequência $\{(\lambda^k, \mu^k)\}$ seja ilimitada. Defina a sequência $\{M_k\}$ por $M_k := \max\{|\lambda_i^k|, \mu_j^k : i = 1, \dots, m; j \in A(x^*)\}$. Claramente, temos que $M_k \rightarrow \infty$. A sequência $\{(\frac{\lambda^k}{M_k}, \frac{\mu^k}{M_k})\}$ é limitada, pois $\|(\frac{\lambda^k}{M_k}, \frac{\mu^k}{M_k})\|_\infty = 1$. Logo, dividindo a expressão (3.11) por M_k e tomando o limite em subsequência convergente, obtemos $\sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j^* \nabla g_j(x^*) = 0$ com $\|(\lambda^*, \mu^*)\|_\infty = 1$, contradizendo a definição de MFCQ. Logo a sequência $\{(\lambda^k, \mu^k)\}$ é limitada. Suponha sem perda de generalidade que $(\lambda^k, \mu^k) \rightarrow (\lambda^*, \mu^*)$. Defina $\mu_j^* = 0$ para todo $j \notin A(x^*)$. Passando ao limite em (3.11), concluímos que x^* é um ponto KKT. Vamos provar que WSONC vale em x^* para o multiplicador (λ^*, μ^*) . Dado $d \in C^W(x^*)$, pelo Lema 3.28 existe uma sequência $\{d^k\} \subset \mathbb{R}^n$ tal que $d^k \in C^W(x^k)$ e $d^k \rightarrow d$. Logo,

$\nabla h_i(x^k)^T d^k = 0$ para todo $i = 1, \dots, m$ e $\nabla g_j(x^k)^T d^k = 0$ para todo $j \in A(x^*)$. Assim, usando (3.12), obtemos

$$(d^k)^T \nabla_x^2 \mathcal{L}(x^k, \lambda^k, \mu^k) d^k \geq -\delta_k \|d^k\|^2.$$

Tomando o limite na desigualdade acima, segue que

$$d^T \left(\nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(x^*) + \sum_{j \in A(x^*)} \mu_j^* \nabla^2 g_j(x^*) \right) d \geq 0,$$

completando a prova. $\square$

Existem muitos algoritmos na literatura que geram seqüências cujos pontos limites satisfazem a condição AKKT2 (ver [4, 58]). A seguir mostramos que o algoritmo Lagrangiano aumentado proposto em [13] gera seqüências AKKT2. Para isso, considere a função Lagrangiano aumentada

$$\mathcal{L}_\rho(x, \lambda, \mu) := f(x) + \frac{\rho}{2} \left[\sum_{i=1}^m \left(h_i(x) + \frac{\lambda_i}{\rho} \right)^2 + \sum_{j=1}^p \left(\max \left\{ 0, g_j(x) + \frac{\mu_j}{\rho} \right\} \right)^2 \right],$$

com $\rho > 0$ e $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$. Defina

$$\bar{\nabla}^2 \left(\max \left\{ 0, g_j(x) + \frac{\mu_j}{\rho} \right\} \right)^2 := \begin{cases} \nabla^2 \left(g_j(x) + \frac{\mu_j}{\rho} \right)^2, & \text{se } g_j(x) + \frac{\mu_j}{\rho} = 0, \\ \nabla^2 \left(\max \left\{ 0, g_j(x) + \frac{\mu_j}{\rho} \right\} \right)^2, & \text{caso contrário.} \end{cases}$$

Algoritmo 3.30. *Sejam $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, $\gamma > 1$, $\rho_1 > 0$ e $\tau \in (0, 1)$. Tome $\{\varepsilon_k\} \subset \mathbb{R}_+$ com $\varepsilon_k \rightarrow 0$. Escolha $x^0 \in \mathbb{R}^n$ como ponto inicial e $\lambda^1 \in [\lambda_{\min}, \lambda_{\max}]^m$, $\mu^1 \in [0, \mu_{\max}]^p$. Defina $V_j^0 = g_j^+(x)$, $j = 1, \dots, p$. Inicialize $k \leftarrow 1$.*

(1) (Minimização). *Determine $x^k \in \mathbb{R}^n$ que satisfaça:*

$$\|\nabla \mathcal{L}_{\rho_k}(x^k, \lambda^k, \mu^k)\| \leq \varepsilon_k \quad \text{e} \quad \bar{\nabla}^2 \mathcal{L}_{\rho_k}(x^k, \lambda^k, \mu^k) \succeq -\varepsilon_k \mathbb{I}.$$

(2) (Atualização do parâmetro de penalidade). *Para todo $j = 1, \dots, p$, defina*

$$V_j^k := \max\{g_j(x^k), -\mu_j^k/\rho_k\}.$$

Se $\max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\} \leq \tau \max\{\|h(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\}$,

defina $\rho_{k+1} = \rho_k$. Caso contrário, coloque $\rho_{k+1} = \gamma \rho_k$;

(3) (Atualização dos multiplicadores de Lagrange). *Calcule*

$$\bar{\lambda}_i^k := \lambda_i^k + \rho_k h_i(x^k), \quad i = 1, \dots, m \quad \text{e} \quad \bar{\mu}_j^k := \max\{0, \mu_j^k + \rho_k g_j(x^k)\}, \quad j = 1, \dots, p.$$

Defina $\lambda_i^{k+1} := \text{proj}_{[\lambda_{\min}, \lambda_{\max}]}(\bar{\lambda}_i^k) \in [\lambda_{\min}, \lambda_{\max}]$, para todo $i = 1, \dots, m$ e $\mu_j^{k+1} := \text{proj}_{[0, \mu_{\max}]}(\bar{\mu}_j^k) \in [0, \mu_{\max}]$, para todo $j = 1, \dots, p$. Faça $k \leftarrow k+1$ e vá para o passo (1).

Teorema 3.31. *Suponha que a sequência $\{x^k\}$ gerada pelo Algoritmo 3.30 admite um ponto limite $x^* \in \Omega$. Então x^* é um ponto AKKT2.*

Demonstração. Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 3.30 tal que $x^k \rightarrow x^*$. Pelo passo (1) do algoritmo, para todo k suficientemente grande, temos

$$\begin{aligned} \|\nabla \mathcal{L}_{\rho_k}(x^k, \lambda^k, \mu^k)\| &= \|\nabla f(x^k) + \sum_{i=1}^m (\lambda_i^k + \rho_k h_i(x^k)) \nabla h_i(x^k) + \\ &\quad + \sum_{j=1}^p \max\{0, \mu_j^k + \rho_k g_j(x^k)\} \nabla g_j(x^k)\| \leq \varepsilon_k. \end{aligned} \quad (3.13)$$

Defina $\hat{\lambda}_i^k := \lambda_i^k + \rho_k h_i(x^k)$ e $\hat{\mu}_j^k := \max\{0, \mu_j^k + \rho_k g_j(x^k)\}$. Usando as mesmas idéias da demonstração do Teorema 2.45, segue que $\hat{\mu}_j^k = 0$, $j \notin A(x^*)$ para todo k suficientemente grande. Além disso,

$$\nabla^2 L(x^k, \hat{\lambda}^k, \hat{\mu}^k) + \rho_k \sum_{i=1}^m \nabla h_i(x^k) \nabla h_i(x^k)^T + \rho_k \sum_{j \in A(x^*)} \nabla g_j(x^k) \nabla g_j(x^k)^T \succeq -\varepsilon_k \mathbb{I}. \quad (3.14)$$

Por (3.13) e (3.14), segue que x^* satisfaz a condição AKKT2. $\square$

Em [4] os autores mostram que o método SQP regularizado de Gill et al. [34] e o método de regiões de confiança de Dennis e Vicente [28] também geram sequências AKKT2.

3.2.1 Propriedade do cone contínuo de segunda ordem

Aqui apresentamos a condição de qualificação de segunda ordem mínima associada à condição sequencial AKKT2. Dado $x \in \mathbb{R}^n$, definimos o cone

$$C^{\mathcal{W}}(x, x^*) := \{d \in \mathbb{R}^n : \nabla h_i(x)^T d = 0, i = 1, \dots, m; \nabla g_j(x)^T d = 0, j \in A(x^*)\}.$$

O cone $C^{\mathcal{W}}(x, x^*)$ pode ser visto como uma perturbação do cone crítico fraco $C^{\mathcal{W}}(x^*)$ ao redor de $x^* \in \Omega$. Um fato importante é que WCR é equivalente à semi-continuidade interior da multifunção $x \in \mathbb{R}^n \rightrightarrows C^{\mathcal{W}}(x, x^*)$ em $x = x^*$. Ou seja, $x^* \in \Omega$ satisfaz WCR se, e somente se,

$$C^{\mathcal{W}}(x^*, x^*) \subset \liminf_{x \rightarrow x^*} C^{\mathcal{W}}(x, x^*).$$

Afim de definir a condição mínima associada à condição AKKT2, considere $x \in \mathbb{R}^n$ e $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ com $\mu_j = 0$, $j \notin A(x^*)$. Defina o seguinte cone

$$K_2^{\mathcal{W}}(x) := \bigcup_{(\lambda, \mu)} S^{\mathcal{W}}(\lambda, \mu, x),$$

onde

$$S^{\mathcal{W}}(\lambda, \mu, x) := \left\{ \begin{array}{l} (\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x), H) \quad \text{tal que} \\ H \preceq \sum_{i=1}^m \lambda_i \nabla^2 h_i(x) + \sum_{j \in A(x^*)} \mu_j \nabla^2 g_j(x) \quad \text{sobre } C^{\mathcal{W}}(x, x^*) \end{array} \right\}.$$

$K_2^{\mathcal{W}}(x)$ é um cone convexo contido em $\mathbb{R}^n \times \text{Sym}(n)$. Usando essa definição, podemos escrever a condição de otimalidade fraca de segunda ordem na forma:

$$(-\nabla f(x^*), -\nabla^2 f(x^*)) \in K_2^{\mathcal{W}}(x^*).$$

Definição 3.32 (CCP2, [4]). Dizemos que $x^* \in \Omega$ satisfaz a *propriedade do cone contínuo de segunda ordem (CCP2)* se a multifunção $x \in \mathbb{R}^n \Rightarrow K_2^{\mathcal{W}}(x)$ é semi-contínua exteriormente em x^* , ou seja,

$$\limsup_{x \rightarrow x^*} K_2^{\mathcal{W}}(x) \subset K_2^{\mathcal{W}}(x^*).$$

Lema 3.33. *Sejam $P \in \text{Sym}(n)$ e $a_1, \dots, a_r \in \mathbb{R}^n$. Considere o subespaço $\mathcal{C} = \{d \in \mathbb{R}^n : a_j^T d = 0, j = 1, \dots, r\}$. Suponha que $v^T P v > 0$ para todo $v \in \mathcal{C}$. Então existem $c_j \geq 0, j = 1, \dots, r$ tais que $P + \sum_{j=1}^r c_j a_j a_j^T$ é definida positiva.*

O Teorema a seguir mostra que CCP2 é a condição mínima para o qual a condição AKKT2 implica a condição WSONC.

Teorema 3.34 ([4], TEOREMA 4.2). *Para toda função objetivo f tal que AKKT2 implica a condição WSONC, então CCP2 vale. Reciprocamente, se vale CCP2, então AKKT2 implica a condição WSONC, para toda função objetivo f .*

Demonstração. Seja $x^* \in \Omega$. Suponha que CCP2 vale em x^* e existe uma função objetivo f tal que AKKT2 vale. Pela definição de AKKT2, existem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\eta^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}_+^p$, $\{\theta^k\} \subset \mathbb{R}_+^p$, $\{\delta_k\} \subset \mathbb{R}_+$ com $\mu_j^k = 0$ para $j \notin A(x^*)$, $\theta_j^k = 0$ para $j \notin A(x^*)$ tais que $x^k \rightarrow x^*$, $\delta_k \rightarrow 0$ e

$$\varepsilon^k := \lim_{k \rightarrow \infty} \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) = 0; \quad (3.15)$$

$$\nabla^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^m \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j \in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succeq -\delta_k \mathbb{I}; \quad (3.16)$$

Usando (3.15) e (3.16), obtemos

$$(-\nabla f(x^k) + \varepsilon^k, -\nabla^2 f(x^k) - \delta_k \mathbb{I}) \in K_2^{\mathcal{W}}(x^k), \quad \forall k.$$

Pela continuidade de $\nabla f(x)$, $\nabla^2 f(x)$ e usando a semi-continuidade exterior de $K_2^{\mathcal{W}}(x)$, segue que $(-\nabla f(x^*), -\nabla^2 f(x^*)) \in K_2^{\mathcal{W}}(x^*)$ e como consequência vale WSONC.

Vamos provar a outra implicação. Seja $(w, W) \in \limsup_{x \rightarrow x^*} K_2^{\mathcal{W}}(x)$. Devemos mostrar que $(w, W) \in K_2^{\mathcal{W}}(x^*)$. Pela definição de limite exterior, existem sequências $\{x^k\}$, $\{\lambda_i^k\}$, $\{\mu_j^k\}$ com $\mu_j^k = 0$ para $j \notin A(x^*)$ e $\{H^k\}$ tais que $x^k \rightarrow x^*$,

$$\left(\sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k), H^k \right) \rightarrow (w, W), \quad (3.17)$$

$$H^k \preceq \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla^2 g_j(x^k) \text{ sobre } C^{\mathcal{W}}(x^k, x^*). \quad (3.18)$$

Defina a seguinte função objetivo:

$$f(x) := -w^T(x - x^*) - \frac{1}{2}(x - x^*)^T W(x - x^*).$$

Mostraremos que para essa função, a condição AKKT2 vale em x^* . Temos que $\nabla f(x) = -w - W(x - x^*)$, $\nabla^2 f(x) = -W$. Assim, segue que

$$\begin{aligned} & \lim_{k \rightarrow \infty} \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) \\ &= \lim_{k \rightarrow \infty} \left(\sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) - w \right) - \lim_{k \rightarrow \infty} W(x^k - x^*) = 0, \end{aligned}$$

o que prova (3.8). Para provar a condição (3.9), vamos usar o Lema 3.33 com

$$P^k := \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla^2 g_j(x^k) - H^k + \frac{1}{k} \mathbb{I}, \quad (3.19)$$

e a_i como sendo as colunas da matriz $(\nabla h_i(x^k), \nabla g_j(x^k), i = 1, \dots, m, j \in A(x^*))$. Usando o Lema 3.33 existem sequências positivas $\{\theta^k\}$ e $\{\eta^k\}$ tais que

$$S^k := P^k + \sum_{i=1}^m \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j \in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succ 0. \quad (3.20)$$

Tome $\theta_j^k = 0$ para $j \notin A(x^*)$. Pelas relações (3.19) e (3.20) e $\nabla^2 f(x) = -W$, obtemos

$$\begin{aligned} \nabla^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^m \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j \in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \\ = -W + H^k + S^k - \frac{1}{k} \mathbb{I}. \end{aligned} \quad (3.21)$$

Usando a definição de norma, temos $-W + H^k = -(W - H^k) \succeq -|\alpha(W - H^k)| \mathbb{I}$, onde α denota o menor autovalor de $W - H^k$. Da relação (3.20), segue que

$$\begin{aligned} -W + H^k + S^k - \frac{1}{k} \mathbb{I} &\succeq -|\alpha(W - H^k)| \mathbb{I} + S^k - \frac{1}{k} \mathbb{I} \\ &\succeq -|\alpha(W - H^k)| \mathbb{I} - \frac{1}{k} \mathbb{I} = -\delta_k \mathbb{I} \end{aligned} \quad (3.22)$$

onde $\delta_k := |\alpha(W - H^k)| + 1/k$. Como $H^k \rightarrow W$, temos que $\delta_k \rightarrow 0$. Substituindo (3.22) em (3.21), obtemos

$$\nabla^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^m \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j \in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T + \delta_k \mathbb{I} \succeq 0$$

e portanto vale AKKT2 em x^* . Logo, usando a hipótese, temos que x^* satisfaz WSONC, ou seja,

$$(w, W) = (-\nabla f(x^*), -\nabla^2 f(x^*)) \in K_2^{\mathcal{W}}(x^*),$$

e o resultado segue. $\square$

Como AKKT2 é uma condição de otimalidade, o Teorema 3.34 garante que CCP2 é uma condição de qualificação (estrita), ou seja, $\text{AKKT2} + \text{CCP2} \Rightarrow \text{WSOnc}$.

Pela Proposição 3.29 e pelo Teorema 3.34, segue que $\text{MFCQ} + \text{WCR} \Rightarrow \text{CCP2}$.

Contra-Exemplo 3.35 (CCP2 não implica $\text{MFCQ} + \text{WCR}$). Considere em $\mathbb{R}$ o ponto $x^* = 0$ e as restrições $g_1(x) = x \leq 0$ e $g_2(x) = -x \leq 0$. Claramente MFCQ não vale em x^* e, conseqüentemente, $\text{MFCQ} + \text{WCR}$ não vale em x^* . Agora, como $\nabla g_1(x) = 1$, $\nabla^2 g_1(x) = 0$, $\nabla g_2(x) = -1$ e $\nabla^2 g_2(x) = 0$, segue que $C^{\mathcal{W}}(x, x^*) = \{0\}$. Desse modo, qualquer matriz $H \in \text{Sym}(n) = \mathbb{R}$ satisfaz

$$H \preceq \mu_1 \nabla^2 g_1(x) + \mu_2 \nabla^2 g_2(x) = 0 \text{ sobre } C^{\mathcal{W}}(x, x^*) \quad \forall \mu_1, \mu_2 \geq 0.$$

Logo, $K_2^{\mathcal{W}}(x) = \mathbb{R} \times \mathbb{R}$ para todo $x \in \mathbb{R}$ e, claramente, $K_2^{\mathcal{W}}(x)$ é semi-contínua exteriormente em x^* . Portanto vale CCP2.

Provou-se em [4] que RCRCQ implica CCP2. O contra-exemplo a seguir mostra que RCRCQ é estritamente mais forte que CCP2.

Contra-Exemplo 3.36 (CCP2 não implica RCRCQ). Considere em $\mathbb{R}^2$, o ponto $x^* = (0, 0)$ e o seguinte conjunto viável

$$\Omega = \{x \in \mathbb{R}^n : h_1(x) = x_1 = 0, g_1(x) = x_2 - x_1^2 \leq 0, g_2(x) = x_2^3 - x_1^2 \leq 0\}.$$

Logo,

$$\nabla h_1(x) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \nabla g_1(x) = \begin{pmatrix} -2x_1 \\ 1 \end{pmatrix}, \quad \nabla g_2(x) = \begin{pmatrix} -2x_1 \\ 3x_2^2 \end{pmatrix}.$$

Dessa forma, RCRCQ falha em x^* . Como $C^{\mathcal{W}}(x, x^*) = \{(0, 0)\}$, segue que $K_2^{\mathcal{W}}(x) = \mathbb{R} \times \mathbb{R}_+ \times \text{Sym}(2)$. Claramente, $K_2^{\mathcal{W}}(x)$ é semi-contínua exteriormente em x^* .

Outra condição de qualificação de segunda ordem conhecida na literatura, introduzida por Baccari e Trad em [17] é a seguinte:

Definição 3.37. Dizemos que a *condição de Baccari-Trad* vale em $x^* \in \Omega$ se MFCQ vale e o posto dos gradientes das restrições ativas é, no máximo, um a menos que o número total das restrições ativas.

Em [17] os autores provam que a condição de Baccari-Trad garante a validade da condição de otimalidade fraca de segunda ordem. Porém, em [38] (exemplo 3.4) mostrou-se que as condições de Baccari-Trad e CCP2 não estão relacionadas entre si. Em [38] (exemplo 3.5), os autores mostram que a condição de Baccari-Trad não é suficiente para garantir que um ponto limite de uma sequência AKKT2 cumpra a condição fraca de segunda ordem. Em outras palavras, a condição de Baccari-Trad não é uma condição de qualificação estrita para a condição sequencial AKKT2.

Podemos definir outra condição semelhante à CCP2 substituindo WSONC por SSONC. Sejam $x^* \in \Omega$ e $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}_+^p$ com $\mu_j = 0$, $j \notin A(x^*)$. Defina o seguinte cone

$$K_2^{\mathcal{S}}(x^*) := \bigcup_{(\lambda, \mu)} \left\{ \begin{array}{l} \left(\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x), H \right) \quad \text{tal que} \\ H \preceq \sum_{i=1}^m \lambda_i \nabla^2 h_i(x) + \sum_{j \in A(x^*)} \mu_j \nabla^2 g_j(x) \quad \text{sobre } C^{\mathcal{S}}(x^*, \mu) \end{array} \right\},$$

$C^{\mathcal{S}}(x^*, \mu)$ é o cone crítico forte.

Definição 3.38 (SCCP2, [4]). Dizemos que a *propriedade forte do cone contínuo de segunda ordem* (SCCP2) vale em $x^* \in \Omega$ se

$$\limsup_{x \rightarrow x^*} K_2^S(x) \subset K_2^S(x^*).$$

Observe que a condição de otimalidade forte de segunda ordem vale em x^* se, e somente se, $(-\nabla f(x^*), -\nabla^2 f(x^*)) \in K_2^S(x^*)$. Note também que $K_2^S(x^*) \subset K_2^W(x^*)$, pois $C^W(x^*) \subset C^S(x^*, \mu)$ para cada $\mu \geq 0$. Como consequência, a condição SCCP2 é mais forte que CCP2.

Como esperado, a condição SCCP2 é a condição mínima que garante que AKKT2 implica a condição SSONC.

Teorema 3.39 ([4], TEOREMA 4.16). *Para toda função objetivo f tal que AKKT2 implica a condição SSONC, então SCCP2 vale. Reciprocamente, se vale SCCP2, então AKKT2 implica a condição SSONC, para toda função objetivo f .*

A condição SCCP2 têm uma desvantagem por ser uma condição de qualificação muito forte, podendo falhar até mesmo em problemas bem comportados onde a condição de independência linear é satisfeita.

Exemplo 3.40 ([4], EXEMPLO 4.17). Considere em $\mathbb{R}^n$ ($n \geq 1$) a restrição de caixa simples $\Omega = \{x \in \mathbb{R}^n : x \geq 0\}$. Vamos mostrar que CCP2 vale em $x^* = 0$ mas SCCP2 falha. O conjunto Ω está definido pelas restrições de desigualdade $g_j(x) = -x_j \leq 0$, $j = 1, \dots, n$. Neste caso, temos $A(x^*) = \{1, \dots, n\}$, $\nabla g_j(x) = -e_j$ e $\nabla^2 g_j(x) = 0$. Segue que $C^W(x, x^*) = \{0\}$ e como consequência $K_2^W(x) = \mathbb{R}_+^n \times \text{Sym}(n)$. Logo, claramente $K_2^W(x)$ é semi-contínua exteriormente em x^* e portanto vale CCP2. Para mostrar que SCCP2 falha em x^* é suficiente que existam $\bar{\mu} \in \mathbb{R}_+^n$ e $H \in \text{Sym}(n)$ tal que $w^T H w > 0$ para algum $w \in C^S(x, \bar{\mu})$, pois nesse caso, $(-\bar{\mu}, H) \in K_2^W(x) = \mathbb{R}_+^n \times \text{Sym}(n)$ e $(-\bar{\mu}, H) \notin K_2^S(x^*)$.

Defina $e := \sum_{j=1}^n e_j$, $\bar{\mu} := e - e_1$ e $H := e_1^T H e_1 = \|\langle e_1, e_1 \rangle\|^2 > 0$. Dessa forma, segue que $(-\bar{\mu}, H) \in \limsup_{x \rightarrow x^*} K_2^W(x)$ com $(-\bar{\mu}, H) \notin K_2^S(x^*)$. Portanto não vale SCCP2.

Note que LICQ é satisfeita em x^* e, portanto, LICQ não implica SCCP2. Mais ainda, o próximo contra-exemplo mostra que essas condições são independentes.

Contra-Exemplo 3.41 ([4], EXEMPLO 4.18). Considere em $\mathbb{R}$ o ponto $x^* = 0$ e as restrições de desigualdade $g_1(x) = 1 - \exp(x) \leq 0$ e $g_2(x) = x \leq 0$. Logo temos $\nabla g_1(x) = -\exp(x)$, $\nabla g_2(x) = 1$, $\nabla^2 g_1(x) = -\exp(x)$ e $\nabla^2 g_2(x) = 0$. Assim, $\nabla g_1(x^*) = -1$ e $\nabla g_2(x^*) = 1$ e não vale LICQ em x^* . Por um cálculo simples, segue que $C^W(x, x^*) = C^S(x^*, \mu) = \{0\}$ para todo $\mu \in \mathbb{R}_+^2$ e, portanto, $K_2^W(x) = \mathbb{R} \times \text{Sym}(1) = K_2^S(x^*)$ para todo $x \in \mathbb{R}$, o que significa que vale SCCP2.

Resumindo, nesta seção apresentamos a condição sequencial de otimalidade de segunda ordem conhecida como AKKT2, bem como a condição de qualificação mínima associada a essa condição. A condição sequencial de segunda ordem AKKT2 é forte no sentido que implica em “WSONC ou não-CCP2”, onde CCP2 é estritamente mais fraca do que a condição RCRCQ. Além disso, muitos métodos numéricos para resolver problemas de otimização restrita geram sequências cujos pontos limites satisfazem a condição AKKT2.

4 Conclusões

Durante o desenvolvimento deste trabalho, concluímos que as condições sequenciais de otimalidade desempenham um papel importante em problemas de otimização restrita, sendo frequentemente utilizadas na análise de convergência de muitos métodos numéricos iterativos. Como foi apresentado neste trabalho, as condições sequenciais são condições necessárias para a otimalidade, ou seja, não dependem de nenhuma condição de qualificação e, portanto, estudar a força da condição sequencial passa a ser relevante. Desse modo, podemos dizer que a eficiência de um método está ligado à força da condição de otimalidade, que é garantido pelos pontos de acumulação das sequências geradas pelo método. A análise da força de uma condição sequencial de otimalidade pode ser feita de forma direta ou indireta. A forma direta consiste em relacionar as condições sequenciais por meio de implicações entre elas. Já a análise indireta corresponde encontrar condições de qualificação estrita para as condições sequenciais, ou seja, significa encontrar uma condição de qualificação para a qual a condição sequencial implica que todo minimizador local é um ponto estacionário. O interesse na análise indireta reside no fato de que as condições de qualificação são propriedades sobre as restrições do problema que não dependem da função objetivo. Fizemos uma análise da força de cada condição sequencial de otimalidade de forma direta e indireta, mostrando as condições de qualificação associadas a cada condição sequencial. Em particular, analisamos uma recente condição sequencial chamada PCAKKT [7] que unifica as condições sequenciais anteriores, no sentido de que é estritamente mais forte do que as condições sequenciais conhecidas atualmente. Entretanto, ainda não conhecemos um método prático com convergência a pontos PCAKKT sem nenhuma hipótese adicional.

Foi feito um estudo detalhado acerca das condições de qualificação de primeira e segunda ordens conhecidas, mostrando as relações entre elas, por meio de demonstrações, exemplos e contra-exemplos. Por fim, nossa principal contribuição neste trabalho foi mostrar o papel que as condições sequenciais desempenham em muitos algoritmos práticos para resolver problemas de otimização restrita.

Pesquisas futuras acerca das condições de qualificação associadas às condições sequenciais podem abordar problemas de otimização não linear com outras características, como por exemplo, com componentes não suaves ou restrições de cone. Estudos recentes foram feitos envolvendo condições sequenciais para otimização cônica em [65]. Por outro lado, muita pesquisa ainda pode ser feita para problemas de otimização multiobjetivo [27], programação semidefinida [12], cálculo variacional e controle ótimo [25], entre outros.

Referências

- [1] ABADIE, J. On the Kuhn-Tucker theorem. In: ABADIE, J. (Ed.). *Nonlinear Programming*. [S.l.]: John Wiley, 1967. p. 21–36.
- [2] ANDREANI, R. et al. A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal of Optimization*, 2016.
- [3] ANDREANI, R. et al. A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programming*, v. 135, p. 255–273, 2012.
- [4] ANDREANI, R. et al. A second-order sequential optimality condition associated to the convergence of optimization algorithms. *IMA Journal of Numerical Analysis*, v. 37, p. 1902–1929, 2017.
- [5] ANDREANI, R. et al. A sequential optimality condition related to the quasi-normality constraint qualification and its algorithmic consequences. *SIAM Journal of Optimization*, v. 29, n. 1, p. 743–766, 2019.
- [6] ANDREANI, R. et al. On second-order optimality conditions in nonlinear optimization. *Optimization Methods & Software*, 2016.
- [7] ANDREANI, R. et al. On the convergence of augmented lagrangian strategies for nonlinear programming. *IMA Journal of Numerical Analysis*, p. 1–31, 2021.
- [8] ANDREANI, R. et al. Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Optimization: A Journal of Mathematical Programming and Operations Research*, p. 1–25, 2018. ISSN 1526-5471.
- [9] ANDREANI, R. et al. Two new weak constraint qualifications and applications. *SIAM Journal of Optimization*, v. 22, n. 3, p. 1109–1135, 2012.
- [10] ANDREANI, R.; ECHAGÜE, C. E.; SCHUVERDT, M. L. Constant-rank condition and second-order constraint qualification. *Journal of Optimization Theory and Applications*, v. 146, p. 255–266, 2010.
- [11] ANDREANI, R.; HAESER, G.; MARTÍNEZ, J. M. On sequential optimality conditions for smooth constrained optimization. *Optimization: A Journal of Mathematical Programming and Operations Research*, v. 60, n. 5, p. 627–641, 2011.
- [12] ANDREANI, R.; HAESER, G.; VIANA, D. S. Optimality conditions and global convergence for nonlinear semidefinite programming. *Mathematical Programming*, 2018. ISSN 0025-5610.

-
- [13] ANDREANI, R.; MARTÍNEZ, J. M.; SCHUVERDT, M. L. On second-order optimality conditions for nonlinear programming. *Optimization: A Journal of Mathematical Programming and Operations Research*, v. 56, n. 5-6, p. 529–542, 2007. ISSN 0233-1934.
- [14] ANDREANI, R.; MARTÍNEZ, J. M.; SCHUVERDT, M. L. On the relation between constant positive linear dependence condition and quasinormality constraint qualification. *Journal of Optimization Theory and Applications*, v. 125, n. 2, p. 473–485, 2005.
- [15] ANDREANI, R.; MARTÍNEZ, J. M.; SVAITER, B. F. A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal of Optimization*, v. 20, n. 6, p. 3533–3554, 2010.
- [16] ARROW, K. J.; HURWITZ, L.; UZAWA, H. Constraint qualifications in maximization problems. *Naval Research Logistics Quarterly*, v. 8, p. 175–191, 1961.
- [17] BACCARI, A.; TRAD, A. On the classical necessary second-order optimality conditions in the presence of equality and inequality constraints. *SIAM Journal on Optimization, Society for Industrial and Applied Mathematics*, v. 15, n. 2, p. 394–408, 2005. ISSN 1052-6234.
- [18] BARTHOLOMEW-BIGGS, M. *Nonlinear Optimization with Engineering Applications*. [S.l.]: Springer, 2008. (Springer Optimization and Its Applications 19).
- [19] BAZARAA, M. S.; SHERALI, H. D.; SHETTY, C. M. *Nonlinear Programming: Theory and Algorithms*. 3. ed. [S.l.]: Wiley-Interscience, 2006.
- [20] BERTSEKAS, D.; NEDIĆ, A.; OZDAGLAR, A. E. *Convex Analysis and Optimization*. Belmont, Massachusetts: Athena Scientific, 2003.
- [21] BERTSEKAS, D. P. *Nonlinear Programming*. 2. ed. Belmont, Massachusetts: Athena Scientific, 1999.
- [22] BIRGIN, E.; MARTÍNEZ, J. M. *Practical Augmented Lagrangian Methods for Constrained Optimization*. [S.l.]: SIAM Publications, 2014.
- [23] BONNANS, J. F.; SHAPIRO, A. *Perturbation Analysis of Optimization Problems*. [S.l.]: Springer, 2000.
- [24] CHEN, L.; GOLDFARB, D. Interior point ℓ_2 -penalty methods for nonlinear programming with strong global convergence properties. *Mathematical Programming*, v. 108, p. 1–26, 2006.
- [25] CLARKE, F. *Functional Analysis, Calculus of Variations and Optimal Control*. London: Springer, 2013. (Graduate Texts in Mathematics).
- [26] CORNUEJOLS, G.; TUTUNCU, R. *Optimization Methods in Finance*. Pittsburgh: Cambridge University Press, 2006. (Mathematics, Finance and Risk).
- [27] DEB, K. *Multi-Objective Optimization using Evolutionary Algorithms*. New York, USA: Wiley, 2001.
- [28] DENNIS, J. E.; VICENTE, L. N. On the convergence theory of trustregion-based algorithms for equality-constrained optimization. *SIAM Journal of Optimization*, v. 7, n. 4, p. 927–950, 1997.

-
- [29] EUSTÁQUIO, R. G. *Condições de otimalidade e de qualificação para problemas de programação não linear*. Dissertação (Mestrado) — Universidade Federal do Paraná, Curitiba, 2007.
- [30] FINKE, G. (Ed.). *Operational Research and Networks*. [S.l.]: John Wiley, 2008. (Geographical Information Systems).
- [31] FLETCHER, R. *Practical methods of Optimization: Constrained Optimization*. New York: Wiley, 1981.
- [32] FOLLAND, G. B. *Real Analysis: modern techniques and their applications*. 2. ed. [S.l.]: Wiley-Interscience, 1999.
- [33] FUJIWARA, O.; HAN, S. P.; MANGASARIAN, O. L. Local duality of nonlinear programs. *SIAM Journal on Control and Optimization*, v. 22, n. 1, p. 162–169, 1984.
- [34] GILL, P. E.; KUNGURTSEV, V.; ROBINSON, D. P. A regularized SQP method with convergence to second-order optimal points. *Optimization Online*, 2013.
- [35] GOULD, F.; TOLLE, J. A necessary and sufficient qualification for constrained optimization. *SIAM Journal on Applied Mathematics*, v. 20, p. 164–172, 1969.
- [36] GUIGNARD, M. Generalized Kuhn-Tucker conditions for mathematical programming problems in Banach space. *SIAM Journal on Control and Optimization*, v. 7, p. 232–241, 1969.
- [37] HAESER, G. *Condições sequenciais de otimalidade*. Tese (Doutorado) — IMECC-UNICAMP, Departamento de Matemática Aplicada, 2009.
- [38] HAESER, G.; RAMOS, A. *Condições de Otimalidade e Algoritmos em Otimização não Linear*. São Carlos: Sociedade Brasileira de Matemática Aplicada e Computacional, 2016.
- [39] HAESER, G.; SCHUVERDT, M. L. On approximate kkt condition and its extension to continuous variational inequalities. *Journal of Optimization Theory and Applications*, v. 149, n. 3, p. 528–539, 2011.
- [40] HESTENES, M. R. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, v. 4, p. 303–320, 1969.
- [41] JANIN, R. Directional derivative of the marginal function in nonlinear programming. *Mathematical Programming Study*, v. 21, p. 110–126, 1984.
- [42] JOHN, F. Extremum problems with inequalities as subsidiary conditions. In: MOSER, J. (Ed.). *Studies and Essays, Courant Anniversary Volume*. [S.l.]: Wiley-Interscience, 1948. p. 187–204.
- [43] KARUSH, W. *Minima of Functions of Several Variables with Inequalities as Side Constraints*. Dissertação (Mestrado) — Department of Mathematics, University of Chicago, 1939.
- [44] KUHN, H.; TUCKER, A. Nonlinear programming. In: NEYMAN, J. (Ed.). *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. [S.l.]: University of California Press, 1951. p. 481–492.

-
- [45] KYPARISIS, J. On uniqueness of Kuhn-Tucker multipliers in nonlinear programming. *Mathematical Programming*, v. 32, p. 242–246, 1985.
- [46] LACHTERMACHER, G. *Pesquisa Operacional na tomada de decisões*. 5. ed. Rio de Janeiro: Grupo GEN, LTC, 2016.
- [47] LIMA, E. L. *Curso de análise vol. 2*. 11. ed. Rio de Janeiro: Coleção Projeto Euclides: IMPA, 2018.
- [48] LIMA, E. L. *Elementos de Topologia Geral*. Rio de Janeiro: SBM, 2009. (Textos Universitários).
- [49] MANGASARIAN, O. L.; FROMOVITZ, S. The Fritz John optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, v. 17, p. 37–47, 1967.
- [50] MARTÍNEZ, J. M.; PILOTTA, E. A. Inexact restoration algorithm for constrained optimization. *Journal of Optimization Theory and Applications*, v. 104, p. 135–163, 2000.
- [51] MARTÍNEZ, J. M.; SVAITER, B. F. A practical optimality condition without constraint qualifications for nonlinear programming. *Journal of Optimization Theory and Applications*, v. 118, n. 1, p. 117–133, 2003.
- [52] MINCHENKO, L.; STAKHOVSKI, S. On relaxed constant rank regularity condition in mathematical programming. *Optimization: A Journal of Mathematical Programming and Operations Research*, v. 60, n. 4, p. 429–440, 2011.
- [53] MINCHENKO, L.; STAKHOVSKI, S. Parametric nonlinear programming problems under the relaxed constant rank condition. *SIAM Journal on Optimization*, v. 21, n. 1, p. 314–332, 2011. ISSN 10526234.
- [54] MONTEIRO, R. N. F. *Modelo de programação não-linear para otimização do padrão de cultivo e retorno financeiro em áreas irrigadas*. Tese (Doutorado) — Faculdade de Ciências Agrônomicas-UNESP, Botucatu, 2014.
- [55] NOCEDAL, J.; WRIGHT, S. *Numerical Optimization*. 2. ed. New York: Springer, 2006.
- [56] POWELL, M. J. D. A method for nonlinear constraints in minimization problems. In: FLETCHER, R. (Ed.). *Optimization*. New York, NY: Academic Press, 1969. v. 19, p. 283–298.
- [57] QI, L.; WEI, Z. On the constant positive linear dependence condition and its applications to SQP methods. *SIAM Journal on Optimization*, v. 10, n. 4, p. 963–981, 2000.
- [58] RAMOS, A. *Tópicos em condições de otimalidade para otimização não linear*. Tese (Doutorado) — IME-USP, Departamento de Matemática Aplicada, 2016.
- [59] RIBEIRO, A. A.; KARAS, E. W. *Otimização Contínua: aspectos teóricos e computacionais*. São Paulo: Cengage Learning, 2014.

-
- [60] ROCKAFELLAR, R. T. Augmented lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control and Optimization*, v. 12, p. 268–285, 1974.
- [61] ROCKAFELLAR, R. T.; WETS, R. J. B. *Variational Analysis*. 1. ed. [S.l.]: Springer-Verlag Berlin Heidelberg, 2009. v. 317.
- [62] SAGAN, C. *Cosmos*. [S.l.]: Ballantine Books, 1985.
- [63] SCHUVERDT, M. L. *Métodos de Lagrangiano aumentado com convergência utilizando a condição de dependência linear positiva constante*. Tese (Doutorado) — IMECC-UNICAMP, Departamento de Matemática Aplicada, 2006.
- [64] SOUSA, J. A. F. *Otimização do peso de uma grande estrutura espacial em órbita baixa da terra com restrição na frequência fundamental de vibração*. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2002.
- [65] VIANA, D. S. *Condições de otimalidade para otimização cônica*. Tese (Doutorado) — IME-USP, Departamento de Matemática Aplicada, 2019.

Apêndice A - Pré-requisitos

As definições e resultados apresentados nas seções a seguir podem ser encontrados nas referências [20, 21, 32, 47, 48, 59].

A.1 Definições

Definição A.1. Um produto interno num espaço vetorial X é uma regra que faz corresponder a cada par de vetores $x, y \in X$ um número real, indicado por $\langle x, y \rangle$, de modo que, para quaisquer $x, x', y \in X$ e $\lambda \in \mathbb{R}$, se tenham:

- (i) $\langle x, y \rangle = \langle y, x \rangle$;
- (ii) $\langle x + x', y \rangle = \langle x, y \rangle + \langle x', y \rangle$;
- (iii) $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle = \langle x, \lambda y \rangle$;
- (iv) $x \neq 0 \Rightarrow \langle x, x \rangle > 0$.

Definição A.2. Uma norma num espaço vetorial X é uma função $\| \cdot \| : X \rightarrow \mathbb{R}_+$ que satisfaz as seguintes condições para quaisquer $x, y \in X$ e $\lambda \in \mathbb{R}$:

- (i) $\|x\| = 0$ se, e somente se, $x = 0$ (positiva definida);
- (ii) $\|\lambda \cdot x\| = |\lambda| \cdot \|x\|$;
- (iii) $\|x + y\| \leq \|x\| + \|y\|$ (desigualdade triangular).

Definição A.3. Em $\mathbb{R}^n$, definimos o *produto interno canônico* e a *norma euclidiana*, respectivamente por

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i \quad \text{e} \quad \|x\|_2 = \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Definição A.4. Um espaço vetorial normado é um par $(X, \| \cdot \|)$ onde X é um espaço vetorial e $\| \cdot \|$ é uma norma em X .

Definição A.5. Seja X um conjunto qualquer. Uma métrica definida sobre X é uma função $d : X \times X \rightarrow \mathbb{R}$ tal que para quaisquer $x, y, z \in X$, vale:

- (i) $d(x, y) \geq 0$ e $d(x, y) = 0$ se, e somente se, $x = y$;
- (ii) $d(x, y) = d(y, x)$;

(iii) $d(x, z) \leq d(x, y) + d(y, z)$.

Um espaço métrico é um par (M, d) formado por um conjunto M e uma métrica d em M .

Definição A.6. Dizemos que uma sequência $\{x_n\}$, num espaço métrico M , é uma *sequência de Cauchy* quando, para todo $\varepsilon > 0$ dado arbitrariamente, é possível obter $n_0 \in \mathbb{N}$ tal que $m, n > n_0 \Rightarrow d(x_m, x_n) < \varepsilon$.

Definição A.7. Dizemos que um espaço métrico M é *completo* quando toda sequência de Cauchy em M é convergente em M .

Definição A.8. Um espaço vetorial X com produto interno é um *Espaço de Hilbert* se X é completo com a norma proveniente do produto interno.

Definição A.9. Um espaço vetorial normado X é chamado *Espaço de Banach* quando for um espaço métrico completo, ou seja, se toda sequência de Cauchy em X é convergente em X .

O espaço $\mathbb{R}^n$ munido da norma euclidiana é um espaço de Banach.

A.2 Topologia do $\mathbb{R}^n$

Definição A.10. A bola aberta de centro no ponto $a \in \mathbb{R}^n$ e raio $r > 0$ é o conjunto

$$B(a; r) = \{x \in \mathbb{R}^n : \|x - a\| < r\}.$$

Analogamente, a bola fechada $B[a; r]$ e a esfera $S[a; r]$, ambas com centro a e raio r , são definidas por:

$$B[a; r] = \{x \in \mathbb{R}^n : \|x - a\| \leq r\} \quad \text{e} \quad S[a; r] = \{x \in \mathbb{R}^n : \|x - a\| = r\}.$$

Denotamos por S^{n-1} a esfera $S[0; 1] = \{x \in \mathbb{R}^n : \|x\| = 1\}$, denominada esfera unitária de $\mathbb{R}^n$.

Definição A.11. Seja $X \subset \mathbb{R}^n$. Um ponto $a \in X$ chama-se *um ponto interior a X* quando é centro de alguma bola aberta contida em X .

O conjunto formado pelos pontos interiores a X é denominado *interior de X* e denotado por $\text{int } X$.

Definição A.12. Dizemos que um conjunto $X \subset \mathbb{R}^n$ é *aberto* quando todos os seus pontos são interiores, isto é, quando $\text{int } X = X$.

Exemplo A.13. Toda bola aberta $B(a; r) \subset \mathbb{R}^n$ é um conjunto aberto. De fato, dado $x \in B(a; r)$, temos $\|x - a\| < r$, logo o número $\delta = r - \|x - a\|$ é positivo. Considere $y \in B(x; \delta)$, assim, $\|y - a\| \leq \|y - x\| + \|x - a\| < \delta + \|x - a\| = r \Rightarrow y \in B(a; r)$. Logo $B(x; \delta) \subset B(a; r)$ e, portanto $B(a; r)$ é aberto.

Definição A.14. Seja $X \subset \mathbb{R}^n$. Dizemos que um ponto $a \in \mathbb{R}^n$ é *um ponto da fronteira de X* se toda bola aberta de centro a contém pontos de X e pontos do complementar de X . Denotaremos por ∂X a fronteira de X .

Observação A.15. A fronteira da bola aberta $B(a; r)$ e da bola fechada $B[a; r]$ é a esfera $S[a; r]$, isto é, $\partial B(a; r) = \partial B[a; r] = S[a; r]$.

Definição A.16. Dizemos que um ponto $a \in \mathbb{R}^n$ é *aderente* a um conjunto $X \subset \mathbb{R}^n$ quando é limite de uma sequência de pontos desse conjunto.

O conjunto dos pontos aderentes a X chama-se o *fecho* de X e é denotado por $\overline{X}$.

Observação A.17. Todo ponto $a \in X$ pertence a $\overline{X}$, basta tomar a sequência $\{x^k\}$ com $x^k = a$ para todo $k \in \mathbb{N}$. Assim, $X \subset \overline{X}$.

Definição A.18. Dizemos que um conjunto $X \subset \mathbb{R}^n$ é *fechado* quando $X = \overline{X}$. Equivalentemente, X é fechado se toda sequência $\{x^k\} \subset X$ converge para um ponto de X .

Observação A.19. O fecho de um conjunto $X \subset \mathbb{R}^n$ pode ser definido como $\overline{X} = X \cup \partial X$. Assim, X é fechado se, e somente se, $\partial X \subset X$.

Definição A.20. Um conjunto $X \subset \mathbb{R}^n$ é *limitado* quando existe $M > 0$ tal que $\|x\| \leq M$ para todo $x \in X$.

Definição A.21. Dizemos que $X \subset \mathbb{R}^n$ é *compacto* quando é fechado e limitado.

Definição A.22. Seja $X \subset \mathbb{R}^n$. Dizemos que um ponto $a \in \mathbb{R}^n$ é um *ponto de acumulação* de X quando toda bola aberta de centro a contém algum ponto de X , diferente do ponto a .

Definição A.23. Sejam X e Y conjuntos não vazios em $\mathbb{R}^n$. Definimos a *distância entre* X e Y por

$$\text{dist}(X, Y) = \inf\{\|x - y\| : x \in X, y \in Y\}.$$

Um caso particular importante da distância entre dois conjuntos ocorre quando um deles se reduz a um ponto.

Definição A.24. Dados $x \in \mathbb{R}^n$ e um conjunto não vazio $Y \subset \mathbb{R}^n$, a distância entre o ponto x e o conjunto Y é dada por

$$\text{dist}(x, Y) = \inf\{\|x - y\| : y \in Y\}.$$

A.3 Análise convexa

Nesta seção, apresentamos alguns resultados importantes em análise convexa que são fundamentais no estudo de problemas de otimização. Com hipóteses de convexidade, as condições necessárias de otimalidade passam a ser suficientes. Em outras palavras, todo ponto estacionário é uma solução do problema.

Definição A.25. Dados $x, y \in \mathbb{R}^n$, o *segmento de reta* de extremos x e y é o conjunto

$$[x, y] = \{(1 - t)x + ty, 0 \leq t \leq 1\}.$$

Definição A.26. Dizemos que um conjunto $\Omega \subset \mathbb{R}^n$ é *convexo* quando dados $x, y \in \Omega$, tivermos $[x, y] \subset \Omega$.

Lema A.27 ([59], p. 52). *Seja $\Omega \subset \mathbb{R}^n$ um conjunto não vazio, convexo e fechado. Dado $z \in \mathbb{R}^n$, existe um único $\bar{z} \in \Omega$ tal que, para qualquer $x \in \Omega$,*

$$\|z - \bar{z}\|_2 \leq \|z - x\|_2.$$

$\bar{z}$ é a projeção do ponto z sobre o conjunto Ω , denotado por $\bar{z} = P_\Omega(z)$.

Teorema A.28. *Sejam $\Omega \subset \mathbb{R}^n$ um conjunto não vazio, convexo e fechado, $z \in \mathbb{R}^n$ e $\bar{z} = P_\Omega(z)$, então para todo $x \in \Omega$,*

$$\langle z - \bar{z}, x - \bar{z} \rangle \leq 0.$$

Demonstração. Considere $x \in \Omega$ e $t \in (0, 1]$. Pela convexidade de Ω , temos $(1-t)\bar{z} + tx \in \Omega$. Portanto,

$$\|z - \bar{z}\| = \|z - (1-t)\bar{z} - tx\| = \|z - \bar{z} + t(\bar{z} - x)\|.$$

Assim,

$$\|z - \bar{z}\|^2 \leq \|z - \bar{z} + t(\bar{z} - x)\|^2 = \|z - \bar{z}\|^2 + 2t\langle z - \bar{z}, \bar{z} - x \rangle + t^2\|\bar{z} - x\|^2.$$

Como $t > 0$, temos $2\langle z - \bar{z}, x - \bar{z} \rangle \leq t\|\bar{z} - x\|^2$. Tomando o limite quando $t \rightarrow 0$, obtemos $\langle z - \bar{z}, x - \bar{z} \rangle \leq 0$. $\square$

Teorema A.29. *Sejam $\Omega \subset \mathbb{R}^n$ um conjunto não vazio, convexo e fechado, $z \in \mathbb{R}^n$. Se $\bar{z} \in \Omega$ satisfaz*

$$\langle z - \bar{z}, x - \bar{z} \rangle \leq 0,$$

para todo $x \in \Omega$, então $\bar{z} = P_\Omega(z)$.

Demonstração. Dado $x \in \Omega$, temos

$$\begin{aligned} \|z - \bar{z}\|^2 - \|z - x\|^2 &= -2\langle z, \bar{z} \rangle + \langle \bar{z}, \bar{z} \rangle + 2\langle z, x \rangle - \langle x, x \rangle \pm \langle x, \bar{z} \rangle \\ &= \langle x - \bar{z}, 2(z - \bar{z}) - (x - \bar{z}) \rangle \\ &= 2\langle x - \bar{z}, z - \bar{z} \rangle - \langle x - \bar{z}, x - \bar{z} \rangle \\ &= 2\langle x - \bar{z}, z - \bar{z} \rangle - \|x - \bar{z}\|^2 \leq 0. \end{aligned}$$

Logo, $\bar{z} = P_\Omega(z)$. $\square$

Teorema A.30. *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável e $C \subset \mathbb{R}^n$ um conjunto convexo e fechado. Se $x \in C$ é minimizador de f em C , então*

$$P_C(x - \alpha \nabla f(x)) = x, \text{ para todo } \alpha \geq 0.$$

Definição A.31. *Seja $\Omega \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é convexa em Ω quando, para todos $x, y \in \Omega$ e $0 \leq t \leq 1$,*

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y).$$

Definição A.32. *Dada uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$, definimos o epígrafo de f por*

$$E_f := \{(x, c) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq c\}.$$

Teorema A.33. *Seja $C \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : C \rightarrow \mathbb{R}$ é convexa em C se, e somente se, o epígrafo de f é um conjunto convexo em $\mathbb{R}^{n+1}$.*

Demonstração. ($\Rightarrow$) Sejam $(x, c_1) \in E_f$ e $(y, c_2) \in E_f$. Como $f(x) \leq c_1$ e $f(y) \leq c_2$, pela convexidade de f , para todo $t \in [0, 1]$ tem-se

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) \leq (1-t)c_1 + tc_2.$$

Logo,

$$(1-t)(x, c_1) + t(y, c_2) = ((1-t)x + ty, (1-t)c_1 + tc_2) \in E_f.$$

Portanto E_f é convexo.

($\Leftarrow$) Agora suponha E_f convexo e sejam $x, y \in C$ e $t \in [0, 1]$. Logo, $(x, f(x)) \in E_f$ e $(y, f(y)) \in E_f$. Como E_f é convexo, temos

$$(1-t)(x, f(x)) + t(y, f(y)) = ((1-t)x + ty, (1-t)f(x) + tf(y)) \in E_f.$$

Pela definição de E_f , segue que $f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$. Portanto f é convexa. $\square$

Teorema A.34 ([59], p. 57). *Sejam $\Omega \subset \mathbb{R}^n$ convexo e $f : \Omega \rightarrow \mathbb{R}$ uma função convexa. Se $x^* \in \Omega$ é minimizador local de f , então x^* é minimizador global de f .*

Demonstração. Seja $x^* \in \Omega$ um minimizador local de f , ou seja, existe $\delta > 0$ tal que $f(x^*) \leq f(x)$ para todo $x \in \Omega \cap B(x^*; \delta)$. Tome $y \in \Omega$, $y \notin B(x^*; \delta)$ e $0 < t < \delta/\|y - x^*\|$. Logo o ponto $x = (1-t)x^* + ty$ satisfaz $\|x - x^*\| = t\|y - x^*\| < \delta$ e assim, $x \in B(x^*; \delta)$. Pela convexidade de f , temos

$$f((1-t)x^* + ty) \leq (1-t)f(x^*) + tf(y).$$

Segue que

$$f(x^*) \leq f((1-t)x^* + ty) \leq (1-t)f(x^*) + tf(y)$$

Logo $tf(x^*) \leq tf(y)$ e, como $t > 0$, segue que $f(x^*) \leq f(y)$. Portanto x^* é minimizador global de f . $\square$

Teorema A.35. *Sejam $\Omega \subset \mathbb{R}^n$ um conjunto convexo e $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável em Ω . Então f é convexa em Ω se, e somente se,*

$$f(y) \geq f(x) + \nabla f(x)^T(y - x), \quad \forall x, y \in \Omega.$$

Demonstração. ($\Rightarrow$) Suponha f convexa. Sejam $x, y \in \Omega$ e $0 < t \leq 1$. Como Ω é convexo, temos que $x + t(y - x) \in \Omega$. Pela diferenciabilidade da função f , temos

$$\lim_{t \rightarrow 0^+} \frac{f(x + t(y - x)) - f(x)}{t} = \nabla f(x)^T(y - x). \quad (\text{A.1})$$

Usando a convexidade de f , temos

$$f((1-t)x + t(y-x)) = f(x + t(y-x)) \leq (1-t)f(x) + tf(y). \quad (\text{A.2})$$

Dividindo a expressão (A.2) por t e tomando o limite quando $t \rightarrow 0^+$, obtemos

$$\nabla f(x)^T(y - x) = \lim_{t \rightarrow 0^+} \frac{f(x + t(y-x)) - f(x)}{t} \leq f(y) - f(x).$$

($\Leftarrow$) Tome $z = (1-t)x + ty$. Assim, pela hipótese temos

$$f(x) \geq f(z) + \nabla f(z)^T(x - z) \quad \text{e} \quad f(y) \geq f(z) + \nabla f(z)^T(y - z).$$

Multiplicando a primeira desigualdade por $(1-t)$, a segunda por t e somando, obtemos

$$(1-t)f(x) + tf(y) \geq f((1-t)x + ty) + \nabla f(z)^T((1-t)x + ty - z) = f((1-t)x + ty).$$

□

Corolário A.36. *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa, diferenciável e $C \subset \mathbb{R}^n$ convexo. Se $\langle \nabla f(x), y - x \rangle \geq 0$, para todo $y \in C$, então x é um minimizador global de f em C . Em particular, todo ponto estacionário é minimizador global.*

Teorema A.37. *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa diferenciável e $C \subset \mathbb{R}^n$ convexo e fechado. Se*

$$P_C(x - \nabla f(x)) = x,$$

então $x \in C$ é minimizador global de f em C .

Demonstração. Como $P_C(x - \nabla f(x)) = x$, usando o Teorema A.28, temos

$$\langle -\nabla f(x), y - x \rangle = \langle x - \nabla f(x) - x, y - x \rangle \leq 0.$$

Portanto, pelo Corolário A.36, temos que x é minimizador global de f em C . □

Teorema A.38 ([59], p. 60). *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função de classe $\mathcal{C}^2$ e $C \subset \mathbb{R}^n$ um conjunto convexo.*

(i) *Se $\nabla^2 f(x) \geq 0$, para todo $x \in C$, então f é convexa em C .*

(ii) *Se f é convexa em C e $\text{int } C \neq \emptyset$, então $\nabla^2 f(x) \geq 0$, para todo $x \in C$.*

Outra ferramenta útil em análise convexa é a noção de cone normal.

Definição A.39. Sejam $\Omega \subset \mathbb{R}^n$ um conjunto convexo e $x^* \in \Omega$. O *cone normal* (cone de direções normais) no ponto x^* em relação ao conjunto Ω é dado por

$$N_\Omega(x^*) = \{d \in \mathbb{R}^n : \langle d, x - x^* \rangle \leq 0, \forall x \in \Omega\}.$$

Proposição A.40. *Seja $\Omega \subset \mathbb{R}^n$ um conjunto convexo fechado e $x \in \Omega$, então*

$$d \in N_\Omega(x) \Leftrightarrow P_\Omega(x + d) = x.$$

Demonstração. ($\Rightarrow$) Dado $d \in N_\Omega(x)$, temos que

$$\langle d, y - x \rangle \leq 0 \Leftrightarrow \langle (x + d) - x, y - x \rangle \leq 0, \text{ para todo } y \in \Omega.$$

Pelo Teorema A.29, segue que $x = P_\Omega(x + d)$.

($\Leftarrow$) Como $P_\Omega(x + d) = x$, então pelo Teorema A.28, temos

$$\langle d, y - x \rangle = \langle (x + d) - x, y - x \rangle \leq 0, \text{ para todo } y \in \Omega.$$

Portanto $d \in N_\Omega(x)$. □

Índice Remissivo

Condição

AKKT de segunda ordem, 81
 aproximadamente KKT (AKKT), 39
 constant rank of the subspace component (CRSC), 34
 de Abadie, 25
 de dependência linear positiva constante (CPLD), 33
 de Error Bound, 36
 de Fritz John, 23
 de independência linear (LICQ), 26
 de Mangasarian-Fromovitz (MFCQ), 27
 de Mangasarian-Fromovitz estrita (SMFCQ), 30
 de posto constante (CRCQ), 31
 de posto constante fraco (WCR), 76
 de Slater, 29
 AGP-regular, 62
 CAKKT-regular, 63
 complementary approximate KKT (CAKKT), 52
 constant positive generator (CPG), 36
 de Baccari-Trad, 87
 de otimalidade, 12
 de qualificação, 25
 gradiente projetado aproximado (AGP), 42
 gradiente projetado aproximado convexo (C-AGP), 49
 gradiente projetado aproximado linear (L-AGP), 52
 necessária forte de segunda ordem (SSONC), 71
 necessária fraca de segunda ordem (WSONC), 71
 PAKKT-regular, 65
 PCAKKT-regular, 67
 positive approximate KKT (PAKKT), 55

positive complementary approximate KKT (PCAKKT), 57
 SAKKT-regular, 64
 sequencial de otimalidade, 38
 strong approximate KKT (SAKKT), 47

Condição relaxada

de dependência linear positiva constante (RCPLD), 34
 de posto constante (RCRCQ), 33

Cone

linearizado, 17
 normal limite, 19
 normal regular, 19
 positivamente gerado, 35
 tangente, 15

Lema

de Carathéodory, 14
 de Farkas, 16

Minimizador

global, 12
 local, 12

Multifunção

domínio, 18
 gráfico, 18
 imagem, 18
 semi-contínua exteriormente, 19
 semi-contínua interiormente, 19

Propriedade

da continuidade do cone (CCP), 61
 do cone contínuo de segunda ordem (CCP2), 85
 forte do cone contínuo de segunda ordem (SCCP2), 88

Teorema

de Karush-Kuhn-Tucker, 21