Notas em Matemática Aplicada e-ISSN 2236-5915

**Volume 83, 2016**



---

**Sociedade Brasileira de Matemática Aplicada e Computacional**

**2016**

A Sociedade Brasileira de Matemática Aplicada e Computacional - SBMAC publica, desde as primeiras edições do evento, monografias dos cursos que são ministrados nos CNMAC.

Para a comemoração dos 25 anos da SBMAC, que ocorreu durante o XXVI CNMAC em 2003, foi criada a série **Notas em Matemática Aplicada** para publicar as monografias dos minicursos ministrados nos CNMAC, o que permaneceu até o XXXIII CNMAC em 2010.

A partir de 2011, a série passa a publicar, também, livros nas áreas de interesse da SBMAC. Os autores que submeterem textos à série Notas em Matemática Aplicada devem estar cientes de que poderão ser convidados a ministrarem minicursos nos eventos patrocinados pela SBMAC, em especial nos CNMAC, sobre assunto a que se refere o texto.

O livro deve ser preparado em **Latex (compatível com o Miktex versão 2.9), as figuras em eps** e deve ter entre **80 e 150 páginas**. O texto deve ser redigido de forma clara, acompanhado de uma excelente revisão bibliográfica e de **exercícios de verificação de aprendizagem** ao final de cada capítulo.

Veja todos os títulos publicados nesta série na página
http://www.sbmac.org.br/p_notas.php

**Sociedade Brasileira de Matemática Aplicada e Computacional**

**2016**

# Condições de Otimalidade e Algoritmos em Otimização não Linear

Gabriel Haeser
ghaeser@ime.usp.br

Departamento de Matemática Aplicada
Instituto de Matemática e Estatística
Universidade de São Paulo, São Paulo-SP

Alberto Ramos
albertoramos@ufpr.br

Departamento de Matemática
Universidade Federal do Paraná, Curitiba-PR

**Sociedade Brasileira de Matemática Aplicada e Computacional**

São Carlos - SP, Brasil
2016

Coordenação Editorial: Igor Leite Freire

Coordenação Editorial da Série: Alexandre L. Madureira

Editora: SBMAC

Capa: Matheus Botossi Trindade

Patrocínio: SBMAC



**Catalogação elaborada pela Biblioteca do IBILCE/UNESP**
**Bibliotecária: Maria Luiza Fernandes Jardim Froner**

Haeser, Gabriel
Condições de Otimalidade e Algoritmos em Otimização não Linear - São Carlos, SP : SBMAC, 2016, 85 p., 21.5 cm - (Notas em Matemática Aplicada; v. 83)

e-ISBN 978-85-8215-075-7

1. Otimização não Linear 2. Condições de otimalidade 3. Condições de qualificação 4. Algoritmos
I. Ramos, Alberto II. Título. III. Série

CDD - 51

# Conteúdo


**Prefácio** **1**

**1 Introdução** **3**

1.1 Definições básicas . . . . . 3
1.2 Condição geométrica de otimalidade . . . . . 4
1.3 Exercícios . . . . . 7

**2 Condições de otimalidade de primeira ordem** **11**

2.1 Condições de qualificação . . . . . 14
2.2 Condições sequenciais de primeira ordem . . . . . 22
2.2.1 Condição Aproximadamente KKT . . . . . 23
2.2.2 Condição Gradiente Projetado Aproximado (AGP) . . . . . 27
2.3 CQs associadas às condições sequenciais . . . . . 31
2.3.1 Resultados básicos de otimização e análise variacional. . . . . 34
2.3.2 CQ associada a AKKT: Propriedade de Continuidade do Cone 36
2.3.3 CQ associada a AGP: AGP-regular . . . . . 41
2.4 Exercícios . . . . . 45

**3 Condições de otimalidade de segunda ordem** **49**

3.1 Condições de qualificação de segunda ordem . . . . . 51
3.2 Condições sequenciais de segunda ordem . . . . . 61
3.3 Condição de qualificação mínima e algoritmos . . . . . 64
3.4 Exercícios . . . . . 77

# Prefácio

Neste trabalho estamos interessados em identificar propriedades de primeira e segunda ordem que são satisfeitas por um minimizador local de um problema geral de otimização não linear. Nosso interesse principal é encontrar condições que possam ser verificadas por algoritmos práticos. Definimos condições de otimalidade de primeira e segunda ordem mais fortes que as usuais, impondo condições menos restritivas sobre o problema, e como consequência mostramos que diversas classes de algoritmos de primeira e segunda ordem têm convergência global para pontos estacionários sob hipóteses mais fracas. Este livro é fortemente baseado nas teses de livre docência e de doutorado do primeiro e segundo autor, respectivamente, e de um mini-curso realizado em Outubro de 2014 na Universidad de Santiago de Compostela (as aulas podem ser acessadas em `www.ime.usp.br/~ghaeser`). As ideias aqui apresentadas são fruto de diversos trabalhos publicados em parceria com Roberto Andreani, Roger Behling, José Mario Martínez, María Laura Schuverdt e Paulo J.S. Silva.

**Palavras chave:** otimização não linear; condições de otimalidade; condições de qualificação; algoritmos.

São Paulo, 31 de janeiro de 2016.

Gabriel Haeser e Alberto Ramos

# Capítulo 1

# Introdução

## 1.1 Definições básicas

Considere o problema geral de otimização não linear

$$\begin{array}{ll} \text{Minimizar} & f(x), \\ \text{Sujeito a} & h_i(x) = 0, i = 1, \ldots, m, \\ & g_j(x) \leq 0, j = 1, \ldots, p. \end{array} \tag{1.1.1}$$

onde $f$, $h_i$, $g_j : \mathbb{R}^n \to \mathbb{R}$ são funções continuamente diferenciáveis.

Muitos problemas científicos e tecnológicos precisam da solução de problemas desta forma, [24, 30, 69, 75]. Tais problemas vêm por exemplo da engenharia, física, química, pesquisa operacional, finanças, tecnologia espacial, entre outros.

Dado um problema prático da forma (1.1.1), a *melhor* solução para esse problema pode ser encontrada usando diferentes esquemas e métodos numéricos. Assim, a resolução de (1.1.1) requer o estudo de condições de otimalidade apropriadas, de uma escolha do método algorítmico adequado, de uma análise da teoria de convergência de algoritmos, e de experimentos numéricos com problemas típicos e problemas da vida real. O objetivo deste trabalho é fornecer um estudo detalhado das condições de otimalidade apropriadas para a análise de convergência de algoritmos.

Para o problema (1.1.1), temos que

- $x$ é o vetor de variáveis desconhecidas, a serem determinadas.
- $f$ é a função objetivo. Isto é, uma função que depende de $x$ e que desejamos minimizar.
- $h_i, i = 1, \ldots, m$ e $g_j, j = 1, \ldots, p$ são as funções de restrição. Elas representam as restrições que devem ser satisfeitas pelo vetor de variáveis $x$. As restrições de igualdades estão representadas por $h_i$ e as as restrições de desigualdades por $g_j$.

Chamamos de conjunto viável ao conjunto de pontos que satisfazem todas as restrições do problema de otimização (1.1.1). Ou seja, o conjunto viável é dado por:

$$\Omega := \{x \in \mathbb{R}^n \mid h_i(x) = 0, i = 1, \ldots, m; \;\; g_j(x) \leq 0, j = 1, \ldots, p\}.$$

Quando a função objetivo e as funções de restrição são funções afins, o problema de minimização (1.1.1) é chamado de *problema de otimização linear*. Se alguma das funções é não linear, o problema é chamado *problema de otimização não linear*.

Suavidade da função objetivo e das funções de restrição é uma importante hipótese para caracterizar as soluções. A suavidade garante que a função objetivo e as restrições se comportam de uma maneira razoável e previsível e como consequência permite aos algoritmos fazer boas escolhas para direções de busca.

Para $x \in \Omega$, denotamos por $A(x) = \{j \in \{1, \ldots, p\} \mid g_j(x) = 0\}$, o conjunto de índices de restrições de desigualdade ativas em $x$.

A seguir, definiremos as diferentes noções de solução para (1.1.1).

**Definição 1.1.** *Soluções para o problema geral de otimização não linear.*

- *Um vetor* $x^* \in \Omega$ *é um minimizador global se* $f(x^*) \leq f(x)$ *para todo* $x \in \Omega$.
- *Um vetor* $x^* \in \Omega$ *é um minimizador local se existe uma vizinhança* $\mathcal{B}$ *de* $x^*$ *tal que* $f(x^*) \leq f(x)$ *para todo* $x \in \Omega \cap \mathcal{B}$.
- *Um vetor* $x^* \in \Omega$ *é um minimizador local estrito se* $f(x^*) < f(x)$ *para todo* $x \in \Omega \cap \mathcal{B}$.

Uma *condição de otimalidade* é qualquer condição satisfeita por uma solução. Neste livro, estamos interessados em estudar as propriedades satisfeitas por um minimizador local do problema. Por exemplo, tanto a mera viabilidade quanto a própria otimalidade local são condições de otimalidade. Entretanto, encontrar um ponto que cumpre apenas a primeira condição não auxilia na tarefa de encontrar uma solução, já que existem muitos pontos viáveis que não são uma solução do problema, enquanto a segunda condição é muito difícil de ser verificada na prática.

Estamos interessados em estudar propriedades analíticas de uma solução local do problema, com o intuito de que tais propriedades possam ser exploradas na prática para guiar um processo iterativo que busca por uma solução do problema, bem como para fornecer um critério de parada. Em outras palavras, nosso interesse é o estudo de condições de otimalidade que possam ser usadas como critério de parada de algoritmos práticos.

Neste livro, estamos interessados em desenvolver condições de otimalidade fortes, no sentido que existam poucas soluções que não cumprem a condição; que sejam de fácil verificação prática e, além disso, que sejamos capazes de desenvolver algoritmos para encontrar pontos que cumprem a condição de otimalidade. Estes serão os candidatos a solução do problema.

## 1.2 Condição geométrica de otimalidade

Supondo que $x^*$ é uma solução (no sentido de minimizador local), uma primeira condição de otimalidade útil, chamada *condição geométrica*, é obtida analisando o ângulo de $\nabla f(x^*)$ com direções que apontam aproximadamente para o interior do conjunto viável. Para formalizar melhor esta ideia, considere a seguinte definição.

**Definição 1.2** (Cone tangente)**.** *Dado* $x \in \Omega$*, denotamos por* $\mathcal{T}_\Omega(x)$ *o cone tangente a* $\Omega$ *em* $x$*, dado por*

$$\mathcal{T}_\Omega(x) = \{0\} \cup \left\{ d \in \mathbb{R}^n \mid \exists \ \ \{x^k\} \in \Omega, \quad x^k \to x, \quad \frac{x^k - x}{\|x^k - x\|} \to \frac{d}{\|d\|} \right\}.$$

O cone tangente $\mathcal{T}_\Omega(x)$ é um cone fechado não-vazio. Este cone nos permite analisar as possíveis direções geradas por todas as sequências de pontos viáveis possíveis que convergem a $x^*$. De fato, seja $\{x^k\}$ é uma sequência de pontos viáveis que converge a $x^*$. Considere, para cada $k \in \mathbb{N}$, a direção $d^k := (x^k - x^*)/\|x^k - x^*\|$. Como $d^k$ é um vetor unitário, podemos tomar uma subsequência convergente de $\{d^k\}$. Ainda mais, podemos supor, sem perda de geralidade, que $d^k$ converge para certa direção $d$. Da construção, temos que $d \in \mathcal{T}_\Omega(x)$.

**Lema 1.1.** *Seja $x^* \in \Omega$ um minimizador local do problema (1.1.1). Então para todo $d \in \mathcal{T}_\Omega(x^*)$, $\nabla f(x^*)^{\mathtt{T}} d \geq 0$.*

*Demonstração.* Seja $d \in \mathcal{T}_\Omega(x^*)$. Da definição de cone tangente, existem sequências $\{t_k\} \in \mathbb{R}_+$ e $\{d^k\}$ tais que $t_k \to 0$, $d^k \to d$ e $x^* + t_k d^k \in \Omega$ para todo $k \in \mathbb{N}$. De fato, basta definir $t_k := \|d\|^{-1}\|x^k - x^*\|$ e $d^k := \|d\|\|x^k - x^*\|^{-1}(x^k - x^*)$. Como $x^* + t_k d^k \to x^*$ e já que $x^*$ é um minimizador local, temos que $f(x^* + t_k d^k) \geq f(x^*)$ para $k$ suficientemente grande. Usando a expansão de Taylor,

$$f(x^*) + t_k \nabla f(x^*)^{\mathtt{T}} d^k + o(t_k) = f(x^* + t_k d^k) \geq f(x^*) \tag{1.2.2}$$

e como consequência

$$t_k \nabla f(x^*)^{\mathtt{T}} d^k + o(t_k) \geq 0, \tag{1.2.3}$$

para $k$ suficientemente grande, onde $o(t_k)/t_k \to 0$. Dividindo a expressão (1.2.3) por $t_k$ e tomando limite concluímos que $\nabla f(x^*)^{\mathtt{T}} d \geq 0$ para todo $d \in T_\Omega(x^*)$. $\square$

Uma definição importante é a seguinte:

**Definição 1.3** (Cone polar)**.** *Seja $\mathcal{C} \subset \mathbb{R}^n$ um cone. Denotamos por $\mathcal{C}^\circ$ o cone polar de $\mathcal{C}$, dado por*

$$\mathcal{C}^\circ = \left\{ v \in \mathbb{R}^n \mid v^{\mathtt{T}} d \leq 0, \quad \forall d \in \mathcal{C} \right\}.$$

Algumas das propriedades desta operação estão descritas nos exercícios do final deste capítulo. Usando a definição de cone polar, o lema pode ser reformulado como:

**Teorema 1.1** (Condição de otimalidade geométrica)**.** *Se $x^* \in \Omega$ é uma solução local do problema* (1.1.1)*, então $-\nabla f(x^*) \in \mathcal{T}_\Omega(x^*)^\circ$.*

Observe que a verificação da condição geométrica na prática pode ser muito difícil. O principal motivo é que o cone tangente é um objeto geométrico, difícil de ser representado no computador (e mais ainda, o seu polar), que depende do conjunto viável $\Omega$, e não das funções $h_i, g_j$ que o definem. De fato, note que um mesmo conjunto $\Omega$ pode possuir descrições analíticas distintas.

Assim, condições de otimalidade que dependem explicitamente dos dados iniciais do problema têm preferência (do ponto de vista prático) sobre aquelas condições que dependem implicitamente dos mesmos dados. O motivo é que, na maioria dos problemas de otimização, o conjunto viável está implicitamente descrito pelas funções de restrição. Assim, só temos acesso direito às funções de restrição e não ao conjunto viável. Note que a condição geométrica depende implicitamente dos dados iniciais, através do cone tangente. Condições que dependem explicitamente da presença de multiplicadores de Lagrange, dizemos que são condições na forma dual. Sob esse critério, a condição geométrica é uma condição na forma primal.

Uma primeira condição analítica (na forma dual) aparece nos trabalhos Kuhn-Tucker [56] e Fritz John [54], onde temos o seguinte resultado:

**Teorema 1.2** (Condição de Fritz John)**.** *Se $x^*$ é uma solução local, então existe $0 \neq (\lambda_0, \lambda, \mu) \in \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^p_+$ tal que*

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

**Prova:** Veja a observação após a prova do Teorema 2.7. □

Os vetores $\lambda_0$, $\lambda$ e $\mu$ são chamados de multiplicadores de Fritz-John. As restrições $\mu_j g_j(x^*) = 0, \forall j$ são chamadas de restrições de complementaridade. Quando $g_j(x^*) < 0$, temos que obrigatoriamente $\mu_j$ deve ser zero. Assim a condição depende apenas das restrições de desigualdade ativas em $x^*$. De fato, as restrições inativas não desempenham nenhum papel na análise local da solução, já que se $g_j(x^*) < 0$, então a restrição $g_j(x) \leq 0$ não é violada em toda uma vizinhança de $x^*$, o que faz com que $x^*$ continue sendo uma solução local do problema, mesmo se esta restrição é removida. Observe ainda que a condição geométrica também não depende das restrições inativas, já que o cone tangente $\mathcal{T}_\Omega(x)$ permanece inalterado independente se incluímos ou não as restrições inativas no problema. Muitas vezes, ao longo do texto, a condição de complementaridade estará implícita ao escrevermos somatórios do tipo $\sum_{j=1}^{p} \mu_j \nabla g_j(x^*)$ da forma $\sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*)$.

Fritz John desenvolveu a condição do Teorema 1.2 estudando conjuntos convexos. Um primeiro problema formulado por ele é o de encontrar a esfera de menor raio que contém um conjunto limitado $S \subset \mathbb{R}^m$. Este problema foi formulado como

$$\begin{array}{ll} \text{Minimizar} & x_{m+1}, \\ \text{Sujeito a} & \sum_{i=1}^{m} (x_i - y_i)^2 \leq x_{m+1}, \forall y \in S. \end{array}$$

Utilizando a condição de Fritz John e uma formulação análoga, em [54], F. John mostrou que todo conjunto convexo compacto $S \subset \mathbb{R}^n$ de interior não vazio, possui um maior elipsóide em seu interior (elipsóide de John), além disso, existe um elipsóide "semelhante" ao elipsóide de John, com razão de semelhança menor ou igual a $n$ e que contém $S$.

Para os problemas estudados por Fritz John, a condição que ele desenvolveu era uma boa ferramenta, entretanto, a condição de Fritz John pode valer com $\lambda_0 = 0$, e então teríamos uma condição de otimalidade pouco útil, pois ela não dependeria da função objetivo. No caso em que $\lambda_0 = 0$, a condição de Fritz John fala mais sobre uma falha na descrição do conjunto viável (com restrições redundantes), do que sobre otimalidade do ponto. Sendo assim, Kuhn e Tucker estudaram propriedades sobre as funções que descrevem o conjunto viável que garantem a condição de Fritz John com $\lambda_0 \neq 0$. Tais condições são chamadas qualificações de restrição, em inglês (*constraint qualifications*), mas em português utilizamos o termo *condições de qualificação.* A condição de Fritz-John com $\lambda_0 \neq 0$ (equivalentemente, $\lambda_0 = 1$) é chamada condição de Karush-Kuhn-Tucker (KKT). Karush [55], de forma independente, também estudou tais condições, mas em seu trabalho ele não enuncia o resultado de Fritz John.

Karush desenvolveu seu resultado em 1939, em sua dissertação de mestrado no Departamento de Matemática da Universidade de Chicago. Tal departamento era muito ativo no problema de otimização análogo em dimensão infinita (cálculo variacional), e a dissertação de Karush não chamou muita atenção por se tratar de um caso marginal do problema de interesse do departamento, e por isso o resultado não foi publicado. Seu trabalho só veio a ser reconhecido pela comunidade científica na década de 70, anos depois de Kuhn e Tucker terem obtido grande atenção pela publicação do teorema em 1951. Tucker, um topólogo de Princeton e Kuhn, doutorando na área de álgebra, utilizaram ideias de teoria dos jogos de Von Neumann para formular a teoria de otimização não linear e dualidade. As ferramentas de teoria dos jogos são utilizadas para estudar o equilíbrio de Nash $L(x^*, \lambda, \mu) \leq L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*)$, $\forall x, \lambda, \mu$, onde $x^*$ é uma solução e $(\lambda^*, \mu^*)$ é um multiplicador de Lagrange associado, onde $L(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \sum_{j=1}^p \mu_j g_j(x)$, $\mu_j \geq 0$, $j = 1, \ldots, p$ é a função Lagrangiano. Na década de 50, com a introdução dos computadores e com os casos de sucesso de implementações do método simplex de Dantzig em problemas lineares advindos da área militar, o financimento e o interesse em otimização cresceu muito, tornando os resultados de (Karush, ) Kuhn e Tucker muito conhecidos.

No Capítulo 2, iremos apresentar diversas condições de otimalidade baseadas em condições de qualificação distintas. Além disso, apresentaremos as chamadas condições sequenciais de otimalidade, que não dependem de uma condição de qualificação e estão mais intimamente ligadas à convergência de algoritmos. No Capítulo 3, desenvolvemos condições de otimalidade mais específicas, utilizando informações de segunda ordem do problema.

## 1.3 Exercícios

Dado um conjunto $X \subset \mathbb{R}^n$, denote por $\mathrm{conv}(X)$ o menor conjunto convexo que contem $X$ e por $\mathrm{cone}(X)$ o menor cone que contem $X$.

1. Dado $\Omega \subset \mathbb{R}^n$ e $x \in \Omega$. Mostre que $T_\Omega(x)$ é sempre um conjunto fechado.

2. Dizemos que um conjunto $X \subset \mathbb{R}^n$ é convexo se $\alpha x + (1-\alpha)y \in X$ para todo $x, y \in X$ e para todo $\alpha \in [0, 1]$.
   Prove que se $\Omega$ é convexo, então $T_\Omega(x)$ é convexo para todo $x \in \Omega$. Dica: Prove que neste caso $T_\Omega(x) := \mathrm{cl}\ \mathrm{cone}(\Omega - x)$, onde cl denota o fecho.

3. Sejam $\Omega_1$ e $\Omega_2$ dois subconjuntos de $\mathbb{R}^n$ e $x^* \in \Omega_1 \cap \Omega_2$ tais que $\Omega_1 \cap \mathbb{B}(x^*, \delta) = \Omega_2 \cap \mathbb{B}(x^*, \delta)$ para algum $\delta > 0$. Então, $T_{\Omega_1}(x^*) = T_{\Omega_2}(x^*)$. Em outras palavras, o cone tangente depende somente do comportamento local do conjunto em questão.

4. Se $\Omega = \Omega_1 \times \cdots \times \Omega_m$ onde cada $\Omega_i$ é um conjunto fechado. Então, mostre que $T_\Omega(x) \subset T_{\Omega_1}(x_1) \times \cdots \times T_{\Omega_m}(x_m)$, $x = (x_1, \ldots, x_m) \in \Omega$. Dê um exemplo onde a inclusão é estrita. Analise o caso onde cada $\Omega_i$ é um convexo fechado.

5. Calcule $T_\Omega(x)$ onde $\Omega = \Omega_1 \times \cdots \times \Omega_m$ e cada $\Omega_i := [a_i, b_i] \subset \mathbb{R}$. Considere o caso onde $a_i$ e $b_i$ possam tomar valores $\pm\infty$.

6. (a) Se $\Omega := \{x \in \mathbb{R}^n : F(x) \in \mathcal{C}\}$ onde $\mathcal{C}$ é um cone fechado em $\mathbb{R}^m$ e $F : \mathbb{R}^n \to \mathbb{R}^m$ é continuamente diferenciável em $x \in \Omega$, conclua que

$$T_\Omega(x) \subset \{d \in \mathbb{R}^n : DF(x)d \in T_\mathcal{C}(F(x))\},$$

   onde $DF(x)$ é a matriz jacobiana de $F$ em $x$. O último cone é chamado de cone linearizado de $\Omega$ em $x$ e é denotado por $\mathcal{L}_\Omega(x)$.

(b) Suponha que $F(x) := (g_1(x), \dots, g_m(x))$ e $\mathcal{C} := \mathbb{R}^m$. Encontre $\mathcal{L}_\Omega(x)$.

7. *Teorema de Lyusternink.* Considere $\Omega := \{x \in \mathbb{R}^n : F(x) = 0\}$, $x^* \in \Omega$ e $F : \mathbb{R}^n \to \mathbb{R}^m$ é continuamente diferenciável. Suponha que $DF(x^*)$ é sobrejetiva (i.e. $DF(x^*)(\mathbb{R}^n) = \mathbb{R}^m$). Então $T_\Omega(x^*) = \{d \in \mathbb{R}^n : DF(x^*)d = 0\}$.

8. Prove as propriedades:

   (a) Considere $C_1, C_2 \subset \mathbb{R}^n$ tais que $C_1 \subset C_2$. Então, $C_2^\circ \subset C_1^\circ$.

   (b) Sejam $C_1 \subset \mathbb{R}^{n_1}$, $C_2 \subset \mathbb{R}^{n_2}$ dois cones. Então
   $$(C_1 \times C_2)^\circ = C_1^\circ \times C_2^\circ.$$

   (c) Sejam $C_1, C_2 \subset \mathbb{R}^n$ dois cones. Então
   $$(C_1 + C_2)^\circ = C_1^\circ \cap C_2^\circ.$$

   (d) (Teorema Bipolar). Se $C$ é um cone convexo fechado. Então,
   $$(C^\circ)^\circ = C.$$

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

   - $\sum_{i=1}^m \beta_i a_i = b$, $0 \le \beta_1, \dots, \beta_m$.
   - $a_i^\mathsf{T} x \ge 0$, $\forall i = 1, \dots, m$ e $b^\mathsf{T} x > 0$, $x \in \mathbb{R}^n$.

   Use o Lema de Farkas para provar que a condição de Fritz John vale em minimizadores locais (Teorema 1.2).

10. Suponha que as funções $f$, $g_j$ são convexas e $h_i$ são afins. Se a condição de Fritz John vale em $x^*$ com $\lambda_0 \neq 0$, então $x^*$ é um minimizador local.

11. Suponha que a condição de Fritz John vale em um ponto viável $x^*$ com multiplicadores $(\lambda_0, \lambda, \mu)$, isto é:
    $$\lambda_0 \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^p \mu_j \nabla g_j(x^*) = 0$$
    com $\lambda_0 \ge 0$, $\mu_j \ge 0$ e $\mu_j g_j(x^*) = 0$ $\forall j = 1, \dots, p$. Suponha agora que os vetores $\{\lambda_0 \nabla f(x^*), \nabla h_i(x^*), \mu_j \nabla g_j(x^*) : j = 1, \dots, p, i = 1, \dots, m\}$ geram o espaço $\mathbb{R}^n$. Então, temos que $x^*$ é um minimizador local.

12. A condição de Fritz John não é suficiente para garantir otimalidade.

    Considere o seguinte problema (com restrições lineares).
    $$\begin{array}{cl} \text{minimizar} & f(x,y) := -x \\ \text{sujeito a} & g_1(x,y) := x + y \le 0 \\ & g_2(x,y) := -x - y \le 0, \\ & g_3(x,y) := -x \le 0 \end{array}$$
    Mostre que a condição de Fritz John vale em $(0,0)$ mas $(0,0)$ não é minimizador local.

13. Considere o seguinte critério de parada de um algoritmo iterativo:
*Fixado $\varepsilon > 0$, pare ao encontrar um ponto $x$ e multiplicadores $\lambda$ e $\mu$ com $\mu \geq 0$ tais que*

$$\|(h_i(x))_{i=1}^m\| \leq \varepsilon, \|(\max\{0, g_j(x)\})_{j=1}^p\| \leq \varepsilon,$$
$$\|\nabla f(x) + \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j=1}^p \mu_j \nabla g_j(x)\| \leq \varepsilon \max\{1, \|\lambda\|_\infty, \|\mu\|_\infty\},$$
$$\text{para } j = 1, \ldots, p, \text{ se } g_j(x) < -\varepsilon, \text{ então } \mu_j = 0.$$

Suponha que $\varepsilon_k \to 0$ e $x^k$ é um ponto que satisfaz o critério de parada com $\varepsilon := \varepsilon_k$. Mostre que qualquer ponto limite de $\{x^k\}$ cumpre a condição de Fritz John.

# Capítulo 2

# Condições de otimalidade de primeira ordem

Sendo o cone tangente $\mathcal{T}_\Omega(x)$ um objeto geométrico difícil de ser simulado numericamente, gostaríamos de definir um objeto analítico que capture o mesmo espírito do cone tangente. A ideia será usar, em lugar de cada função de restrição, uma aproximação de primeira ordem para cada uma das funções de restrição. Assim, em torno de $x \in \Omega$, as restrições não lineares de igualdade, $h_i(x+d) = 0, i = 1, \ldots, m$, serão substituídas pelas restrições que definem o hiperplano tangente, a saber, $\nabla h_i(x)^\mathtt{T} d = 0$ e de forma análoga, as restrições de desigualdade $g_j(x+d) \leq 0, j \in A(x)$ serão substituídas pelas restrições que definem o hiperplano tangente $\nabla g_j(x)^\mathtt{T} d \leq 0$. Desta maneira, podemos definir um cone que captura a informação, até de primeira ordem, contida no cone tangente. Este cone é chamado de cone linearizado conforme a definição abaixo:

**Definição 2.1** (Cone linearizado)**.** *Dado $x \in \Omega$, denotamos por $\mathcal{L}_\Omega(x)$ o cone linearizado de $\Omega$ em $x$, dado por*

$$\mathcal{L}_\Omega(x) = \left\{ d \in \mathbb{R}^n \mid \nabla h_i(x)^\mathtt{T} d = 0, i = 1, \ldots, m; \nabla g_j(x)^\mathtt{T} d \leq 0, j \in A(x) \right\}.$$

O cone linearizado, em muitos casos, coincide com o cone tangente, como por exemplo, quando as restrições vêm dadas por funções lineares. Como veremos ao longo do texto, a relação exata entre o cone linearizado e o cone tangente terá um papel importante em nossa análise. Continuamos com a seguinte proposição.

**Proposição 2.1.** *Sempre se cumpre que $\mathcal{T}_\Omega(x^*) \subset \mathcal{L}_\Omega(x^*)$.*

*Demonstração.* Seja $d \in \mathcal{T}_\Omega(x^*)$, por definição, existem sequências $\{t_k\} \in \mathbb{R}_+$ e $\{d^k\}$ tais que $t_k \to 0$, $d^k \to d$ e $x^* + t_k d^k \in \Omega$ para todo $k \in \mathbb{N}$. Usando a expansão de Taylor e o fato que $h_i(x^*) = 0$, $g_j(x^*) = 0$, para todo $i = 1, \ldots, m$, $j \in A(x^*)$ temos

$$h_i(x^* + t_k d^k) = t_k \nabla h_i(x^*)^\mathtt{T} d^k + o(t_k) = 0; \qquad i \in \{1, \ldots, m\} \tag{2.0.1}$$

$$g_j(x^* + t_k d^k) = t_k \nabla g_j(x^*)^\mathtt{T} d^k + o(t_k) \leq 0; \qquad j \in A(x^*). \tag{2.0.2}$$

Dividindo por $t_k$ as expressões (2.0.1) e (2.0.2) e tomando limite, obtemos que $d$ pertence ao cone linearizado $\mathcal{L}_\Omega(x^*)$. □

Em muitos casos, a inclusão $\mathcal{T}_\Omega(x^*) \subset \mathcal{L}_\Omega(x^*)$ é estrita. De fato, considere o seguinte exemplo.

**Exemplo 2.1.** *(A inclusão $\mathcal{T}_\Omega(x^*) \subset \mathcal{L}_\Omega(x^*)$ pode ser estrita)*

Considere o ponto $x^* = (0,0)$ e a região viável

$$\Omega := \{(x_1, x_2) \in \mathbb{R}^2 : x_1 x_2 = 0, x_1 \geq 0, x_2 \geq 0\}.$$

Nesse caso, calculando, temos que o cone linearizado é o ortante positivo, i.e. $\mathcal{L}_\Omega(x^*) = \mathbb{R}_+ \times \mathbb{R}_+$, enquanto o cone tangente é formado pela união dos semi-eixos positivos, ainda mais $\mathcal{T}_\Omega(x^*) = \Omega$.

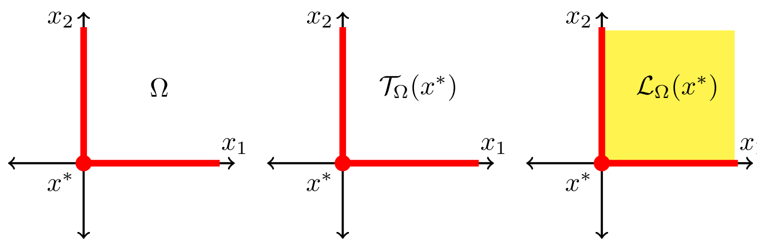



Figura 2.1: Conjunto viável $\Omega = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 x_2 = 0, x_1 \geq 0, x_2 \geq 0\}$, cone tangente $\mathcal{T}_\Omega(x^*)$ e cone linearizado $\mathcal{L}_\Omega(x^*)$.

Embora os cones não coincidam, é fácil verificar que o polar de ambos os cones é o ortante negativo. Assim, mesmo que a inclusão $\mathcal{T}_\Omega(x^*) \subset \mathcal{L}_\Omega(x^*)$ possa ser estrita, é possível que $\mathcal{T}_\Omega(x^*)^\circ = \mathcal{L}_\Omega(x^*)^\circ$. Da Proposição 2.1, sempre temos que $\mathcal{L}_\Omega(x)^\circ \subset \mathcal{T}_\Omega(x)^\circ$. O seguinte exemplo mostra que esta inclusão também pode ser estrita.

**Exemplo 2.2.** *(A inclusão $\mathcal{L}(x^*)^\circ \subset \mathcal{T}(x^*)^\circ$ pode ser estrita)*

Considere o ponto $x^* = (0,0)$ e a região viável

$$\Omega := \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \leq -x_1^3, x_2 \geq x_1^3\}.$$

Podemos verificar que $\mathcal{L}(x)^\circ = \{0\} \times \mathbb{R}$ e $\mathcal{T}(x)^\circ = \mathbb{R}_+ \times \mathbb{R}$. Antes de continuar,

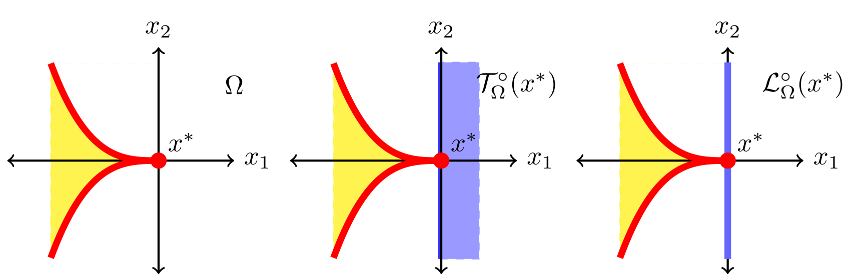



Figura 2.2: Conjunto viável $\Omega = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \leq -x_1^3, x_2 \geq x_1^3\}$, o polar do cone tangente $\mathcal{T}_\Omega^\circ(x^*)$ e o polar do cone linearizado $\mathcal{L}_\Omega^\circ(x^*)$

lembremos o Lema de Farkas.

**Lema 2.1.** *(Lema de Farkas) Considere vetores $v_1, v_2, \ldots, v_s$ e $w$ em $\mathbb{R}^n$. Então, exatamente um dos seguintes sistemas possui solução.*

1. $\sum_{i=1}^{s} \alpha_i v_i = w, \quad 0 \leq \alpha_i$ *para todo* $i = 1, \ldots, s$.

*2.* $w^{\mathsf{T}}x > 0$, $v_i^{\mathsf{T}}x \leq 0$ *para todo* $i = 1, \ldots, s$.

Supondo que $\mathcal{T}_\Omega(x)^\circ = \mathcal{L}_\Omega(x)^\circ$, a condição de otimalidade geométrica se reduz a $-\nabla f_0(x) \in \mathcal{L}_\Omega(x)^\circ$. Calculando o polar do cone linearizado, obtemos:

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

Com efeito, seja $S(x)$ o cone do lado direito da igualdade acima. Da definição de $S(x)$ temos que $S(x)$ é um cone convexo. O Exercício 3 do Capítulo 2 garante que $S(x)$ é um conjunto fechado. O Lema de Farkas assegura que $S(x)^\circ = \mathcal{L}_\Omega(x)$. Logo, temos que $\mathcal{L}_\Omega(x)^\circ = (S(x)^\circ)^\circ$. Já que $S(x)$ é um cone convexo fechado, o Exercício 8d do Capítulo 1 garante que $(S(x)^\circ)^\circ = S(x)$. Assim, $\mathcal{L}_\Omega(x)^\circ = S(x)$.

Note que a condição $-\nabla f(x) \in \mathcal{L}_\Omega(x)^\circ$ é precisamente a condição KKT.

**Teorema 2.1** (Karush-Kuhn-Tucker)**.** *Seja $x$ uma solução para o problema (1.1.1) e assuma que $\mathcal{T}_\Omega(x)^\circ = \mathcal{L}_\Omega(x)^\circ$. Então, existem multiplicadores de Lagrange $\lambda_i \in \mathbb{R}$, $i = 1, \ldots, m$ e $\mu_j \in \mathbb{R}_+$ com $\mu_j = 0$, $j \notin A(x)$ tais que*

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{j \in A(x)} \mu_j \nabla g_j(x) = 0. \tag{2.0.3}$$

Os pontos viáveis $x \in \Omega$ que cumprem (2.0.3) são chamados pontos estacionários (ou pontos KKT) do problema, e são os principais candidatos à solução quando o problema cumpre uma condição de qualificação.

**Definição 2.2.** *(Condição de Qualificação) Uma condição de qualificação (CQ) é uma proposição acerca da descrição analítica do conjunto $\Omega$, tal que sob essa proposição, todo minimizador local é um ponto estacionário.*

Note que, por definição, uma condição de qualificação depende somente do conjunto $\Omega$ e da sua representação analítica, e não da função objetivo.

Sendo assim, a igualdade dos cones polares é uma condição de qualificação, chamada condição de Guignard [50]. Mais do que isso, a condição de Guignard é a condição de qualificação mais fraca possível [48], 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 estacionário, então a condição de Guignard é satisfeita em $x^*$. O resultado segue da inclusão

$$\mathcal{T}_\Omega(x^*)^\circ \subset \{-\nabla f(x^*) \mid f \text{ assume um minimizador local restrito a } \Omega \text{ em } x^*\}$$

e será feita no Teorema 2.4 no final desta sessão, onde corrigimos uma pequena imprecisão de [48]. Note que a inclusão recíproca

$$\mathcal{T}_\Omega(x^*)^\circ \supset \{-\nabla f(x^*) \mid f \text{ assume um minimizador local restrito a } \Omega \text{ em } x^*\}$$

também é satisfeita, e é precisamente uma re-escrita da condição de otimalidade geométrica do Teorema 1.1.

A condição KKT fornece uma relação simples entre a função objetivo e as restrições do problema. Ainda mais, ela pode ser facilmente verificada e implementada em algoritmos. Por outro lado, a condição KKT não é uma condição de otimalidade, considere por exemplo o problema de minimizar $x$, sujeito a $x^2 = 0$. Neste caso, a única solução do problema é $x = 0$ e não é um ponto estacionário.

Apesar disso, podemos associar à condição KKT, condições de otimalidade. As condições de otimalidade clássicas associadas à condição KKT são da forma

"KKT ou não-CQ",

onde CQ é uma condição de qualificação qualquer.

Quanto menos exigente a condição de qualificação, mais forte é a condição de otimalidade associada, e mais interessante se torna na prática, pois possui maior chance de ser verificada. Com efeito, considere duas condições de qualificação $\mathrm{CQ}_1$ e $\mathrm{CQ}_2$ e suponha que $\mathrm{CQ}_1$ implica $\mathrm{CQ}_2$. Em outras palavras, $\mathrm{CQ}_2$ é mais fraca que $\mathrm{CQ}_1$. Por conseguinte, se $\mathrm{CQ}_2$ não se verifica, $\mathrm{CQ}_1$ também não se verifica. Assim, obtemos a seguinte implicação:

$$\text{“KKT ou não-CQ}_2\text{”} \Rightarrow \text{“KKT ou não-CQ}_1\text{.”}$$

Em outra palavras, a condição de otimalidade "KKT ou não-$\mathrm{CQ}_2$" é mais forte que "KKT ou não-$\mathrm{CQ}_1$".

Embora conheçamos a condição de qualificação mais fraca possível (Guignard), não se conhece um algoritmo capaz de encontrar pontos estacionários, a menos que se exija mais regularidade do problema. Em outras palavras, não se conhece algoritmos que geram pontos que satisfazem a condição de otimalidade "KKT ou não-CQ", onde CQ é a condição de qualificação de Guignard. Sendo assim, o estudo de condições de qualificação mais exigentes que Guignard, mas que permitam a prova de convergência global de algoritmos, é o tema central deste trabalho. Em mais detalhes, temos tipicamente que algoritmos de otimização são iterativos e geram sequências possivelmente divergentes, entretanto, o que somos capazes de provar é que, para qualquer escolha do ponto inicial, se uma sequência gerada por um algoritmo possui um ponto de acumulação, então sob uma condição de qualificação (mais forte que Guignard), este ponto é estacionário para o problema. A um resultado deste tipo chamamos de *convergência global*, ou *convergência global de primeira ordem.*

## 2.1 Condições de qualificação

A condição de igualdade entre os cones $\mathcal{T}_\Omega(x) = \mathcal{L}_\Omega(x)$, mais exigente que Guignard, é chamada condição de Abadie [1], e é frequentemente satisfeita sob a maioria das condições de qualificação clássicas. Entretanto, supor Abadie ainda não é suficiente para a convergência global de algoritmos.

A seguir, vamos listar as condições de qualificação associadas a algoritmos, iniciando das mais restritivas, embora a conexão com algoritmos seja postergada para a próxima seção.

**Definição 2.3.** *(Condição de Independência Linear) Dizemos que a condição de independência linear (LICQ) - ou regularidade - vale no ponto viável $x$, se os gradientes das restrições de igualdade e de desigualdade ativas são linearmente inde-*

*pendentes, isto é,*

$$\{\nabla h_i(x), \nabla g_j(x) : i = 1 \ldots, m, \quad j \in A(x)\} \text{ são linearmente independentes.}$$

LICQ possui várias propriedades importantes. Por exemplo, sob LICQ, temos que os multiplicadores de Lagrange são únicos. De fato, se $(\lambda, \mu)$ e $(\widehat{\lambda}, \widehat{\mu})$ são dois multiplicadores associados a (1.1.1). Então temos que $\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{j \in A(x)} \mu_j \nabla g_j(x) = 0$ e $\nabla f(x) + \sum_{i=1}^{m} \widehat{\lambda}_i \nabla h_i(x) + \sum_{j \in A(x)} \widehat{\mu}_j \nabla g_j(x) = 0$. Subtraindo as expressões temos que

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

Da independência linear de $\{\nabla h_i(x), \nabla g_j(x) : i = 1 \ldots, m, \quad j \in A(x)\}$, concluímos que $\lambda = \widehat{\lambda}$ e $\mu = \widehat{\mu}$. Provando a unicidade das multiplicadores.

Uma abordagem clássica para a prova da existência de multiplicadores de Lagrange em uma solução do problema consiste em supor LICQ na solução e usar o Teorema da Função Implícita para provar a igualdade dos cones tangente e linearizado (condição de Abadie).

Exigindo LICQ, há pouco espaço para redundância na descrição do problema. De fato, se os gradientes de duas restrições são coincidentes, LICQ não é satisfeita.

Na definição a seguir podemos enfraquecer LICQ de modo a permitir este tipo de redundância. Explorando o fato que os multiplicadores de Lagrange associados às restrições de desigualdade são não-negativos, é possível definir uma condição de qualificação menos exigente que LICQ, denominada condição de Mangasarian-Fromovitz (MFCQ, [61]).

**Definição 2.4.** *(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 $i = 1, \ldots, m$, $j \in A(x)$.*

Neste caso diremos que os gradientes $\{\nabla h_i(x), \nabla g_j(x) : i = 1 \ldots, m, \quad j \in A(x)\}$ são positivo-linearmente independentes.

Utilizando o Lema de Farkas, podemos ver que a condição MFCQ, ver [73], é equivalente à existência de uma direção estritamente viável, no seguinte sentido:

1. $\{\nabla h_i(x^*), i = 1, \ldots, m\}$ é linearmente independente;

2. Existe uma direção $d \in \mathbb{R}^n$ tal que $\nabla h_i(x)^{\mathtt{T}} d = 0$ para todo $i = 1, \ldots, m$ e $\nabla g_j(x)^{\mathtt{T}} d < 0$ para todo $j \in A(x)$.

De fato, esta é a maneira como a condição foi originalmente proposta.

Embora sob MFCQ não se possa garantir a unicidade dos multiplicadores de Lagrange, sabe-se que ela é equivalente ao fato do conjunto de multiplicadores de Lagrange ser um conjunto fechado, limitado e não-vazio [45]. Mais especificamente, MFCQ garante a compacidade do conjunto de multiplicadores de Lagrange independente da função objetivo, e, se existe uma função objetivo tal que o conjunto de multiplicadores de Lagrange é compacto, então vale MFCQ. É interessante observar que a condição de otimalidade oriunda de MFCQ, isto é, "KKT ou não-MFCQ" é precisamente a condição de Fritz John. Isto é, sob MFCQ, todos os multiplicadores de Fritz John são também multiplicadores de Lagrange, o que não é verdade sob uma CQ mais fraca.

Em muitos casos de interesse, a unicidade dos multiplicadores de Lagrange simplifica a análise de convergência, assim como a análise de estabilidade para vários métodos numéricos. Assim, saber sob quais condições os multiplicadores de Lagrange são únicos, torna-se importante.

**Definição 2.5.** *(Condição de Mangasarian-Fromovitz estrita, [57]) 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* $\{\nabla h_i(x), \nabla g_j(x) : i \in \{1, \dots, m\}, j \in A(x) \text{ com } \mu_j > 0\}$ *e* $\{\nabla g_j(x) : j \in A(x) \text{ com } \mu_j = 0\}$ *são positivo-linearmente independentes (a parte "positiva" está associada ao segundo conjunto de vetores).*

Note que, sob SMFCQ, as restrições de desigualdade ativas associadas a um multiplicador positivo são tratadas como restrições de igualdade na definição de MFCQ.

**Proposição 2.2.** *A unicidade dos multiplicadores de Lagrange é equivalente à condição SMFCQ.*

*Demonstração.* Para mostrar a equivalência de SMFCQ com a unicidade do multiplicador, assuma que $(\bar{\lambda}, \bar{\mu}) \neq (\lambda, \mu)$ seja outro multiplicador de Lagrange associado a $x \in \Omega$, então, da condição KKT temos

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

o que contradiz SMFCQ.

Agora provaremos a reciproca. Seja $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$ um multiplicador de Lagrange. Suponha que não vale SMFCQ. Então, existe $0 \neq (\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^m \times \mathbb{R}^p_+$ com $\bar{\mu}_j \geq 0$ para $j \in A(x)$ e $\mu_j = 0$ tal que $\sum_{i=1}^m \bar{\lambda}_i \nabla h_i(x) + \sum_{j \in A(x)} \bar{\mu}_j \nabla g_j(x) = 0$. Claramente, temos que $\nabla f(x) + \sum_{i=1}^m (\bar{\lambda}_i + \lambda_i) \nabla h_i(x) + \sum_{j \in A(x)} (\bar{\mu}_j + \mu_j) \nabla g_j(x) = 0$. Tomando $(\bar{\lambda}, \bar{\mu})$ pequeno o suficiente de modo que $\bar{\mu}_j + \mu_j \geq 0$ para $j \in A(x)$ temos que $(\bar{\lambda} + \lambda, \bar{\mu} + \mu)$ é um multiplicador de Lagrange diferente de $(\lambda, \mu)$. Assim, não há unicidade dos multiplicadores de Lagrange. □

Observe que a definição de SMFCQ depende previamente da existência de um multiplicador de Lagrange (o que torna confusa a sua definição como uma "condição de qualificação"), e portanto depende implicitamente da função objetivo. Uma condição de qualificação que é equivalente à existência e unicidade de multiplicadores de Lagrange para qualquer função objetivo é a LICQ [77], isto é, se assumimos que para qualquer função objetivo, existem multiplicadores de Lagrange que cumprem SMFCQ, então vale LICQ.

**Teorema 2.2.** *LICQ vale em* $x^*$ *se, e somente se para toda função objetivo suave* $f$ *que possua um minimizador local* $x^*$, *existe um único multiplicador de Lagrange associado a* $f$.

*Demonstração.* Só mostraremos que se para qualquer função objetivo, existem multiplicadores de Lagrange que cumprem SMFCQ, então vale LICQ.

Com efeito, sendo $x^* \in \Omega$ e definindo $f(x) := -\sum_{j\in A(x^*)} g_j(x)$ temos que $x^*$ é um minimizador local de $f$ em $\Omega$. Suponha que existem únicos multiplicadores de Lagrange, associados a $f$, a saber $\mu_j^* = 1, j \in A(x^*)$, $\mu_j^* = 0$ caso contrário e $\lambda_i^* = 0$, $\forall i$. Seja $(\alpha, \beta)$ tal que $\sum_{i=1}^m \alpha_i \nabla h_i(x^*) + \sum_{j\in A(x^*)} \beta_j \nabla g_j(x^*) = 0$. É fácil ver que para $C > 0$ grande o suficiente, $(\bar{\lambda}, \bar{\mu}) := (\lambda^*, \mu^*) + C^{-1}(\alpha, \beta)$ é também um multiplicador de Lagrange associado a $f$, donde concluímos que $(\bar{\lambda}, \bar{\mu}) = (\lambda^*, \mu^*)$. Assim, temos que $(\alpha, \beta) = 0$ e vale LICQ em $x^*$. □

Uma outra maneira de enfraquecer LICQ é, ao invés de requerer posto completo para o conjunto de gradientes, permitir que o posto seja deficiente, desde que permaneça deficiente em uma vizinhança do ponto. Esta condição é chamada *dependência linear constante*. Para que este enfraquecimento seja de fato uma condição de qualificação, é necessário exigir tal propriedade para todos os subconjuntos de gradientes, da maneira a seguir:

**Definição 2.6.** *Dizemos que vale a condição de posto constante (CRCQ, [53]) em $x \in \Omega$ se existe uma vizinhança $\mathcal{B}$ de $x$ tal que para todo $\mathcal{I} \subset \{1, \ldots, m\}$, $\mathcal{J} \subset A(x)$, 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 $y \in \mathcal{B}$.*

Equivalentemente, a condição de dependência linear constante pode ser substituída pela condição de posto constante, isto é, vale CRCQ se, e somente se, existe uma vizinhança $\mathcal{B}$ de $x$ tal que para todo $\mathcal{I} \subset \{1, \ldots, m\}$, $\mathcal{J} \subset A(x)$ o posto de $\{\nabla h_i(y), \nabla g_j(y) : i \in \mathcal{I}, j \in \mathcal{J}\}$ se mantém constante para cada $y \in \mathcal{B}$.

Em uma primeira análise, CRCQ parece ter pouca relação com LICQ, já que CRCQ faz exigências em toda uma vizinhança de $x$, além de ser descrita para todo subconjunto de restrições. Entretanto, ao exigirmos LICQ, estamos exigindo independência linear para todos os subconjuntos de restrições e em toda uma vizinhança do ponto, em outras palavras, LICQ pode ser equivalentemente descrita como: Existe uma vizinhança $\mathcal{B}$ de $x$ tal que para todo $\mathcal{I} \subset \{1, \ldots, m\}$, $\mathcal{J} \subset A(x)$, o conjunto $\{\nabla h_i(y), \nabla g_j(y) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é linearmente independente para cada $y \in \mathcal{B}$. Está observação, mostra que CRCQ é mais fraca que LICQ. Não é difícil encontrar exemplos onde CRCQ vale mas LICQ falhe.

Um caso frequente em modelagem é quando uma restrição de igualdade $h(x) = 0$ é descrita analiticamente como duas restrições de desigualdade $g_1(x) := h(x) \leq 0$ e $g_2(x) := -h(x) \leq 0$. Neste caso, embora os gradientes sejam positivo-linearmente dependentes, o que faz com que MFCQ falhe, a dependência linear se mantém em uma vizinhança, o que faz com que CRCQ seja satisfeita. Outro caso frequente em que CRCQ é satisfeita é quando as restrições são descritas por funções afins (restrições lineares), o que justifica o fato que em otimização linear, sempre existem multiplicadores de Lagrange (dados pela solução do problema dual).

Embora as condições CRCQ e MFCQ não tenham nenhuma relação de implicação, é sabido que sob CRCQ, é possível reformular o problema de modo a cumprir MFCQ (ver [58]).

Da mesma maneira que MFCQ enfraquece LICQ olhando para o sinal dos multiplicadores, a condição de dependência linear positiva constante (CPLD, [71]) enfraquece CRCQ no mesmo sentido.

**Definição 2.7.** *Dizemos que a condição de dependência linear positiva constante (CPLD) vale em $x \in \Omega$ se existe uma vizinhança aberta $\mathcal{B}$ de $x$ tal que para todo $\mathcal{I} \subset \{1, \ldots, m\}$, $\mathcal{J} \subset A(x)$, temos que se $\{\nabla h_i(x), \nabla g_j(x) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é positivo-linearmente dependente, então $\{\nabla h_i(y), \nabla g_j(y) : i \in \mathcal{I}, j \in \mathcal{J}\}$ deve permanecer positivo-linearmente dependente para todo $y \in \mathcal{B}$.*

Note que na definição de dependência linear positiva, devemos considerar os gradientes associados às restrições de desigualdade ativas como os índices associados aos escalares não-negativos.

Mais recentemente, outro tipo de enfraquecimento destas condições surgiu na literatura. A ideia é não considerar todos os subconjuntos de restrições, em lugar, só devemos considerar certos subconjuntos.

**Definição 2.8.** *Dizemos que vale Relaxed-CRCQ (RCRCQ, [65]) no ponto viável $x$, se existe uma vizinhança $\mathcal{B}$ de $x \in \Omega$ tal que para todo $\mathcal{J} \subset A(x)$, o posto de $\{\nabla h_i(y), \nabla g_j(y) : i = 1, \ldots, m; j \in \mathcal{J}\}$ permanece constante para todo $y \in \mathcal{B}$.*

Observe que na condição RCRCQ, só é necessário verificar a propriedade de posto constante para subconjuntos que incluem todas as restrições de igualdade. Assim, a verificação de RCRCQ é mais simples que a verificação de CRCQ.

Para definirmos uma relaxação análoga para CPLD, chamada Relaxed-CPLD, (RCPLD, [9]), vamos interpretar a condição RCRCQ em termos de dependência linear constante.

Considere um conjunto $B \subset \{1, \ldots, m\}$ tal que $\{\nabla h_i(x), i \in B\}$ forma uma base para o subespaço gerado pelos gradientes $\{\nabla h_i(x) : i = 1 \ldots, m\}$. Tal subespaço é denotado como $\operatorname{span}\{\nabla h_i(x) : i = 1, \ldots, m\}$.

A condição RCRCQ pode ser equivalentemente definida como a existência de uma vizinhança $\mathcal{B}$ de $x$ tal que:

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

A definição de RCPLD é como a definição acima, mas utilizando a propriedade de dependência linear positiva constante.

**Definição 2.9.** *Dizemos que vale Relaxed-CPLD (RCPLD, [9]) no ponto viável $x$, se fixado $B \subset \{1, \ldots, m\}$ tal que $\{\nabla h_i(x), i \in B\}$ forma uma base para span$\{\nabla h_i(x) : i = 1 \ldots, m\}$ as seguintes condições valem:*

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

Como mostra [9], a definição é independente da escolha da base $B$.

As relaxações acima são incompletas no sentido que ainda exigem a verificação da condição para todos os subconjuntos de restrições de desigualdade ativas. A seguir, vamos definir a condição CRSC (*constant rank of the subspace component*, [10]) que detecta exatamente qual é o subconjunto de gradientes que deve manter o posto, para garantir a existência de multiplicadores de Lagrange.

Considere o polar do cone linearizado

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

e considere o maior subespaço contido em $\mathcal{L}_\Omega(x)^\circ$. Já que $\mathcal{L}_\Omega(x)^\circ$ é um cone fechado convexo, o maior subespaço contido em $\mathcal{L}_\Omega(x)^\circ$ é dado por $\mathcal{L}_\Omega(x)^\circ \cap -\mathcal{L}_\Omega(x)^\circ$ e é chamado de *subespaço componente*.

Devido a que $\mathcal{L}_\Omega(x)^\circ$ é finitamente gerado, o subespaço $\mathcal{L}_\Omega(x)^\circ \cap -\mathcal{L}_\Omega(x)^\circ$ é também finitamente gerado. Vejamos quais são os vetores que geram este subespaço. Certamente, os vetores $\nabla h_i(x)$ devem pertencer ao conjunto de geradores deste espaço. Agora, considere os vetores $\nabla g_j(x)$. O vetor $\nabla g_j(x)$ encontra-se em $\mathcal{L}_\Omega(x)^\circ \cap -\mathcal{L}_\Omega(x)^\circ$ se e somente se $-\nabla g_j(x) \in \mathcal{L}_\Omega(x)^\circ$. Isto nos leva a considerar o seguinte conjunto de índices, $J_-(x) := \{j \in A(x) : -\nabla g_j(x) \in \mathcal{L}_\Omega(x)^\circ\}$. Com essa notação não é difícil ver que o subespaço $\mathcal{L}_\Omega(x)^\circ \cap -\mathcal{L}_\Omega(x)^\circ$ é o subespaço gerado pelos vetores $\{\nabla h_i(x), \nabla g_j(x) : i = 1, \ldots, m, j \in J_-(x)\}$.

As restrições de desigualdade com índices em $J_-(x)$ correspondem a restrição de desigualdade que se comporta como uma restrição de igualdade, pelo menos na maneira de compor o polar do cone linearizado, já que os escalares correspondentes a esta restrição tem sinal irrestrito.

**Definição 2.10.** *Dizemos que $x \in \Omega$ satisfaz CRSC quando existe uma vizinhança $\mathcal{B}$ de $x$ tal que os gradientes $\{\nabla h_i(y), \nabla g_j(y) : i = 1, \ldots, m, j \in J_-(x)\}$ têm posto constante para todo $y \in \mathcal{B}$.*

De fato, sob CRSC, é possível mostrar (ver [10]) que as restrições em $J_-(x)$ atuam localmente como igualdades, isto é, existe uma vizinhança $\mathcal{B}$ de $x$ tal que para todo $y \in \mathcal{B} \cap \Omega$ e para todo $i \in J_-(x)$, vale $f_i(y) = 0$.

A condição CRSC unifica todas as demais, no sentido que é implicada pelas condições apresentadas anteriormente (LICQ, MFCQ, (R)CRCQ e (R)CPLD). O conceito de posto constante é adequado para lidar com restrições de igualdade, enquanto o conceito de dependência linear positiva se adapta melhor às desigualdades, enquanto a CRSC incorpora de maneira simples os dois conceitos.

Em certo sentido, todas as condições de qualificação anteriores estão garantindo que a decomposição de $\mathcal{L}_\Omega(x)^\circ$ na soma direta entre seu maior subespaço e um cone pontudo, seja tal que a componente de subespaço mantenha localmente a dimensão.

Para ter uma melhor compreensão geométrica de $\mathcal{L}_\Omega(x)^\circ$, podemos analisar as possíveis decomposições de $\mathcal{L}_\Omega(x)^\circ$ na soma direta entre seu maior subespaço e um cone pontudo. No caso em que o espaço é $\mathbb{R}$, temos as seguintes possibilidades, dependendo da forma de $\mathcal{L}_\Omega(x)^\circ$.

- Um ponto. Nesse caso, tanto o cone pontudo como o maior subespaço coincidem e ambos têm dimensão 0.
- Um raio. O subespaço componente tem dimensão 0 e o cone pontudo tem dimensão 1. Nesse caso, o cone pontudo coincide com o polar do cone linearizado.
- Uma linha. Neste caso, o subespaço componente tem dimensão 1 e o cone pontudo tem dimensão 0.

Em $\mathbb{R}^2$, além das possibilidades já mencionadas, temos:

- Um setor angular. Aqui, o subespaço componente tem dimensão 0 e o cone pontudo tem dimensão 2.
- Um semi-plano. O subespaço componente tem dimensão 1 e o cone pontudo tem dimensão 1 ou dimensão 2 (apontando para o interior do semi-plano).
- Todo o espaço. Nesse caso, o subespaço componente tem dimensão 2 e o cone pontudo tem dimensão 0.

A Figura 2.3 mostra uma decomposição em uma soma direta entre o subespaço componente e o máximo cone pontudo.

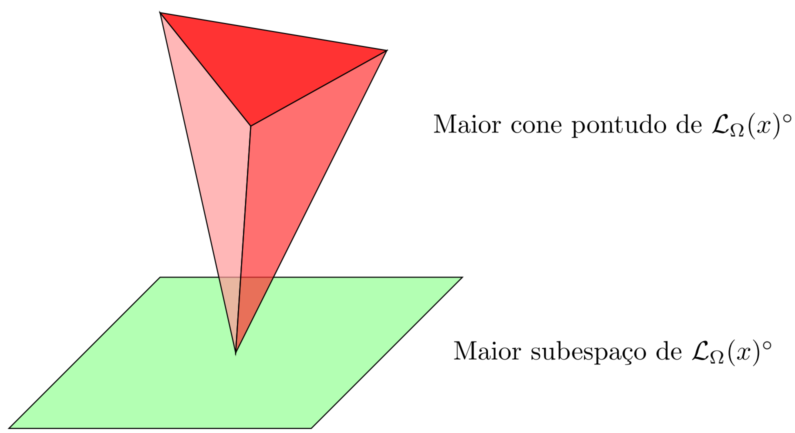



Figura 2.3: Decomposição de um cone em um subespaço e em um cone pontudo.

Todos os enfraquecimentos acima são possíveis ainda mantendo o fato de ser uma condição de qualificação, além disso, as condições apresentadas são *estáveis*, no sentido que se valem em um ponto, então valem em uma vizinhança deste ponto.

Mais importante, as condições de qualificação apresentadas fornecem condições de otimalidade mais fortes e possíveis de serem verificadas na prática, no sentido que somos capazes de construir algoritmos eficientes com convergência global para um ponto estacionário sob estas condições. Além disto, todas as condições apresentadas garantem a validade da *propriedade de error bound*, isto é, a distância ao conjunto viável pode ser medida, em uma vizinhança de $x$, utilizando uma medida analítica de inviabilidade.

**Definição 2.11.** *Dizemos que vale a condição de Error Bound em $x \in \Omega$ se existe uma vizinhança $\mathcal{B}$ de $x$ e uma constante $\alpha > 0$ tal que para todo $y \in \mathcal{B}$ vale:*

$$\min_{z\in\Omega} \|z - y\| \leq \alpha \max\{|h_1(y)|, \ldots, |h_m(y)|, \max\{0, g_1(y)\}, \ldots, \max\{0, g_p(y)\}\}.$$

**Teorema 2.3.** *A propriedade de Error Bound é uma condição de qualificação.*

*Demonstração.* Seja $d \in \mathcal{L}_\Omega(x)$. Pelo teorema de Taylor, temos que $h_i(x + td) = o(t)$, $i = 1, \ldots, m$, $g_j(x + td) \leq o(t)$, $j \in A(x)$ e $g_j(x + td) < 0$, $j \notin A(x)$, para $t > 0$ suficientemente pequeno, onde $o(t)/t \to 0$, quando $t \to 0$. Logo, usando a propriedade do error boud, temos que $\text{dist}(x + td, \Omega) = o(t)$. Logo, não é difícil ver que $d \in \mathcal{T}_\Omega(x)$. Assim, a condição de Abadie é satisfeita. $\square$

A Figura 2.4 mostra a relação entre as condições até aqui apresentadas.

É interessante observar que o error bound é a condição de qualificação mais fraca possível que garante a validade da condição KKT para todas as funções objetivo da forma $f := f_s + f_c$, onde $f_s$ é continuamente diferenciável e $f_c$ é convexa possivelmente não-suave [20]. Para mais informações acerca do significado das condições KKT para o caso não-suave, ver [26, 33, 66, 75].

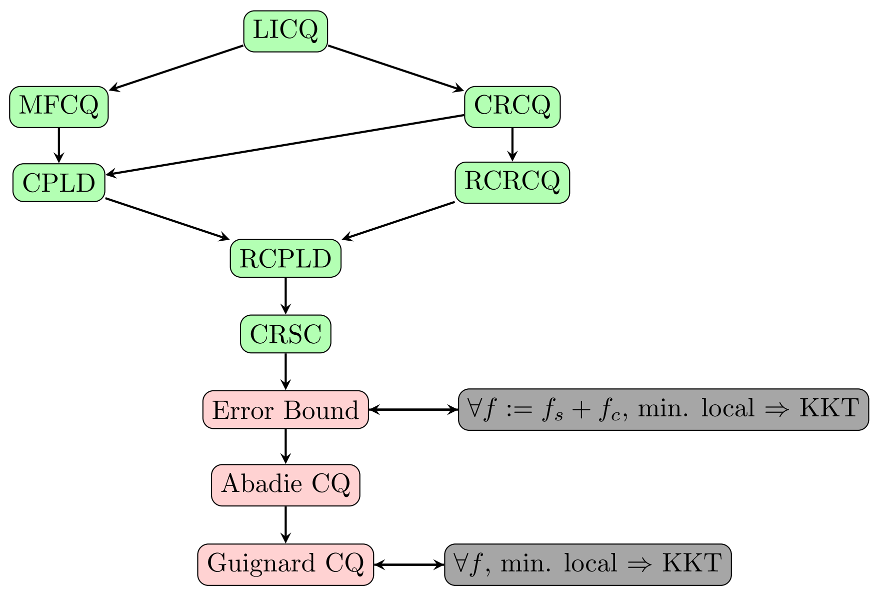



Figura 2.4: Relação entre as condições de qualificação apresentadas. Onde $f_s$ é uma função $C^1$ e $f_c$ é uma função convexa (possivelmente não-suave) quaisquer.

No caso suave, a condição mais fraca possível que garante KKT é a condição de Guignard. Terminamos esta sessão com a prova deste resultado de [48], onde corrigimos algumas pequenas imprecisões do trabalho original.

**Teorema 2.4.** *Seja $x \in \Omega$. Então $\mathcal{L}_\Omega(x)^\circ = \mathcal{T}_\Omega(x)^\circ$ (condição de Guignard) é equivalente ao fato que para qualquer função $f$ que assume mínimo restrito a $\Omega$ em $x$, é tal que $x$ é um ponto KKT.*

*Demonstração.* O fato que a condição de Guignard é uma condição de qualificação já foi mostrado ao longo do texto. Sabemos também que $\mathcal{L}_\Omega(x)^\circ \subset \mathcal{T}_\Omega(x)^\circ$.

Agora, seja $d \neq 0 \in \mathcal{T}_\Omega(x)^\circ$. Vamos construir uma função suave $f$ que assume um mínimo local em $x$ (restrito a $\Omega$), de modo que $d = -\nabla f(x)$. A hipótese de $x$ ser estacionário para este problema resulta em $d \in \mathcal{L}_\Omega(x)^\circ$. Definimos $C_k, k \geq 1$ o cone de direções não nulas que formam ângulo entre 0 e $\frac{\pi}{2} - \frac{\pi}{k+3}$ com $d$. Então, para cada $k \geq 1$, existe $\hat{\varepsilon}_k > 0$, tal que $\Omega \cap \mathcal{B}(x, \hat{\varepsilon}_k) \subset \mathbb{R}^n \backslash C_k$. Com efeito, se existisse $k$ tal que $x^\ell \in C_k \cap \Omega, x^\ell \to x$, então $\left\langle d, \frac{x^\ell}{\|x^\ell\|} \right\rangle \geq \|d\| \cos(\frac{\pi}{2} - \frac{\pi}{k+3}) > 0$. Tomando o limite em $\ell$ para uma subsequência temos $\langle d, y \rangle > 0$ com $y \in \mathcal{T}_\Omega(x)$, o que contradiz o fato de $d \in \mathcal{T}_\Omega(x)^\circ$. Definimos $\varepsilon_1 = \min(\hat{\varepsilon}_1, 1)$ e $\varepsilon_k = \min(\hat{\varepsilon}_k, \frac{\varepsilon_{k-1}}{2}), k > 1$.

Vamos assumir sem perda de generalidade que $x = 0$ e $d = (0, \ldots, 0, 1)$, assim, definimos $P : \mathbb{R}^{n-1} \to \mathbb{R}$ no subespaço ortogonal a $d$ da seguinte maneira:

$$P(z) = \begin{cases} 0, \quad \text{se } z = 0, \\ \tan(\frac{\pi}{3}), \quad \text{se } \|z\| \geq \varepsilon_2, \\ \tan(\frac{\pi}{k+2})\varepsilon_{k+1} + \frac{\tan(\frac{\pi}{k+1})\varepsilon_k - \tan(\frac{\pi}{k+2})\varepsilon_{k+1}}{\varepsilon_k - \varepsilon_{k+1}}(\|z\| - \varepsilon_{k+1}), \text{ se } \varepsilon_{k+1} \leq \|z\| < \varepsilon_k. \end{cases}$$

Observamos que $P$ é linear por partes e contínua. Além disso, para $\varepsilon_{k+1} \leq \|z\| < \varepsilon_k$, observamos que $\tan(\frac{\pi}{k+2})\|z\| \leq P(z) \leq \tan(\frac{\pi}{k+1})\|z\|$. Este fato é uma propriedade geométrica de funções afins crescentes que assumem um valor negativo na origem. Isso mostra que $P$ é diferenciável na origem com $\nabla P(0) = 0$. Agora, seja $f : \mathbb{R}^n \to \mathbb{R}$ definida por $f(y) = P(z) - d^\mathrm{T} y$, onde $y = (z, w) \in \mathbb{R}^{n-1} \times \mathbb{R}$ e vamos mostrar que $f$ restrita a $\Omega$ assume um mínimo local em 0. Com efeito, seja $y = (z, w) \neq (0, 0) \in \mathbb{R}^{n-1} \times \mathbb{R}$ com $y \in \Omega \cap \mathcal{B}(0, \varepsilon_3)$. Como $d = (0, \ldots, 0, 1)$ e $P(0) = 0$, basta mostrar que $w < P(z)$. Se $z = 0$, caso $w > 0$ teríamos que $y$ seria um múltiplo positivo de $d \in \mathcal{T}_\Omega(0)^\circ$, o que contradiz o fato de $y \in \Omega$. Então $w < 0 = P(z)$.

Sendo $z \neq 0$, existe $k \geq 2$ tal que $\varepsilon_{k+1} \leq \|z\| < \varepsilon_k$. Logo $P(z) \geq \|z\| \tan(\frac{\pi}{k+2})$. Note que $y \notin C_r$ é equivalente a dizer que $w < \tan(\frac{\pi}{r+3})\|z\|$. Sendo $y \in \Omega \cap \mathcal{B}(0, \varepsilon_3) \subset \mathbb{R}^n \backslash C_3$ temos $w < \tan(\frac{\pi}{3+3})\|z\| < \|z\|$. Assumindo $w > 0$, já que caso contrário o resultado é trivialmente verdadeiro, temos $\|y\| = \sqrt{\|z\|^2 + |w|^2} < \sqrt{2}\|z\| < 2\varepsilon_k \leq \varepsilon_{k-1}$. Portanto $y \in \Omega \cap \mathcal{B}(0, \varepsilon_{k-1}) \subset \mathbb{R}^n \backslash C_{k-1}$. Sendo $y \notin C_{k-1}$, temos $w < \tan(\frac{\pi}{k-1+3})\|z\|$. Segue que $w < P(z)$.

Sendo $f(0) = 0$ e $f(y) > 0$ para $y \neq 0 \in S$ suficientemente pequeno, segue da hipótese que $x = 0$ é um ponto KKT, ou seja $d = -\nabla f(0) \in \mathcal{L}_\Omega(x)^\circ$. □

A construção pode ser feita com $f$ diferenciável em todo $\mathbb{R}^n$ e de modo que $x$ seja o único minimizador global de $f$ restrito a $\Omega$. Ver [74, Teorema 6.11].

## 2.2 Condições sequenciais de primeira ordem

Na prática, usando algoritmos iterativos, não verificamos condições do tipo "KKT ou não-CQ" onde CQ é alguma condição de qualificação. Isso porque um algoritmo iterativo gera uma sequência $\{x^k\}$, onde apenas no limite esta condição é verificada. Numericamente, é preciso utilizar condições de otimalidade que possam ser verificadas no ponto $x^k$ disponível na iteração $k$ do método. Sendo assim, o que se verifica, tipicamente, é se $x^k$ satisfaz, aproximadamente, a condição KKT. Isso fornece certa confiança de que $x^k$ é um bom candidato para a solução do problema e o algoritmo para. Condições sequenciais de otimalidade fornecem a ferramenta teórica que justifica esta prática.

Dizemos que $x$ satisfaz a condição sequencial de otimalidade associada à proposição matemática $\mathcal{P}$ se existe uma sequência $x^k \to x$ tal que $\mathcal{P}(\{x^k\})$ é satisfeita. A proposição deve ser de tal forma que sempre que $x$ é uma solução do problema, deve existir uma sequência $x^k \to x$ que cumpre $\mathcal{P}$. Tipicamente o cumprimento de $\mathcal{P}$ está associado a um parâmetro $\varepsilon_k \to 0$. O parâmetro $\varepsilon_k$ pode ser considerado como tolerância ao erro. Assim, se sabemos que um algoritmo gera sequência que cumpre $\mathcal{P}$, é seguro terminar a execução do algoritmo quando o parâmetro $\varepsilon_k$ associado é pequeno o suficiente. No texto mencionamos as duas condições sequenciais mais conhecidas, a condição AKKT e a condição AGP.

### 2.2.1 Condição Aproximadamente KKT

A condição sequencial de otimalidade mais frequentemente utilizada na prática é a condição Aproximadamente-KKT que definimos a seguir:

**Definição 2.12** (Aproximadamente-KKT, [7]). *A condição Aproximadamente KKT (AKKT) é satisfeita no ponto viável $x \in \Omega$, se existem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{(\lambda^k, \mu^k)\} \subset \mathbb{R}^m \times \mathbb{R}^p_+$ com $\mu^k_j = 0$, $j \notin A(x)$ e $x^k \to x$ tais que*

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda^k_i \nabla h_i(x^k) + \sum_{j \in A(x)} \mu^k_j \nabla g_j(x^k) \to 0.$$

Quando $\{x^k\}$ é uma sequência como na definição acima dizemos que $\{x^k\}$ é uma sequência AKKT. Note que a sequência não precisa ser formada por pontos viáveis.

O teorema a seguir mostra como verificar AKKT na prática e como é usada como critério de parada.

**Teorema 2.5.** *O ponto $x$ satisfaz AKKT se, e só se, existem sequências $\{x^k\}$, $\{(\lambda^k, \mu^k)\} \subset \mathbb{R}^m \times \mathbb{R}^p_+$ e $\{\varepsilon_k\} \subset \mathbb{R}_+$ tais que $x^k \to x$, $\varepsilon_k \to 0$ e*

$$\begin{gathered}
\|(h_i(x^k))_{i=1}^m\| \leq \varepsilon_k, \|(\max\{0, g_j(x^k)\})_{j=1}^p\| \leq \varepsilon_k, \\
\|\nabla f(x^k) + \sum_{i=1}^{m} \lambda^k_i \nabla h_i(x^k) + \sum_{j=1}^{p} \mu^k_j \nabla g_j(x^k)\| \leq \varepsilon_k, \\
\forall j = 1, \ldots, p, \text{ se } g_j(x^k) < -\varepsilon_k \text{ então } \mu^k_j = 0.
\end{gathered}$$

*Demonstração.* Suponha que $x$ satisfaz AKKT. Procedemos a definir $\varepsilon_k$ satisfazendo as condições descritas acima. Defina

$$\begin{aligned}
\varepsilon_k := \; & \max\{\|(h_i(x^k))_{i=1}^m\|, \|(\max\{0, g_j(x^k)\})_{j=1}^p\|, \\
& \|\nabla f(x^k) + \textstyle\sum_{i=1}^m \lambda^k_i \nabla h_i(x^k) + \sum_{j=1}^p \mu^k_j \nabla g_j(x^k)\|, -g_j(x^k) : j \in A(x)\}.
\end{aligned}$$

Da definição de AKKT, temos que $\varepsilon_k \to 0^+$. Note que se $j_0 \in \{1, \ldots, p\}$ é tal que $g_{j_0}(x^k) < -\varepsilon_k$, então $-g_{j_0}(x^k) > \varepsilon_k \geq -g_j(x^k), \forall j \in A(x)$. Em particular, $j_0 \notin A(x)$. Portanto, basta re-definir $\mu^k_{j_0} = 0$ e temos o resultado.

Suponha agora que as condições valem. A continuidade das funções envolvidas garante que $x \in \Omega$. Agora observe que se $j \notin A(x)$, então $g_j(x) < 0$, e portanto, para $k$ suficientemente grande vale $g_j(x^k) < -\varepsilon_k$, logo $\mu^k_j = 0$ e vale AKKT. □

**Observação:** Se um algoritmo gera uma sequência que cumpre as condições descritas acima, temos que os pontos limites da sequência são AKKT, e portanto candidatos à solução. Na prática, o algoritmo para em uma iteração $k$ ao encontrar um iterando $x^k$ que cumpre as três condições do Teorema 2.5 para um $\varepsilon_k \leq \varepsilon$, uma tolerância pequena previamente estabelecida. Assim, é seguro dizer que se o algoritmo não fosse interrompido, a sequência gerada seria, provavelmente, AKKT. Portanto, $x^k$ é retornado como um bom candidato à uma aproximação para a solução do problema.

No restante da seção, mostraremos que AKKT é uma condição de otimalidade, e como tal deve ser satisfeita em qualquer minimizador local do problema, independente da validade de condições de qualificação.

Para isto, vamos primeiro considerar o algoritmo de penalidade externa para o problema

$$\begin{array}{ll} \text{Minimizar} & f(x), \\ \text{Sujeito a} & h_i(x) = 0, i = 1, \ldots, m, \\ & g_j(x) \leq 0, j = 1, \ldots, p, \\ & x \in \bar{\Omega}, \end{array}$$

onde $\bar{\Omega} \subset \mathbb{R}^n$ é um conjunto fechado e não vazio. Considere a função

$$P(x) := \sum_{i=1}^{m} h_i(x)^2 + \sum_{j=1}^{p} \max\{0, g_j(x)\}^2.$$

A função $P(x)$ pode ser considerada como uma medida de inviabilidade associada às restrições de igualdade e desigualdade.

**Teorema 2.6.** *Escolha uma sequência $\{\rho_k\} \subset \mathbb{R}$ com $\rho_k \to +\infty$ e para cada $k$, seja $x^k$ a solução (global), caso exista, para o problema*

$$\begin{array}{ll} \textit{Minimizar} & f(x) + \rho_k P(x), \\ \textit{sujeito a} & x \in \bar{\Omega}. \end{array}$$

*Então, todo ponto limite de $\{x^k\}$ é minimizador global do problema original.*

*Demonstração.* Seja $x^*$ um ponto limite arbitrário de $\{x^k\}$, caso exista. Como $\bar{\Omega}$ é fechado, temos $x^* \in \bar{\Omega}$. Para mostrar que $x^*$ é um minimizador global de $P(x)$, sujeito a $x \in \bar{\Omega}$, assuma $P(x^*) > P(z), z \in \bar{\Omega}$. Em uma subsequência apropriada temos $P(x^k) > P(z) + c$ para alguma constante $c > 0$, o que podemos re-escrever como $f(x^k) + \rho_k P(x^k) > f(z) + \rho_k P(z) + (\rho_k c + f(x^k) - f(z))$. Assim para $k$ suficientemente grande temos $(\rho_k c + f(x^k) - f(z)) > 0$ e portanto $f(x^k) + \rho_k P(x^k) > f(z) + \rho_k P(z)$, o que contradiz a definição de $x^k$.

Assumindo que o problema admite um ponto viável $z \in \bar{\Omega}, P(z) = 0$, temos que $x^*$ é um ponto viável. Mas neste caso podemos mostrar ainda mais. Podemos mostrar que $x^*$ é uma solução global. Já que para todo $z \in \bar{\Omega}$, temos que $f(x^k) \leq f(x^k) + \rho_k P(x^k) \leq f(z) + \rho_k P(z)$. Em particular, para todo $z$, tal que $P(z) = 0$. Assim, para $z$ viável temos $f(x^k) \leq f(z)$. Tomando o limite para uma subsequência apropriada segue que $f(x^*) \leq f(z)$, e portanto todos os pontos limites, caso existam, são soluções (globais) do problema quando a região viável é não vazia. □

Com o teorema anterior à disposição, provaremos que a condição AKKT é uma condição de otimalidade.

**Teorema 2.7.** *Se $x^*$ é uma solução local, então $x^*$ satisfaz AKKT.*

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

$$\begin{array}{ll} \text{Minimizar} & f(x) + \frac{1}{2}\|x - x^*\|^2, \\ \text{sujeito a} & h_i(x) = 0, i = 1, \ldots, m, \\ & g_j(x) \leq 0, j = 1, \ldots, p, \\ & \|x - x^*\|^2 \leq \delta, \end{array}$$

onde $\delta > 0$ é um parâmetro suficientemente pequeno dado pela definição de minimizador local. Defina $\bar{\Omega} := \{x \in \mathbb{R}^n \mid \|x - x^*\|^2 \leq \delta\}$. Claramente, $\bar{\Omega}$ é um conjunto não vazio, limitado e fechado (i.e. compacto).

Note que $x^*$ é a única solução global. Pela compacidade de $\bar{\Omega}$, a sequência $\{x^k\}$ gerada pelo algoritmo de penalidade externa está bem definida, além disso, sendo o conjunto viável não vazio, como consequência do Teorema 2.6, todos seus pontos limites são soluções globais. Da compacidade de $\bar{\Omega}$ temos a existência de pelo menos um ponto limite, e da unicidade da solução, temos que há um único ponto limite, a saber, $x^*$. Portanto, a sequência $\{x^k\}$ é uma sequência convergente com limite $x^*$.

Para mostrar que $\{x^k\}$ é uma sequência AKKT, basta observar que $\|x^k - x^*\| < \delta$ para $k$ suficientemente grande, e portanto, usando a definição da função $P(x)$ e o fato que o gradiente de $f(x) + \frac{1}{2}\|x - x^*\|^2 + \rho_k P(x)$ se anula em $x = x^k$, temos

$$\nabla f(x^k) + x^k - x^* + \sum_{i=1}^{m}(2\rho_k h_i(x^k))\nabla h_i(x^k) + \sum_{j=1}^{p}(2\rho_k \max\{0, g_j(x^k)\})\nabla g_j(x^k) = 0.$$

Defina $\lambda_i^k = 2\rho_k h_i(x^k)$, para todo $i = 1, \ldots, m$ e $\mu_j^k = 2\rho_k \max\{0, g_j(x^k)\} \geq 0$, para todo $j = 1, \ldots, p$. Ainda mais, temos que para $j$ tal que $g_j(x^*) < 0$, se cumpre que $\mu_j^k = 0$ para $k$ suficientemente grande.

Assim, obtemos

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

Tomando o limite temos que $x^*$ satisfaz AKKT. □

**Observação** A existência de multiplicadores Fritz John (Teorema 1.2) pode ser provada como uma consequência do Teorema 2.7. De fato, sendo $x^*$ uma solução local, considere sequências $\{x^k\}$, $\{(\lambda^k, \mu^k)\} \subset \mathbb{R}^m \times \mathbb{R}_+^p$ com $\mu_j^k = 0$, $j \notin A(x^*)$ como na Definição 2.12 de AKKT, de modo que $\varepsilon_k := \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \to 0$. Dividindo $\varepsilon_k$ por $\|(1, \lambda^k, \mu^k)\|$, e já que $\|\|(1, \lambda^k, \mu^k)\|^{-1}\varepsilon_k\| \leq \|\varepsilon_k\|$ temos que

$$\alpha_0^k \nabla f_0(x^k) + \sum_{i=1}^{m}\lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{p}\mu_j^k \nabla g_j(x^k) \to 0$$

onde $\alpha_0^k := \|(1, \lambda^k, \mu^k)\|^{-1}$, $\alpha^k := \|(1, \lambda^k, \mu^k)\|^{-1}\lambda^k$, $\beta^k := \|(1, \lambda^k, \mu^k)\|^{-1}\mu^k$ e $\|(\alpha_0^k, \alpha^k, \beta^k)\| = 1$. Tomando o limite em uma subsequência tal que $\{(\alpha_0^k, \alpha^k, \beta^k)\}$ é convergente, temos que tal limite é um multiplicador que satisfaz as condições do Teorema 1.2.

Dada a equivalência da condição de Fritz John com a condição de otimalidade "KKT ou não-MFCQ", a demonstração anterior também prova que um ponto AKKT que satisfaz MFCQ deve ser um ponto KKT. Os Teoremas 2.14 e 2.16 generalizam este resultado.

Existem várias razões para a popularidade da condição AKKT, não somente pela sua simplicidade, mas também devido a sua associação com muitos algoritmos práticos (ver [10]). Entre os diferentes métodos mencionamos o método de Lagrangiano aumentado [4,31], alguns métodos de programação quadrática sequencial [71], alguns métodos de pontos interiores [37] e métodos de restauração inexata [63]. Outra importante característica é a capacidade de provar convergência global destes métodos sob condições de qualificação mais fracas que as usuais. Por outro lado, existem outros algoritmos práticos que não geram sequências AKKT, como o caso do método de Newton-Lagrange, [13].

A seguir, vamos mostrar que o método de Lagrangiano aumentado definido no Algoritmo 2.1 gera sequências AKKT. O método de Lagrangiano aumentado é um dos métodos mais populares para resolver problemas de otimização, amplamente estudado na literatura desde sua introdução por Hestenes e Powell. Ver [24, 28, 29, 31, 44, 52, 69, 70] e as referências ali mencionadas,

O método de Lagrangiano aumentado está baseado no método de penalidade, mas as estimativas dos multiplicadores de Lagrange são usadas para evitar o mal-condicionamento inerente aos métodos de penalidade quadrática, uma vez que o parâmetro de penalidade deve ir a infinito. O método de Lagrangiano aumentado (ou *método dos multiplicadores*) serve como base para implementações de alta qualidade como LANCELOT, PENOPT ou ALGENCAN.

Dado $\rho > 0$ e vetores $(\lambda, \mu) \subset \mathbb{R}^m \times \mathbb{R}^p, \mu \geq 0$, considere a função Lagrangiano aumentado:

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

O método é definido da seguinte maneira:

**Algoritmo 2.1 (Método de Lagrangiano aumentado).** *Sejam* $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, $\gamma > 1$, $\rho_1 > 0$ *e* $\tau \in (0, 1)$. *Tome* $\{\varepsilon_k\} \subset \mathbb{R}_+$ *uma sequência de escalares positivos tal que* $\lim \varepsilon_k = 0$.
*Defina* $\lambda_i^1 \in [\lambda_{\min}, \lambda_{\max}], i \in \{1, \ldots, m\}$ *e* $\mu_j^1 \in [0, \mu_{\max}], j \in \{1, \ldots, p\}$. *Escolha* $x^0 \in \mathbb{R}^n$ *como ponto inicial arbitrário. Defina* $V^0 = \max\{0, g(x^0)\}$. *Inicialize com* $k = 1$.

1. *(Minimização)*
   *Encontre um minimizador aproximado* $x^k$ *de* $L_{\rho_k}(x, \lambda^k, \mu^k)$. *Isto é, determine* $x^k$ *que satisfaça:*
   $$\|\nabla L_{\rho_k}(x^k, \lambda^k, \mu^k)\| \leq \varepsilon_k$$
2. *(Atualização do parâmetro de penalidade)*
   *Defina para todo* $j \in \{1, \ldots, p\}$
   $$V_j^k := \max\{g_j(x^k), -\mu_j^k/\rho_k\}.$$
   *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*
   $\hat{\lambda}_i^k := \lambda_i^k + \rho_k h_i(x^k)$, $i = 1, \ldots, m$ *e* $\hat{\mu}_j^k := \max\{0, \mu_j^k + \rho_k g_j(x^k)\}$, $j = 1, \ldots, p$.
   *Defina* $\lambda_i^{k+1} := proj_{[\lambda_{\min}, \lambda_{\max}]}(\hat{\lambda}_i^k) \in [\lambda_{\min}, \lambda_{\max}]$ *para todo* $i \in \{1, \ldots, m\}$ *e* $\mu_j^{k+1} := proj_{[0, \mu_{\max}]}(\hat{\mu}_j^k) \in [0, \mu_{\max}]$, *para todo* $j \in \{1, \ldots, p\}$. *Faça* $k \leftarrow k + 1$ *e vá para o Passo 1.*

A seguir, estudaremos o comportamento da sequência $\{x^k\}$ gerada pelo método de Lagrangiano aumentado.

**Teorema 2.8.** *Seja $x^*$ um ponto limite da sequência $\{x^k\}$ gerada pelo Algoritmo 2.1. Então $x^*$ é um ponto estacionário para o problema de otimização:*

$$\textit{Minimizar} \quad \tfrac{1}{2}\textstyle\sum_{i=1}^{m} h_i(x)^2 + \sum_{j=1}^{p} \max\{0, g_j(x)\}^2$$

*Demonstração.* Se $\{\rho_k\}$ tem alguma subsequência limitada, temos que $V^k \to 0$ e $x^*$ é um ponto viável. Assim, $h_i(x^*) = 0$, $i = 1, \ldots, m$; $g_j(x^*) \leq 0$, $j = 1, \ldots, p$ e portanto solução do problema Minimizar $\frac{1}{2}\sum_{i=1}^{m} h_i(x)^2 + \sum_{j=1}^{p} \max\{0, g_j(x)\}^2$.

Se toda subsequência de $\{\rho_k\}$ é ilimitada. Neste caso, defina

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

onde $\hat{\lambda}_i^k := \lambda_i^k + \rho_k h_i(x^k)$, $i = 1, \ldots, m$ e $\hat{\mu}_j^k := \max\{0, \mu_j^k + \rho_k g_j(x^k)\}$, $j = 1, \ldots, p$. Claramente do Passo 1 do Algoritmo 2.1, temos que $\|\delta_k\| \leq \varepsilon_k$. Divida $\delta_k$ por $\rho_k$. Como $\{(\lambda^k, \mu^k)\}$ é limitada, temos que $\hat{\lambda}_i^k/\rho_k$ converge para $h_i(x^*)$, quando $i = 1, \ldots, m$ e $\hat{\mu}_j^k/\rho_k$ converge para $\max\{0, g_j(x^*)\}$ quando $j = 1, \ldots, p$. Assim, o resultado segue. □

**Teorema 2.9.** *Suponha que a sequência $\{x^k\}$ gerada pelo algoritmo de Lagrangiano aumentado admite um ponto limite viável $x^*$. Então $x^*$ é um ponto AKKT.*

*Demonstração.* Pela definição de $L_\rho$ e pelo Passo 1 do Algoritmo 2.1 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 $\hat{\lambda}_i^k := \lambda_i^k + \rho_k h_i(x^k)$, $i = 1, \ldots, m$; $\hat{\mu}_j^k := \max\{0, \mu_j^k + \rho_k g_j(x^k)\}$, $j = 1, \ldots, p$. Agora provaremos que $\hat{\mu}_j^k = 0$ para $j \notin A(x^*)$ e para todo $k$ suficientemente grande. Suponha que $g_{j_0}(x^*) < 0$ para algum $j_0 = 1, \ldots, p$. Temos dois casos: (i) Se $\rho_k \to +\infty$. Neste caso, já que a sequência $\{\mu_{j_0}^k\}$ é limitada, concluímos que $\hat{\mu}_{j_0}^k = 0$. (ii) Se $\rho^k$ é limitada. Neste caso, $V_{j_0}^k = \max\{g_{j_0}(x^k), -(\rho^k)^{-1}\mu_{j_0}^k\} \to 0$ e portanto $-\mu_{j_0}^k/\rho^k \to 0$. Logo $\hat{\mu}_j^k = \rho^k \max\{0, g_{j_0}(x^k) + \mu_{j_0}^k/\rho^k\} = 0$. Segue de ambos os casos, que $x^*$ satisfaz a condição AKKT. □

A condição AKKT é uma condição de otimalidade não trivial. De fato, AKKT implica "KKT ou não-CQ", para condições de qualificação mais fracas que LICQ ou MFCQ, como por exemplo a CPLD ou CRSC. Desta forma "KKT ou não-CRSC" fornece uma medida de força da condição AKKT como critério de parada. Na próxima seção iremos mostrar uma medida de força mais precisa para a condição AKKT.

### 2.2.2 Condição Gradiente Projetado Aproximado (AGP)

Na procura por condições de otimalidade que encaixem bem no comportamento de métodos práticos, Martínez e Svaiter, em [64], introduziram uma nova condição de otimalidade baseada nos gradientes projetados. Esta condição é chamada de AGP (*Approximate Gradient Projection*). A condição AGP é uma condição de otimalidade que encaixa naturalmente como critério de parada para os métodos de restauração inexata, [32, 35, 43, 62, 63].

**Definição 2.13.** *Dado um número não positivo* $\gamma \in [-\infty, 0]$ *e um ponto viável* $x^* \in \Omega$. *Dizemos que* $x^*$ *satisfaz a condição AGP(*$\gamma$*) se existe uma sequência* $\{x^k\}$ *com limite* $x^*$ *tal que*

$$P_{\Omega(x^k,\gamma)}(x^k - \nabla f(x^k)) - x^k \to 0, \tag{2.2.4}$$

*onde* $P_{\Omega(x^k,\gamma)}$ *é a projeção ortogonal sobre o conjunto convexo fechado* $\Omega(x^k,\gamma)$. *O conjunto* $\Omega(x^k,\gamma)$ *é definido como o conjunto dos* $z \in \mathbb{R}^n$ *tal que*

$$\left\{ \begin{array}{l} \nabla h_i(x^k)^{\mathtt{T}}(z - x^k) = 0, \ \textit{para todo } i \in \{1,\ldots,m\} \\ \nabla g_j(x^k)^{\mathtt{T}}(z - x^k) \le 0, \ \textit{se } 0 \le g_j(x^k), j = 1,\ldots,p \\ g_j(x^k) + \nabla g_j(x^k)^{\mathtt{T}}(z - x^k) \le 0, \ \textit{se } \gamma < g_j(x^k) < 0, j = 1,\ldots,p \end{array} \right\}. \tag{2.2.5}$$

A sequência $\{x^k\}$ é chamada de sequência AGP($\gamma$) associada a $x^*$. O parâmetro $\gamma$ serve para decidir, dependendo da viabilidade de $x^k$, quando levar em consideração uma determinada restrição na construção do conjunto linearizado. A ideia é que se uma restrição está satisfeita "com folgas", é provável que esta restrição continuará a ser satisfeita automaticamente numa próxima iteração, podendo ser descartada.

A validade da condição AGP($\gamma$) é independente do parâmetro $\gamma$ sempre que $\gamma \in [-\infty, 0)$, isto é, se a condição AGP($\gamma$) vale em $x^* \in \Omega$ para algum $\gamma \in [-\infty, 0)$ então AGP($\gamma^{'}$) também vale em $x^*$ para qualquer $\gamma^{'} \in [-\infty, 0)$, ver [64], por esse motivo escrevemos simplesmente AGP em vez de AGP($\gamma$).

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

$$\left\{ z \in \mathbb{R}^n : \begin{array}{ll} h_i(z) = h_i(x^k), & \text{para todo } i \in \{1,\ldots,m\} \\ g_j(z) \le g_j(x^k), & \text{se } 0 \le g_j(x^k), j = 1,\ldots,p \\ g_j(z) \le 0, & \text{se } \gamma < g_j(x^k) < 0, j = 1,\ldots,p \end{array} \right\}. \tag{2.2.6}$$

O conjunto (2.2.6) é interpretado como o conjunto de pontos $z \in \mathbb{R}^n$ que são tão inviáveis quanto o ponto atual $x^k$.

Para entender melhor a condição AGP, falaremos um pouco acerca dos métodos de restauração inexata. Os métodos de restauração inexata baseiam-se na ideia que em cada iteração, a viabilidade e a otimalidade são abordadas em diferentes fases sem deteriorar muito o avanço obtido em cada fase anterior. Na fase de restauração, o método visa em melhorar a viabilidade enquanto na fase de otimalidade o objetivo é melhorar a otimalidade, possivelmente deteriorando um pouco a viabilidade obtida. Usualmente, para algoritmos práticos de restauração inexata, a fase de viabilidade é realizada por algum procedimento relacionado com as características próprias do problema e a fase de otimalidade está relacionada com a direção de busca

$$d_\gamma(x^k) := P_{\Omega(x^k,\gamma)}(x^k - \nabla f(x^k)) - x^k$$

que não deteriora muito a viabilidade obtida devido a que para todo $t \in [0,1]$, $x^k + td_\gamma(x^k)$ pertence a $\Omega(x^k,\gamma)$, a aproximação linear de (2.2.6).

Observe que se o gradiente projetado é zero ($d_\gamma(x) = 0$), então as condições KKT são satisfeitas em $x$. Como mencionamos anteriormente, cada condição sequencial tem associado um critério natural de parada. Esse critério relaciona diretamente a condição sequencial com os algoritmos. Suponha que $\varepsilon_1$ e $\varepsilon_2$ são duas tolerâncias ao erro, referentes à viabilidade e à otimalidade, respectivamente. O critério natural de parada associado à condição AGP é declarar convergência em $x$ quando

$$\|(h_i(x))_{i=1}^m\| \le \varepsilon_1, \quad \|(\max\{0, g_j(x)\})_{j=1}^p\| \le \varepsilon_1 \tag{2.2.7}$$

e

$$\|x - P_{\Omega(x,\gamma)}(x - \nabla f(x))\| \leq \varepsilon_2. \tag{2.2.8}$$

Perceba que no critério associado à condição AGP, não é necessário nenhum cálculo referente aos multiplicadores de Lagrange, como no caso da condição AKKT.

O Teorema 2.10 mostra que a condição AGP é de fato uma genuína condição de otimalidade, [64, Teorema 2.1].

**Teorema 2.10.** *Seja $x^* \in \Omega$ um minimizador local e $\gamma \in [-\infty, 0]$. Então a condição AGP($\gamma$) é satisfeita em $x^*$*

*Demonstração.* Seja $x^*$ um minimizador local, $\{\rho_k\}$ uma sequência de números positivos tal que $\rho_k \to \infty$ e $\gamma \in [-\infty, 0]$. Seguindo as mesmas linhas da demonstração do Teorema 2.7, temos que existe uma sequência $\{x^k\}$ com $x^k \to x^*$ tal que

$$\nabla f(x^k) + \sum_{i=1}^{m} \rho_k h_i(x^k) \nabla h_i(x^k) + \sum_{j:g_j(x^k)\geq 0} \rho_k \max\{0, g_j(x^k)\} \nabla g_j(x^k) = -\delta^k, \tag{2.2.9}$$

onde $\delta^k := \|x^k - x^*\|(x^k - x^*) \to 0$. Denote $\lambda_i^k := \rho_k h_i(x^k)$, para $i = 1, \ldots, m$ e $\mu_j^k := \rho_k \max\{0, g_j(x^k)\}$ para $j = 1, \ldots, p$.

Adicionando e subtraindo $x^k$ adequadamente no lado esquerdo de (2.2.9), tomando a projeção sob o conjunto $\Omega(x^k, \gamma)$ e usando a não expansividade da projeção segue

$$\|P_{\Omega(x^k,\gamma)}(x^k - \nabla f(x^k)) - P_{\Omega(x^k,\gamma)}(x^k + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j:g_j(x^k)\geq 0} \mu_j^k \nabla g_j(x^k))\| \leq \|\delta^k\|. \tag{2.2.10}$$

Como $\Omega(x^k, \gamma)$ é definido através de restrições lineares, o cone normal de $\Omega(x^k, \gamma)$ em $x = x^k$, $N_{\Omega(x^k,\gamma)}(x^k)$ (veja (2.3.25) mais a frente), é facilmente calculado. De fato, isso é consequência do Lema de Farkas. Assim, vemos que

$$\sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j:g_j(x^k)\geq 0} \mu_j^k \nabla g_j(x^k) \in N_{\Omega(x^k,\gamma)}(x^k). \tag{2.2.11}$$

Concluímos que: (veja Proposição 2.3 na próxima seção)

$$P_{\Omega(x^k,\gamma)}(x^k + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j:g_j(x^k)\geq 0} \mu_j^k \nabla g_j(x^k)) = x^k. \tag{2.2.12}$$

Substituindo (2.2.12) em (2.2.10) temos $\|P_{\Omega(x^k,\gamma)}(x^k - \nabla f(x^k)) - x^k\| \leq \|\delta^k\|$, tomando limite na última expressão a condição AGP($\gamma$) vale. □

Já que ambas condições AGP e AKKT são condições de otimalidade, as relações de implicação entre elas são interessante. O próximo teorema mostra que a condição AGP implica a condição AKKT.

**Teorema 2.11.** *A condição AGP implica a condição AKKT.*

*Demonstração.* Suponha que a condição AGP vale em $x^*$, então pela definição existe uma sequência $\{x^k\}$ tal que $y^k := P_{\Omega(x^k,\gamma)}(x^k - \nabla f(x^k)) - x^k \to 0$ 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:

$$\text{Minimizar } \frac{1}{2}\|z - x^k + \nabla f(x^k)\|^2 \quad \text{sujeito a } z \in \Omega(x^k, \gamma). \tag{2.2.13}$$

Desde que $\Omega(x^k, \gamma)$ é dado por restrições lineares, as condições KKT valem em $y^k + x^k$. Assim, existem multiplicadores $\lambda \in \mathbb{R}^m$ e $\mu \in \mathbb{R}^p_+$ tais que

$$\nabla f(x^k) + y^k + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) = 0, \tag{2.2.14}$$

e satisfazem as seguintes relações de complementaridade, para $j = 1, \ldots, p$:

$$\mu_j^k[g_j(x^k) + \langle \nabla g_j(x^k), y^k \rangle] = 0, \qquad \text{se } \gamma < g_j(x^k) < 0 \tag{2.2.15}$$

$$\mu_j^k[\langle \nabla g_j(x^k), y^k \rangle] = 0, \qquad \text{se } 0 \le g_j(x^k) \tag{2.2.16}$$

$$\mu_j^k = 0, \qquad \text{caso contrário.} \tag{2.2.17}$$

Da expressão (2.2.14) e de $y^k \to 0$, temos que $\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \to 0$. Para provar que a condição AKKT vale, só falta demonstrar que $\mu_j^k = 0$ para $j \notin A(x^*)$. Seja $j \notin A(x^*)$, logo $g_j(x^*) < 0$. Da continuidade temos que para $k$ suficientemente grande que $g_j(x^k) + \langle \nabla g_j(x^k), y^k \rangle < 0$ e como consequência de (2.2.15), $\mu_j^k$ deve ser zero para todo $j \notin A(x^*)$. □

Surpreendentemente a condição sequencial AGP é estritamente mais forte que a condição AKKT, como mostra o próximo contra-exemplo.

**Contra-Exemplo 2.1.** *(A condição AKKT não implica AGP, [7])*

Em $\mathbb{R}^2$, consideremos $x^* = (0, 1)$ e o problema de otimização não linear

$$\text{Minimizar } f(x_1, x_2) = -x_2 \text{ s.a. } h(x_1, x_2) := x_1 x_2 = 0, \quad g(x_1, x_2) := -x_1 \le 0.$$

Nesse problema de otimização, a condição AGP não vale em $x^* = (0, 1)$ mas vale a condição AKKT. Podemos medir a força de condição AGP através das proposições que ela implica. De fato, a condição sequencial AGP é forte no sentido que implica "KKT ou não-CPLD", [51]. A expressão "KKT ou não-CPLD" pode ser considerado como uma medida da força da condição AGP. Na próxima subseção mostraremos a medida mais precisa da força da condição AGP como critério de parada para os algoritmos práticos.

No caso convexo temos o seguinte resultado, [64, Teorema 3.1], que diz que a condição AGP é uma condição suficiente para otimalidade.

**Teorema 2.12.** *[64, Teorema 3.1].*
*Suponha que as funções $f$, $g_j$, $j = 1, \ldots, p$ são convexas e que as funções $h_i, i = 1, \ldots, m$ são funções afins. Seja $\gamma \in [-\infty, 0]$. Suponha que $x^* \in \Omega$ e que $\{x^k\}$ é uma sequência AGP($\gamma$) para $x^*$, com a propriedade adicional $h_i(x^k) = 0, i = 1, \ldots, m$ para todo $k \in \mathbb{N}$. Então $x^*$ é um minimizador global.*

## 2.3 CQs associadas às condições sequenciais

Nesta seção, nosso principal interesse é o estudo de condições de qualificação, de preferência fracas, que possam ser usadas na análise de convergência de vários métodos numéricos. Lembre que condições de qualificação fracas geram condições de otimalidade fortes da forma "KKT ou não-CQ".

Estes tipos de condições de qualificação são importantes e tem consequências práticas relevantes. Consideremos como paradigma deste tipo de condição de qualificação a condição CPLD (dependência linear positiva constante, em inglês) introduzida por Qi e Wei em [71]. Qi e Wi apresentaram um método particular de programação quadrática sequencial com convergência a pontos limites que satisfazem a condição "KKT ou não-CPLD". Então sob a condição CPLD, este método de programação quadrática sequencial tem convergência a pontos KKT. O fato de CPLD ser uma condição de qualificação foi provado posteriormente por Andreani, Martínez e Schuverdt [15], que ainda mostraram as relações existentes entre as outras condições de qualificação conhecidas na época. Em artigos posteriores, Andreani, Birgin, Martínez e Schuverdt, [3, 4], provaram que a condição de qualificação CPLD pode também ser usada para a análise de convergência do método de lagrangiano aumentado, um avanço significativo pois as provas anteriores eram baseadas na condição de regularidade (LICQ), claramente mais restritiva.

A procura por condições de qualificação que possam ser usadas na análise de convergência de métodos numéricos motivou uma série de artigos [9–11] dando origem às condições de qualificação RCPLD, CRSC, CPG e CCP.

Para entender melhor a relação entre condições de qualificação e algoritmos, começaremos analisando a condição sequencial de otimalidade AKKT.

Considere uma proposição $\mathcal{P}$ que depende somente da descrição analítica das restrições tal que

$$\text{AKKT} + \mathcal{P} \Longrightarrow \text{KKT}. \tag{2.3.18}$$

O que podemos dizer de $\mathcal{P}$? Primeiro, devido ao fato que AKKT é uma condição de otimalidade, a condição $\mathcal{P}$ é uma condição de qualificação. Ainda mais, se temos algum algoritmo prático que gera sequências cujos pontos limites satisfazem a condição AKKT temos que, sob $\mathcal{P}$, o algoritmo gera sequências cujos pontos limites são pontos KKT. O interessante é que, de fato, existem algoritmos práticos que naturalmente geram sequências cujos pontos limites viáveis são pontos AKKT. Assim, podemos dizer que a condição $\mathcal{P}$ é uma condição de qualificação com claras implicações algorítmicas.

Uma pergunta natural é a seguinte: Dado que a condição AKKT é uma genuína condição de otimalidade obviamente associada com o critério de parada de algoritmos práticos, por quê devemos nos preocupar com condições de qualificações nos quais AKKT implica KKT? O motivo é que para famílias de problemas importantes, a condição AKKT não é suficientemente forte para fornecer uma confiável detecção da otimalidade. O exemplo clássico é uma versão simplificada dos MPCC (*mathematical programs with complementarity constraints*). Consideremos o problema de minimizar $x_2$ sujeito a $x_1 \geq 0$ e $x_1x_2 = 0$. O ponto viável $x^* = (0, 1)$ é um ponto AKKT (basta considerar a sequência $x_1^k = 1/k$, $x_2^k = 1$ e multiplicadores $\lambda_1^k = \lambda_2^k = k$) mas o ponto $x^*$ não tem nenhuma relação com o problema a minimizar, ele não é nem minimizador nem um ponto KKT. Mas passaria qualquer critério de parada baseado na condição AKKT. Certamente, isso não aconteceria se no ponto $x^* \in \Omega$ valesse qualquer condição $\mathcal{P}$ satisfazendo (2.3.18).

Tipicamente, a convergência global de um algoritmo é dada mostrando que sob uma condição de qualificação, os pontos limites gerados pelo algoritmo são pontos estacionários. Resta saber qual é a relação entre a condição de otimalidade sequencial AKKT e condições de otimalidade "pontuais" do tipo "KKT ou não-CQ". Vamos mostrar que AKKT é mais forte que "KKT ou não-CPG" onde a condição de qualificação *constant positive generators* (CPG, [10]) é mais fraca que CRSC. Em particular, algoritmos que geram sequências AKKT tem pontos limites estacionários exigindo CPG ou qualquer outra condição de qualificação mais forte como CRSC, (R)CPLD, MFCQ, etc.

Para definir a condição CPG, considere vetores $\mathcal{V} = \{v_1, \dots, v_{m+r}\} \in \mathbb{R}^n$ e vamos estudar cones da forma

$$\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}) = \{w \in \mathbb{R}^n \mid w = \sum_{i \in \mathcal{I} \cup \mathcal{J}} \lambda_i v_i, \lambda_i \geq 0, i \in \mathcal{J}\},$$

onde $\mathcal{I} = \{1, \dots, m\}$, $\mathcal{J} = \{m+1, \dots, m+r\}$ e os vetores $v_1, \dots, v_{m+r}$ estarão claros do contexto. Note que quando $v_i := \nabla h_i(x)$, para $i = 1, \dots, m$ e $\{v_{m+1}, \dots, v_{m+r}\} := \{\nabla g_i(x), i \in A(x)\}$, temos $\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V})$ coincide com $\mathcal{L}(x)^\circ$.

Para cones desta forma (cones positivamente gerados), podemos definir uma noção de base. Diremos que o par ordenado $(\mathcal{I}', \mathcal{J}')$, $\mathcal{I}', \mathcal{J}' \subset \{1, \dots, m + r\}$ 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 $v_i, i \in \mathcal{J}'$ são positivo-linearmente independentes. Mostramos a seguir a existência de uma base. Para isso, note que vetores originalmente em $\mathcal{J}$ podem estar em $\mathcal{I}'$. No nosso contexto, isto representa uma restrição de desigualdade que será considerada como uma restrição de igualdade. Se os vetores $v_i, i \in \mathcal{I}$ e $v_i, i \in \mathcal{J}$ são positivo-linearmente dependentes, isto é, existe $0 \neq \alpha \in \mathbb{R}^{m+r}$ tal $\sum_{i=1}^{m+r} \alpha_i v_i = 0, \alpha_i \geq 0, i = m+1, \dots, m+r$, temos duas possibilidades: se $v_i, i \in \mathcal{I}$ é linearmente dependente, considere $\mathcal{I}'$ obtido de $\mathcal{I}$ removendo um índice redundante. Claramente temos $\text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}) = \text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V})$ com $\mathcal{J}' = \mathcal{J}$. Se $v_i, i \in \mathcal{I}$ é linearmente independente, então existe $j \in \mathcal{J}$ tal que $-v_j \in \text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V})$ e definindo $\mathcal{I}' = \mathcal{I} \cup \{j\}$ e $\mathcal{J}' = \mathcal{J} - \{j\}$ é fácil ver que $\text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}) = \text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V})$. Repetindo esta construção até obter vetores $v_i, i \in \mathcal{I}'$ e $v_i, i \in \mathcal{J}'$ positivo-linearmente independentes obtemos uma base para $\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V})$.

A seguir procederemos a definir a condição CPG.

**Definição 2.14.** *(CPG) Considere* $\mathcal{V}(x) = \{\nabla h_i(x)\}_{i=1}^m \cup \{\nabla g_i(x)\}_{i \in A(x^*)}$, $\mathcal{I} = \{1, \dots, m\}$ *e* $\mathcal{J} = A(x^*)$. *Dizemos que a condição CPG vale em* $x^* \in \Omega$ *quando existe uma base* $(\mathcal{I}', \mathcal{J}')$ *de* $\text{span}_+(\mathcal{I}, \mathcal{J}, \mathcal{V}(x^*)) = \mathcal{L}(x^*)^\circ$ *que é suficiente para gerar o cone em uma vizinhança, isto é,*

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

*para todo* $x$ *em alguma vizinhança de* $x^*$.

**Observação:** Considerando a decomposição de $\mathcal{L}(x^*)^\circ$ na soma direta de seu maior subespaço, o subespaço componente,

$$\{v \in \mathbb{R}^n \mid v = \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{i \in J_-(x^*)} \mu_i \nabla g_i(x^*), \lambda_i, \mu_i \in \mathbb{R}\}$$

com o cone pontudo

$$\{v \mid v = \sum_{i \notin J_-(x^*)} \mu_i \nabla g_i(x^*), \mu_i \geq 0, \forall i\},$$

podemos considerar $\mathcal{I}' \subset \{1,\ldots,m\} \cup J_-(x)$ tal que $\{v_i, i \in \mathcal{I}'\}$ forma uma base para a componente do subespaço, enquanto necessariamente $\mathcal{J}' = \{i \mid i \notin J_-(x)\}$.

**Contra-Exemplo 2.2.** *(CPG não implica CRSC)*
*Considere em $\mathbb{R}^2$, o ponto $x = (0,0)$ e a região viável definida por*

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

Em $x = (0,0)$, temos que $J_-(x) = \{1,2\}$. Escolha $\mathcal{I}' = \{1\}$ e $\mathcal{J}' = \{3\}$. Então, temos que $\nabla g_1(x)$ e $\nabla g_3(x)$ são positivo-linearmente independentes tais que $\text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}(x)) = \text{span}_+(\emptyset, \{1,2,3\}; \mathcal{V}(x))$, e ainda, para todo $y \neq 0$, $\text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}(y))$ é o semi-espaço que contém propriamente o cone pontudo $\text{span}_+(\emptyset, \{1,2,3\}; \mathcal{V}(y))$. Portanto vale CPG, entretanto, CRSC não vale, já que o posto de $\{\nabla g_1(0), \nabla g_2(0)\}$ é 1, enquanto para qualquer $y \neq 0$ o posto aumenta.

O Teorema a seguir mostra que CPG é de fato mais fraca que CRSC.

**Teorema 2.13.** *Se $x \in \Omega$ satisfaz CRSC, então $x$ satisfaz CPG.*

*Demonstração.* Seja $(\mathcal{I}', \mathcal{J}')$ 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 que o posto do conjunto $\{\nabla f_i(y), i \in J_-(x)\}$ é constante . $\square$

Incluindo a restrição $g_4(x_1, x_2) := x_2^3 \leq 0$ no exemplo analisado, observamos que a condição Error Bound não é satisfeita. Para ver isto, basta considerar $x^k = (-(1/k)^{1/3}, 1/k)$. A distância de $x^k$ para o conjunto viável é $1/k$, enquanto a medida de inviabilidade é $1/k^3$. Embora CPG não cumpra com Error Bound, temos que CPG implica Abadie. Ver [10].

Mostraremos a seguir que sob CPG, um ponto AKKT é de fato KKT. Como todo minimizador é AKKT, isso mostra que CPG é de fato uma condição de qualificação, e além disso, garante a convergência global para algoritmos que geram sequência AKKT sob CPG.

**Teorema 2.14.** *Se $x \in \Omega$ é AKKT e satisfaz CPG, então $x$ cumpre KKT.*

*Demonstração.* Da definição de AKKT, existem $\{x^k\} \subset \mathbb{R}^n$, $x^k \to x$ e $\{\lambda^k\} \subset \mathbb{R}^m, \{\mu^k\} \subset \mathbb{R}^p$ com $\mu_i^k \geq 0, i \in A(x)$ tal que

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{i \in A(x)} \mu_i^k \nabla g_i(x^k) \to 0.$$

Como a condição CPG vale, existe $(\mathcal{I}', \mathcal{J}')$ uma base para $\mathcal{L}(x)^\circ$ tal que

$$\underset{+}{\text{span}}(\{1,\ldots,m\}, A(x); \mathcal{V}(y)) \subset \underset{+}{\text{span}}(\mathcal{I}', \mathcal{J}'; \mathcal{V}(y))$$

para $y$ em uma vizinhança de $x$. Logo para $k$ suficientemente grande, existem $\bar{\lambda}_i^k, i \in \mathcal{I}' \cup \mathcal{J}'$ com $\bar{\lambda}_i^k \geq 0, i \in \mathcal{J}'$ tais que $\nabla f(x^k) + \sum_{i \in \mathcal{I}' \cup \mathcal{J}'} \lambda_i^k \nabla f_i(x^k) \to 0$, onde $\nabla f_i := \nabla h_i, i = 1,\ldots,m$ e $\{\nabla f_{m+1}, \ldots, f_{m+r}\} := \{\nabla g_i : i \in A(x)\}$.

Defina $M_k = \max\{|\bar{\lambda}_i^k|, i \in \mathcal{I}' \cup \mathcal{J}'\}$. Se $\{M_k\}$ é ilimitada, dividindo por $M_k$ podemos tomar uma subsequência tal que $M_k^{-1} \bar{\lambda}_i^k \to \alpha_i$ com $\alpha_i$ não todos nulos e $\alpha_i \geq 0$, $i \in \mathcal{J}'$. Assim, tomando o limite temos $\sum_{i \in \mathcal{I}' \cup \mathcal{J}'} \alpha_i \nabla f_i(x) = 0$ o que contradiz a independência linear positiva. No caso que $\{M_k\}$ é limitada, podemos tomar uma subsequência tal que $\bar{\lambda}_i^k \to \lambda_i$, $\lambda_i \geq 0, i \in \mathcal{J}'$. Tomando o limite temos $\nabla f(x) + \sum_{i \in \mathcal{I}' \cup \mathcal{J}'} \lambda_i \nabla f_i(x) = 0$. Sendo $(\mathcal{I}', \mathcal{J}')$ base para $\mathcal{L}(x)^\circ$ temos que $x$ é KKT. $\square$

Acreditava-se que CPG era a condição de qualificação mais fraca possível que garante que AKKT implica KKT. Entretanto, em [11, 72], é definida a condição de qualificação CCP (*cone continuity property*) com tal propriedade. CCP é estritamente mais fraca que CPG, embora CCP implique Abadie.

A condição de qualificação CCP desempenha para a condição sequencial AKKT o mesmo papel que a condição de Guignard desempenha para a otimalidade. A definição de CCP requer várias ferramentas da análise variacional. Assim, procederemos a descrever alguns resultados básicos de análise variacional.

### 2.3.1 Resultados básicos de otimização e análise variacional.

Nesta subseção, começaremos revisando os objetos geométricos mais importantes em otimização e na análise variacional [66, 74].

Dizemos que $F : \mathbb{R}^s \rightrightarrows \mathbb{R}^d$ é uma multifunção se para cada $x \in \mathbb{R}^s$, $F(x)$ é um subconjunto de $\mathbb{R}^d$. Dependendo da situação, $F(x)$ pode ser um conjunto convexo, um cone ou simplesmente um conjunto fechado. Definimos o domínio, a imgem e o gráfico de $F$ como $\mathrm{dom}F := \{x \in \mathbb{R}^s : F(x) \neq \emptyset\}$, $\mathrm{range}F := \{y \in \mathbb{R}^d : y \in F(x) \text{ para algum } x \in \mathbb{R}^s\}$ e $\mathrm{graf}F := \{(x, y) : y \in F(x)\}$, respectivamente. Note que uma multifunção está completamente caracterizada por seu gráfico. Dizemos que $F$ é fechada, se $\mathrm{graf}F$ é um conjunto fechado.

A ideia de multifunção aparece naturalmente em diferentes áreas da Matemática, como em otimização, análise convexa, teoria de controle, análise variacional, etc. Exemplos clássicos de multifunções são os cones normais, os cones tangentes, conjunto de soluções de problemas de otimização, etc. A noção de multifunção estende naturalmente a noção usual de função.

A capacidade de usar conjuntos, funções e multifunções é uma das grandes vantagens da análise variacional. Tendo definido as multifunções procederemos a definir os limites e a continuidade.

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

$$\limsup_{z \to z^*} F(z) := \{w^* \in \mathbb{R}^d : \exists \ (z^k, w^k) \to (z^*, w^*) \text{ tal que } w^k \in F(z^k)\} \tag{2.3.19}$$

e o limite interior por

$$\liminf_{z \to z^*} F(z) := \{w^* \in \mathbb{R}^d : \forall z^k \to z^* \ \ \exists \ \ w^k \to w^* \text{ tal que } w^k \in F(z^k)\}. \tag{2.3.20}$$

Dizemos que $F$ é *semi-contínua exteriormente* (osc) em $z^*$ se

$$\limsup_{z \to z^*} F(z) \subset F(z^*) \tag{2.3.21}$$

e $F$ é *semi-contínua interiormente* (isc) em $z^*$ se

$$F(z^*) \subset \liminf_{z \to z^*} F(z). \tag{2.3.22}$$

Quando $F$ é simultaneamente semi-contínua exteriormente e semi-contínua interiormente em $z^*$ dizemos que $F$ é *contínua* em $z^*$.

Lembre que usamos a notação $\phi(t) \leq o(t)$ para qualquer função $\phi : \mathbb{R}_+ \to \mathbb{R}^s$ tal que $\limsup_{t \to 0_+} t^{-1}\phi(t) \leq 0$ (note que este é o lim sup clássico para funções).

Seja $\Omega \subset \mathbb{R}^n$ e $z^* \in \Omega$ um ponto fixo arbitrário. O *cone tangente* de $\Omega$ em $z^*$ pode ser definido como

$$T_\Omega(z^*) = \limsup_{t \downarrow 0} \frac{\Omega - z^*}{t} = \{d \in \mathbb{R}^n : \exists \ \ t_k \downarrow 0, d^k \to d \text{ tal que } z^* + t_k d^k \in \Omega\}. \tag{2.3.23}$$

Outro objeto geométrico de vital importância é o cone normal. Antes de definir o cone normal, necessitamos a ideia de cone normal regular.

O *cone normal regular* de $\Omega$ em $z^* \in \Omega$ é o conjunto definido por

$$\widehat{N}_\Omega(z^*) := \{w \in \mathbb{R}^n : \langle w, z - z^* \rangle \leq o(|z - z^*|) \text{ para } z \in \Omega\}. \tag{2.3.24}$$

O *cone normal limite* de $\Omega$ em $x^* \in \Omega$ é dado por

$$N_\Omega(z^*) := \limsup_{z \to z^*, z \in \Omega} \widehat{N}_\Omega(z). \tag{2.3.25}$$

Quando $\Omega$ é um conjunto convexo, o cone regular normal (2.3.24) e cone normal limite (2.3.25) coincidem com a noção clássica de cone normal usado na análise convexa e nesse caso usamos a notação comum $N_\Omega(z^*)$. Em geral, temos a inclusão $\widehat{N}_\Omega(z) \subset N_\Omega(z)$ para qualquer $z \in \Omega$.

**Contra-Exemplo 2.3** (A inclusão $\widehat{N}_\Omega(z) \subset N_\Omega(z)$ pode ser estrita)**.**

Considere $\Omega := \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0, x_1 x_2 = 0\}$ e o ponto $x^* = (0,0)$. Neste caso, temos que $\widehat{N}_\Omega((0,0)) = \mathbb{R}_- \times \mathbb{R}_-$ e $N_\Omega((0,0)) = \mathbb{R}_- \times \mathbb{R}_- \cup \mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R}$.

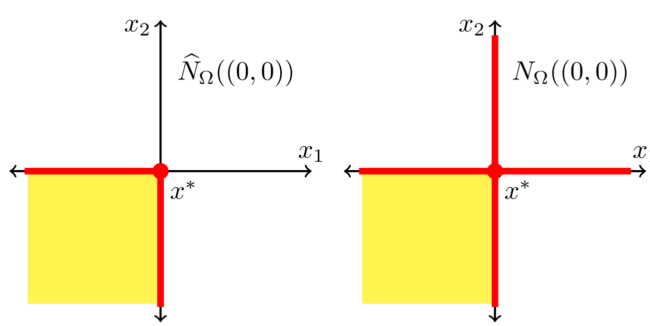



Figura 2.5: Cone normal regular e cone normal limite no ponto $(0,0)$.

Algumas propriedades adicionais

**Teorema 2.15.** *[74, Teorema 6.28]*

*a) O cone regular normal $\widehat{N}_\Omega(z^*)$ é um cone fechado convexo. Além disso, $\widehat{N}_\Omega(z^*) = T^\circ_\Omega(z^*)$.*

*b) Seja $\widehat{T}_\Omega(z^*) := \liminf_{t \downarrow 0}(\Omega - z^*)/t$ então $\widehat{T}_\Omega(z^*) = N^\circ_\Omega(z^*)$ para $\Omega$ localmente fechado em $z^*$. Como consequência $N^\circ_\Omega(z^*) \subset T_\Omega(z^*)$.*

Quando $\Omega$ é convexo, existe uma relação próxima entre o cone normal e a projeção euclidiana (projeção ortogonal).

**Proposição 2.3.** *[74, Proposição 6.17]*
*Seja $\Omega$ um conjunto convexo fechado e $x \in \Omega$. Então o cone normal $N_\Omega$ e a projeção euclidiana $P_\Omega$ cumprem a seguinte relação:*

$$\omega \in N_\Omega(x) \text{ se, e somente se } P_\Omega(x+\omega) = x, \tag{2.3.26}$$

*onde $P_\Omega(x)$ é a única solução do problema de otimização: minimizar $\|x-\omega\|^2$ sujeito a $\omega \in \Omega$.*

### 2.3.2 CQ associada a AKKT: Propriedade de Continuidade do Cone

Agora, vamos analisar em mais detalhes a condição CPG, com o intuito de enfraquecê-la. O intuito é fornecer uma decomposição mais natural do polar do cone linearizado, $\mathcal{L}^\circ(x^*)$. Seja $A_-$ o subconjunto de índices $\ell \in A(x^*)$ tal que existem $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ e $\mu_j \in \mathbb{R}_+$ para $j \in A(x^*)$, tal que

$$-\nabla g_\ell(x^*) = \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*). \tag{2.3.27}$$

Finalmente, $A_+ := A(x^*) \setminus A_-$. Ambos subconjuntos de índices $A_-$ e $A_+$ dependem explicitamente do ponto $x^*$, assim o razoável é escrever $A_-(x^*)$ e $A_+(x^*)$, mas quando é claro do contexto simplesmente escrevemos $A_-$ e $A_+$.

A condição de qualificação CPG vale em $x^*$ se existem subconjuntos (possivelmente vazios) $I' \subset \{1, \ldots, m\}$ e $J' \subset A_-$ e uma vizinhança $V$ de $x^*$ tais que

1. Os gradientes $\nabla h_i(x^*)$ e $\nabla g_j(x^*)$ indexadas por $i \in I'$ e $j \in J'$ são linearmente independentes

2. Para todo $x \in V$, se

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

com $\mu_j' \geq 0$ para todo $j \in A(x^*)$, então para todo $i \in I'$, $\ell \in J'$, e $j \in J_+$, existem $\lambda_i'' \in \mathbb{R}$, $\lambda_\ell''' \in \mathbb{R}$, e $\mu_j'' \in \mathbb{R}_+$ tal que

$$z = \sum_{i \in I'} \lambda_i'' \nabla h_i(x) + \sum_{\ell \in J'} \lambda_\ell''' \nabla g_\ell(x) + \sum_{j \in J_+} \mu_j'' \nabla g_j(x).$$

A condição CPG depende explicitamente dos subconjuntos de índices $I' \subset I$ e $J' \subset A_-$, isto é, o item 2 pode não ser satisfeito para outros subconjuntos que cumprem o item 1.

**Observação**: Devido à definição de $A_-$, $\{\nabla h_i(x^*), \nabla g_j(x^*) : i \in I', j \in J'\}$ é linearmente independente se, e somente se, $\{\nabla h_i(x^*), \nabla g_\ell(x^*), \nabla g_j(x^*) : i \in I', j \in J', j \in A_+\}$ é positivo-linearmente independente, isto é, a única solução de

$$\sum_{i \in I'} \lambda_i \nabla h_i(x^*) + \sum_{\ell \in J'} \gamma_\ell \nabla g_\ell(x^*) + \sum_{j \in A_+} \mu_j \nabla g_j(x^*) = 0, \tag{2.3.28}$$

onde $\lambda_i \in \mathbb{R}, i \in I'$, $\gamma_\ell \in \mathbb{R}, \ell \in J'$ e $\mu_j \geq 0, j \in A_+$ é a trivial. De fato, se algum $\mu_{j_0}$ para algum $j_0 \in A_+$ fosse estritamente positivo, obtemos, da expressão (2.3.28) que $j_0$ deve estar em $A_-$, o que é impossível.

Associado aos subconjuntos de índices $I' \subset I$, $J' \subset A_-$ e $A_+ = A(x^*)\backslash A-$, temos o cone

$$K_{I',J'}(x) := \left\{ \sum_{i\in I'} \lambda_i \nabla h_i(x) + \sum_{\ell \in J'} \gamma_\ell \nabla g_\ell(x) + \sum_{j\in A_+} \mu_j \nabla g_j(x) \colon \mu_j \in \mathbb{R}_+, \lambda_i, \gamma_r \in \mathbb{R} \right\}. \tag{2.3.29}$$

**Lema 2.2.** *Seja $x^* \in \Omega$ um ponto viável e sejam $I' \subset I, J' \subset A_-$ e $A_+ = A(x^*)\backslash A_-$ subconjuntos de índices tais que $\{\nabla h_i(x^*), \nabla g_j(x^*) : i \in I', j \in J'\}$ é linearmente independente. Então a multifunção $x \in \mathbb{R}^n \rightrightarrows K_{I',J'}(x)$ é semi-contínua exteriormente em $x^*$.*

*Demonstração.* Seja $\omega^* \in \limsup_{x\to x^*} K_{I',J'}(x)$, então existem sequências $\{x^k\}$ e $\{\omega^k\}$ tais que $x^k \to x^*$, $\omega^k \to \omega^*$ e $\omega^k \in K_{I',J'}(x^k)$, onde

$$\omega^k = \sum_{i\in I'} \lambda_i^k \nabla h_i(x^k) + \sum_{\ell\in J'} \gamma_\ell^k \nabla g_\ell(x^k) + \sum_{j\in A_+} \mu_j^k \nabla g_j(x^k), \tag{2.3.30}$$

para certas sequências $\{\lambda_i^k \in \mathbb{R}, i \in I'\}$, $\{\gamma_\ell^k \in \mathbb{R}, \ell \in J'\}$ e $\{\mu_j^k \in \mathbb{R}_+, j \in A_+\}$. Agora, defina $M_k := \max\{|\lambda_i^k|, i \in I'; |\gamma_\ell^k|, \ell \in J'; \mu_j^k, j \in A_+\}$. Temos as seguintes duas possibilidades:

(i) Se $\{M_k\}$ tem alguma subsequência limitada. Podemos considerar, possivelmente depois de extrair uma subsequência adequada, que para todo $i \in I', \ell \in J'$ e $j \in A_+$ as sequências $\lambda_i^k$, $\gamma_\ell^k$ e $\mu_j^k$ têm como limites $\lambda_i^*$, $\gamma_\ell^*$ e $\mu_j^*$ respectivamente. Tomando limite na expressão (2.3.30) obtemos

$$\omega^* = \sum_{i\in I'} \lambda_i^* \nabla h_i(x^*) + \sum_{\ell\in J'} \gamma_\ell^* \nabla g_\ell(x^*) + \sum_{j\in A_+} \mu_j^* \nabla g_j(x^*) \in K_{I',J'}(x^*).$$

(ii) Caso contrário, temos $M_k \to \infty$. Dividindo (2.3.30) por $M_k$, chegamos a

$$\frac{\omega^k}{M_k} = \sum_{i\in I'} \frac{\lambda_i^k}{M_k} \nabla h_i(x^k) + \sum_{l\in J'} \frac{\gamma_\ell^k}{M_k} \nabla g_\ell(x^k) + \sum_{j\in A_+} \frac{\mu_j^k}{M_k} \nabla g_j(x^k). \tag{2.3.31}$$

Desde $\max\{|\lambda_i^k/M_k|, i \in I'; |\gamma_\ell^k/M_k|, \ell \in J'; \mu_j^k/M_k, j \in A_+\} = 1$ para todo $k \in \mathbb{N}$, podemos extrair uma subsequência convergente. Por conseguinte, tomando limites na expressão (2.3.31), obtemos uma contradição com o fato que $\{\nabla h_i(x^*), \nabla g_\ell(x^*), \nabla g_j(x^*) : i \in I', \ell \in J', j \in A_+\}$ é um conjunto positivamente independente.

$\square$

Seja $x^* \in \Omega$ um ponto viável, defina

$$K(x) := \left\{ \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j\in A(x^*)} \mu_j \nabla g_j(x) : \quad \mu_j \in \mathbb{R}_+, \lambda_i \in \mathbb{R} \right\}. \tag{2.3.32}$$

O conjunto $K(x)$ é um cone fechado convexo que coincide com o cone polar linearizado $\mathcal{L}(x^*)^\circ$ em $x^* \in \Omega$. Da expressão (2.3.32) temos que o cone $K(x)$ pode ser considerado como uma pertubação do cone polar linearizado ao redor do $x^* \in \Omega$. Já que $K(x^*)$ coincide com $\mathcal{L}(x^*)^\circ$, as condições KKT podem ser escritas como $-\nabla f(x^*) \in K(x^*)$. Mais ainda, usando o cone $K(x)$ podemos re-escrever a condição CPG de uma forma mais geométrica. Dado os subconjuntos de índices $I' \subset I$, $J' \subset A_-$ e $A_+ = A(x^*) \setminus A_-$, a notação $K_{I',J'}(x)$ representa o seguinte cone

$$K_{I',J'}(x) := \left\{ \sum_{i \in I'} \lambda_i \nabla h_i(x) + \sum_{\ell \in J'} \gamma_\ell \nabla g_\ell(x) + \sum_{j \in A_+} \mu_j \nabla g_j(x) \colon \mu_j \in \mathbb{R}_+, \lambda_i, \gamma_r \in \mathbb{R} \right\}.$$

Assim, a condição CPG vale em $x^*$ se existem $I' \subset \{1, \ldots, m\}$, $J' \subset A_-$ e uma vizinhança $V$ de $x^*$ tal que:

1. Os gradientes $\{\nabla h_i(x^*), \nabla g_j(x^*) : i \in I', j \in J'\}$ são linearmente independentes;

2. Para todo $x \in V$
$$K(x) \subset K_{I',J'}(x). \tag{2.3.33}$$

Claramente os cones coincidem em $x^*$, isto é, $K(x^*) = K_{I',J'}(x^*)$.

Assim, a multifunção $K(x)$ permite uma descrição mais geométrica da condição CPG. Ainda mais, como $K(x^*) = \mathcal{L}_\Omega(x^*)^\circ$, o comportamento de $K(x)$ quando $x$ se aproxima de $x^*$ merece uma atenção especial. Esta observação nos leva a considerar a seguinte propriedade

**Definição 2.15.** *Seja $x^*$ um ponto viável. A Propriedade de Continuidade do Cone (CCP) vale em $x^* \in \Omega$ se a multifunção $x \in \mathbb{R}^n \rightrightarrows K(x)$ é semicontínua exteriormente em $x^*$, isto é:*
$$\limsup_{x \to x^*} K(x) \subset K(x^*). \tag{2.3.34}$$

Note que devido à continuidade dos gradientes e da própria definição de $K(x)$, a multifunção $x \in \mathbb{R}^n \rightrightarrows K(x)$ é sempre semicontínua interiormente. Por essa razão, a semi-continuidade exterior é suficiente para a continuidade de $K(x)$ em $x^*$. A condição sequencial AKKT está naturalmente associada com a Propriedade do Cone Contínuo. Da definição de AKKT, existem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$ e $\{\mu^k\} \subset \mathbb{R}^p_+$ com $\mu^k_j = 0$ para $j \notin A(x^*)$, tais que $\lim_{k \to \infty} x^k = x^*$ e

$$\lim_{k \to \infty} \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) = 0. \tag{2.3.35}$$

A expressão (2.3.35) diz que se a condição AKKT é satisfeita, os gradientes $-\nabla f(x^k)$ se aproximam do cone $K(x^k)$ quando $k$ vai para infinito.

O seguinte teorema mostra que a condição CCP desempenha, com respeito à condição AKKT, o mesmo papel que a condição de Guignard desempenha com respeito à otimalidade local. Como mencionamos durante o texto, a condição de Guignard é a condição de qualificação mínima para garantir que todo mínimo local implica as condições KKT. Do mesmo modo temos que CCP é a condição mínima com respeito às restrições que garante que a condição AKKT implica as condições KKT.

**Teorema 2.16.** *A condição CCP é a propriedade mínima com respeito às restrições sob a qual a condição AKKT implica a validade das condições KKT independentemente da função objetivo.*

*Demonstração.* Primeiro mostraremos que se a condição CCP vale, então a condição sequencial AKKT implica as condições KKT independentemente da função objetivo. Seja $f$ uma função objetivo tal que a condição sequencial AKKT vale em $x^*$, então pela definição existem sequências $\{x^k\} \to x^*$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p_+$ com $\mu^k_j = 0$ para $j \notin J(x^*)$ e $\{\zeta^k\} \subset \mathbb{R}^n$ tais que

$$\zeta^k := \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla g_j(x^k) \to 0. \tag{2.3.36}$$

Defina

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

De (2.3.36) e da definição do cone $K(x)$ temos que

$$\omega^k \in K(x^k) \ \text{ e } \ \omega^k = \zeta^k - \nabla f(x^k). \tag{2.3.37}$$

Tomando limites em (2.3.37), obtemos que

$$-\nabla f(x^*) = \lim_{k \to \infty} \omega^k \in \limsup_{x \to x^*} K(x) \subset K(x^*), \tag{2.3.38}$$

onde a última inclusão segue da propriedade do cone contínuo. Desta forma, $-\nabla f(x^*)$ pertence ao polar do cone linearizado $K(x^*) = \mathcal{L}(x^*)^\circ$, ou equivalentemente temos que as condições KKT valem em $x^*$.

A seguir, provaremos que se a implicação AKKT $\Rightarrow$ KKT, independente da função objetivo, então a condição CCP vale.

Seja $\omega^* \in \limsup_{x \to x^*} K(x)$. Da definição de limite exterior, existem sequências $\{x^k\}$ e $\{\omega^k\}$ tais que $x^k \to x^*$, $\omega^k \to \omega^*$ e $\omega^k \in K(x^k)$. Defina $f(x) := -\langle w^*, x\rangle$ para todo $x \in \mathbb{R}^n$. Note que a condição AKKT vale em $x^*$ para esta função com $\{x^k\}$ como a sequência AKKT associada porque $\nabla f(x^k) + \omega^k = -\omega^* + \omega^k \to 0$. Logo, por hipótese, as condições KKT valem em $x^*$, isto é, $-\nabla f(x^*) = \omega^* \in K(x^*)$. Assim, a multifunção $K(x)$ é semicontínua exteriormente em $x^*$. Portanto, CCP vale. □

Já que AKKT é uma condição de otimalidade, temos o seguinte corolário.

**Corolário 2.1.** *A condição CCP é uma condição de qualificação.*

A convergência para pontos KKT de vários algoritmos práticos que geram sequências AKKT (como por exemplo, o método de programação quadrática sequencial de Qi e Wei [71], o método de pontos interiores de Chen e Goldfarb [37] e os métodos de lagrangiano aumentado de [4,31], entre outros) tem sido provada assumindo diferentes condições de qualificação. Com o Teorema 2.16 à disposição, em todos os casos, podemos substituir a condição de qualificação usada pela condição CCP, que é mais fraca, e esse tipo de resultado não pode ser melhorado, em princípio, usando outras condições de qualificação.

Note que se a condição CCP falha em um ponto viável $x$, é porque podemos definir uma função objetivo suave tal que $x$ é um ponto AKKT, mas não um ponto KKT. Na Figura 2.6, a área sombreada representa a região viável formada pela

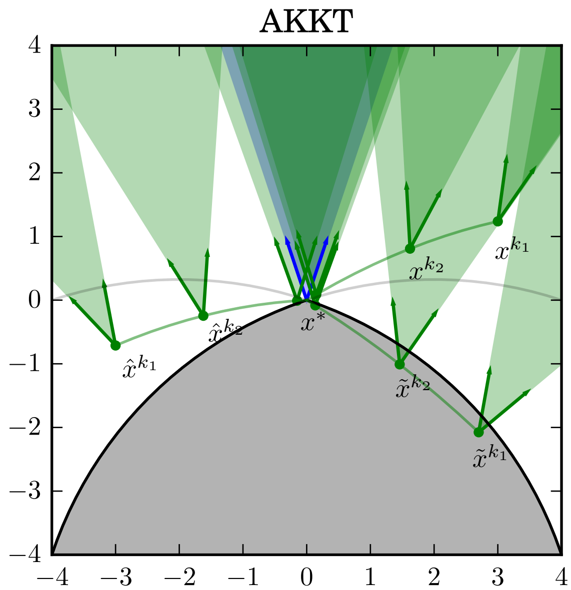



Figura 2.6: Exemplo do cone associado à condição AKKT.

interseção transversal de dois círculos. A figura mostra o comportamento do cone $K(x)$ quando $x$ vai para a origem, através de diferentes sequências.

Como consequência do Teorema 2.16, "KKT ou não-CCP" fornece a maior medida de força da condição AKKT. Assim, a exata posição da CCP com respeito às outras CQs, é relevante. Por [11] temos que CCP é mais fraca que CPG e mais forte que a condição de Abadie. Procedemos provando que CPG implica CCP.

**Teorema 2.17.** *CPG implica CCP.*

*Demonstração.* Devido à definição de CPG, existem subconjuntos de índices $I'$, $J'$ e $A_+$ tais que os gradientes $\nabla h_i(x^*)$ e $\nabla g_j(x^*)$ com $(i,j) \in (I', J')$ são linearmente independentes e uma vizinhança $V$ de $x^*$ tal que

$$K(x) \subset K_{I',J'}(x) \quad \forall x \in V. \tag{2.3.39}$$

Agora, tomando limites em (2.3.39) e usando a semi-continuidade externa de $K_{I',J'}(x)$ em $x^*$, ver Lema 2.2, temos que

$$\limsup_{x \to x^*} K(x) \subset \limsup_{x \to x^*} K_{I',J'}(x) \subset K_{I',J'}(x^*) = K(x^*). \tag{2.3.40}$$

O que equivale a dizer que CCP vale em $x^*$. □

O próximo exemplo mostra que CCP é estritamente mais fraca que CPG.

**Exemplo 2.3.** *(CCP não implica CPG.)*

Em $\mathbb{R}^2$, consideremos $x^* = (0,0)$ e a região viável

$$\Omega := \{(x_1, x_2) : g_1(x_1,x_2) = x_1 \leq 0, g_2(x_1,x_2) = (x_1^+)^2 \exp((x_2^+)^2) \leq 0\} \tag{2.3.41}$$

onde $x_i^+ := \max\{0, x_i\}, i = 1, 2$. Claramente, $x^* = (0, 0)$ é um ponto viável e ambas restrições são ativas em $x^*$. Calculando os gradientes das restrições temos que

$$\nabla g_1(x_1, x_2) = (1, 0) \text{ para } x = (x_1, x_2) \in \mathbb{R}^2$$

e

$$\nabla g_2(x_1, x_2) = (2x_1^+ \exp((x_2^+)^2), 2x_2^+(x_1^+)^2 \exp((x_2^+)^2)) \text{ para } x = (x_1, x_2) \in \mathbb{R}^2$$

Como $\nabla g_1(x^*) = (1, 0)$ e $\nabla g_2(x^*) = (0, 0)$, temos que $K(x^*) = \mathbb{R}_+ \times \{0\}$ e a única possível escolha para uma base positiva de $K(x^*)$ é $\{I^{'} = \emptyset, J^{'} = \emptyset, A_+ = \{1\}\}$. Assim, $K_{I', J'}(x) = \mathbb{R}_+ \times \{0\}$ para qualquer $x = (x_1, x_2) \in \mathbb{R}^2$.

Por outro lado, temos

$$K(x) = \{(\mu_1 + 2\mu_2 x_1^+ \exp((x_2^+)^2), 2\mu_2 x_2^+(x_1^+)^2 \exp((x_2^+)^2)) \mid \quad \mu_1, \mu_2 \geq 0\}.$$

Claramente $K(x)$ não pode ser um subconjunto de $K_{I', J'}(x)$ em nenhuma vizinhança de $x^* = (0, 0)$ (escolha qualquer $x_1 > 0$, $x_2 > 0$ próximos a zero e $\mu_2 > 0$) e como consequência CPG falha.

Agora procederemos a provar que $K(x)$ é semicontínua exterior em $x^*$. Seja $\omega^* \in \limsup_{x \to x^*} K(x)$. Por conseguinte, existem sequências $\{x^k\}$ e $\{\omega^k\}$ tais que $x^k = (x_1^k, x_2^k) \to x^*$, $\omega^k = (\omega_1^k, \omega_2^k) \to \omega^*$ e

$$\omega^k = \mu_1^k(1, 0) + \mu_2^k(2x_1^+ \exp((x_2^+)^2), 2x_2^+(x_1^+)^2 \exp((x_2^+)^2)) \in K(x^k), \tag{2.3.42}$$

para certos $\mu_1^k, \mu_2^k \geq 0$. Agora, suponha por contradição que $\omega^* = (\omega_1^*, \omega_2^*)$ não pertence ao cone $K(x^*) = \mathbb{R}_+ \times \{0\}$. Assim, $\omega_2^*$ deve ser diferente de zero. De (2.3.42) temos que existe $\rho > 0$ tal que

$$|\omega_2^k| = 2\mu_2^k |(x_2^k)^+((x_1^k)^+)^2 \exp(x_2^k)^+)^2| > \rho > 0 \tag{2.3.43}$$

para $k$ suficientemente grande. Usando $\mu_1^k \geq 0$ e (2.3.43), temos

$$\omega_1^k = \mu_1^k + 2\mu_2^k (x_1^k)^+ \exp((x_2^k)^+)^2 \geq \frac{|\omega_2^k|}{(x_1^k)^+(x_2^k)^+} > \frac{\rho}{(x_1^k)^+(x_2^k)^+} > 0 \tag{2.3.44}$$

Tomando limites em (2.3.44) chegamos a uma contradição, já que $\omega_1^k$ converge para $\omega_1^*$. Assim, $\omega^* = (\omega_1^*, \omega_2^*)$ pertence a $K(x^*)$.

Observe que neste exemplo o cone $K(x)$ é o cone positivo gerado pelos vetores $(1, 0)$ e por $(1, x_1^+ x_2^+)$.

### 2.3.3 CQ associada a AGP: AGP-regular

A seguir, vamos descrever a condição de qualificação mínima associada à força da condição de otimalidade sequencial AGP.

**Definição 2.16.** *Seja $x^* \in \Omega$ dizemos que $x^*$ satisfaz a condição AGP-regular se a multifunção*

$$(x, \varepsilon) \in \mathbb{R}^n \times \mathbb{R}^n \rightrightarrows N_{\Omega(x, -\infty)}(x + \varepsilon) \tag{2.3.45}$$

*é semicontínua exteriormente em $(x^*, 0)$. Isto é:*

$$\limsup_{(x,\varepsilon) \to (x^*, 0)} N_{\Omega(x, -\infty)}(x + \varepsilon) \subset N_{\Omega(x^*, -\infty)}(x^*) = K(x^*). \tag{2.3.46}$$

Como consequência do Lema de Farkas, o cone normal $N_{\Omega(x,-\infty)}(x+\varepsilon) = \mathcal{L}_{\Omega(x,-\infty)}(x+\varepsilon)^{\circ}$ pode ser calculado explicitamente. De fato, temos a seguinte proposição.

**Proposição 2.4.** *Todo elemento do cone $N_{\Omega(x,-\infty)}(x+\varepsilon)$ é da forma*

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{j:g_j(x)\geq 0} \mu_{1j} \nabla g_j(x) + \sum_{j:g_j(x)<0} \mu_{2j} \nabla g_j(x),$$

*onde $\lambda_i \in \mathbb{R}$, $\mu_{1j} \in \mathbb{R}_+$, $\mu_{2j} \in \mathbb{R}_+$ e se satisfazem as relações*

$$\mu_{1j}(\langle \nabla g_j(x), \varepsilon \rangle) = 0, \ \text{se } g_j(x) \geq 0 \quad e \quad \mu_{2j}(g_j(x) + \langle \nabla g_j(x), \varepsilon \rangle) = 0, \ \text{se } g_j(x) < 0.$$

Da proposição anterior, temos que a igualdade $N_{\Omega(x,-\infty)}(x) = \mathcal{L}_{\Omega}(x)^{\circ}$ vale para todo ponto viável $x$.

Alternativamente a condição AGP-regular é equivalente a

$$\limsup_{(x,y)\to(x^*,x^*)} N_{\Omega(x,-\infty)}(y) \subset N_{\Omega(x^*,-\infty)}(x^*) = K(x^*)$$

Como $N_{\Omega(x,-\infty)}(y)$ é um conjunto vazio para $y \notin \Omega(x,-\infty)$, temos que $N_{\Omega(x,-\infty)}(y)$ só tem sentido, se $y \in \Omega(x,-\infty)$. Assim, $y$ é aproximadamente tão inviável como é o ponto $x$.

O Teorema 2.18 mostra que a semi-continuidade exterior de $N_{\Omega(x,-\infty)}(x+\varepsilon)$ em $(x^*,0)$ é a condição mínima para garantir que a condição sequencial AGP implica as condições KKT independentemente da função objetivo. Assim, a condição AGP regular desempenha, com respeito à condição AGP, o mesmo papel que a condição de Guignard desempenha com respeito à otimalidade local.

**Teorema 2.18.** *A condição AGP-regular é a mínima condição com respeito às restrições para o qual a condição AGP implica a condição KKT, independentemente da função objetivo.*

*Demonstração.* Começaremos por demonstrar que, sob a condição AGP-regular, AGP implica KKT. Seja $\gamma \in [-\infty, 0)$ e seja $f$ uma função objetivo para o qual a condição AGP($\gamma$) vale. Pela definição de AGP($\gamma$) existe uma sequência $\{x^k\} \in \mathbb{R}^n$, tal que $x^k \to x^*$ e $P_{\Omega(x^k,\gamma)}(x^k - \nabla f(x^k)) - x^k \to 0$.

Defina $y^k := P_{\Omega(x^k,\gamma)}(x^k - \nabla f(x^k))$ e $\varepsilon^k := y^k - x^k$. Note que da condição AGP, $\varepsilon^k \to 0$. Da Proposição 2.3, temos que

$$\omega^k := x^k - \nabla f(x^k) - y^k \in N_{\Omega(x^k,\gamma)}(y^k = x^k + \varepsilon^k). \tag{2.3.47}$$

Devido à inclusão $\Omega(x,-\infty) \subset \Omega(x,\gamma)$, temos que o cone normal $N_{\Omega(x^k,\gamma)}(y^k)$ é um subconjunto de $N_{\Omega(x^k,-\infty)}(y^k)$. A sequência $\{\omega^k\}$ satisfaz as seguintes relações:

$$\omega^k \in N_{\Omega(x^k,-\infty)}(x^k + \varepsilon^k) \ \text{ e } \ \omega^k = x^k - \nabla f(x^k) - y^k = -\nabla f(x^k) - \varepsilon^k. \tag{2.3.48}$$

Tomando limite na expressão anterior e usando a continuidade do gradiente de $f$

$$-\nabla f(x^*) = \lim_{k\to\infty} \omega^k \in \limsup_{(x,\varepsilon)\to(x^*,0)} N_{\Omega(x,-\infty)}(x+\varepsilon) \subset N_{\Omega(x^*,-\infty)}(x^*). \tag{2.3.49}$$

Portanto, concluímos que $-\nabla f(x^*)$ pertence ao cone normal $N_{\Omega(x^*,-\infty)}(x^*) = K(x^*)$ ou equivalentemente as condições KKT valem em $x^*$.

Agora, provaremos que se a condição sequencial AGP implica as condições KKT para toda função objetivo, então a condição AGP-regular vale.

Seja $\omega^* \in \limsup_{(x,\varepsilon)\to(x^*,0)} N_{\Omega(x,-\infty)}(x+\varepsilon)$, logo existem sequências $\{x^k\}$, $\{\omega^k\}$ e $\{\varepsilon^k\}$ tais que $x^k \to x^*$, $\varepsilon^k \to 0$, $\omega^k \to \omega^*$ e $\omega^k \in N_{\Omega(x^k,-\infty)}(x^k+\varepsilon^k)$. Defina a função $f(x) := -\langle w^*, x\rangle$ para todo $x \in \mathbb{R}^n$.

Mostraremos que a condição sequencial AGP$(-\infty)$ vale em $x^*$ para essa escolha de $f$. Nosso objetivo é provar $\lim_{k\to\infty} P_{\Omega(x^k,-\infty)}(x^k - \nabla f(x^k)) - x^k = 0$. Defina $y^k := x^k+\varepsilon^k$ e $z^k := P_{\Omega(x^k,-\infty)}(x^k - \nabla f(x^k)) = P_{\Omega(x^k,-\infty)}(x^k+\omega^*)$. Note que, como $\omega^k \in N_{\Omega(x^k,-\infty)}(y^k)$, temos, pela Proposição 2.3, que $P_{\Omega(x^k,-\infty)}(\omega^k + y^k) = y^k$. Usando a desigualdade triangular e a não expansividade da projeção ortogonal, segue que

$$\|z^k - y^k\| = \|P_{\Omega(x^k,-\infty)}(x^k+\omega^*) - P_{\Omega(x^k,-\infty)}(\omega^k + y^k)\| \le \|\omega^* - \omega^k\| + \|y^k - x^k\|. \tag{2.3.50}$$

Tomando limite em (2.3.50) e usando $\lim_{k\to\infty}(y^k - x^k) = \lim_{k\to\infty} \varepsilon^k = 0$ temos $\lim_{k\to\infty} z^k - y^k = 0$. Como consequência

$$\lim_{k\to\infty} P_{\Omega(x^k,-\infty)}(x^k - \nabla f(x^k)) - x^k = \lim_{k\to\infty} z^k - x^k = \lim_{k\to\infty}(z^k - y^k) + \lim_{k\to\infty}(y^k - x^k) = 0. \tag{2.3.51}$$

Portanto a condição sequencial AGP vale em $x^*$ e da hipótese concluímos que as condições KKT valem em $x^*$, isto é, $-\nabla f(x^*) = \omega^*$ pertence a $N_{\Omega(x^*,-\infty)}(x^*) = K(x^*)$. □

Já que que AGP é uma condição de otimalidade temos

**Corolário 2.2.** *A condição AGP-regular é uma condição de qualificação.*

Observe que como a condição sequencial AKKT é implicada pela condição sequencial AGP, temos o seguinte resultado

**Proposição 2.5.** *CCP implica AGP-regular*

A Figura 2.7 mostra um exemplo do cone associado à condição AGP. O conjunto viável é formado pela interseção de dois círculos e o ponto de interesse é $x^* = (0,0)$. Note que AGP-regular não só olha o ponto atual $x$, mas também pontos próximos a $x$ que são quase tão viáveis quanto $x$.

Na Figura 2.8 apresentamos um diagrama completo de relações entre diversas condições de qualificação apresentadas ao longo do texto. Uma análise sobre as relações existentes entre outras condições de qualificação e condições de otimalidade, é desenvolvida em [12].

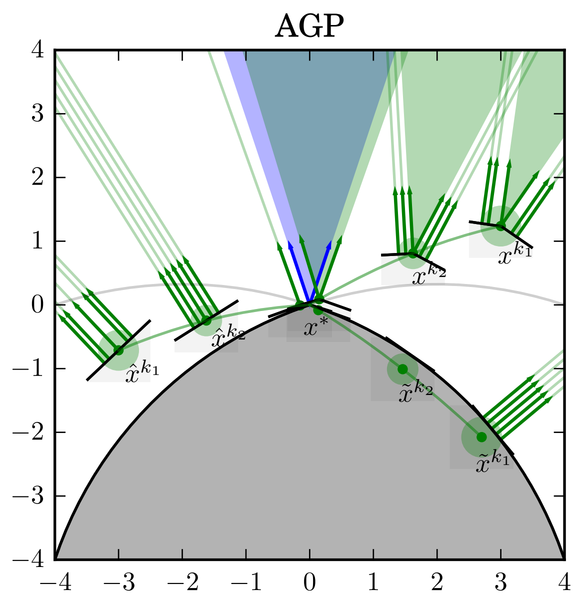



Figura 2.7: Exemplo do cone associado à condição AGP.

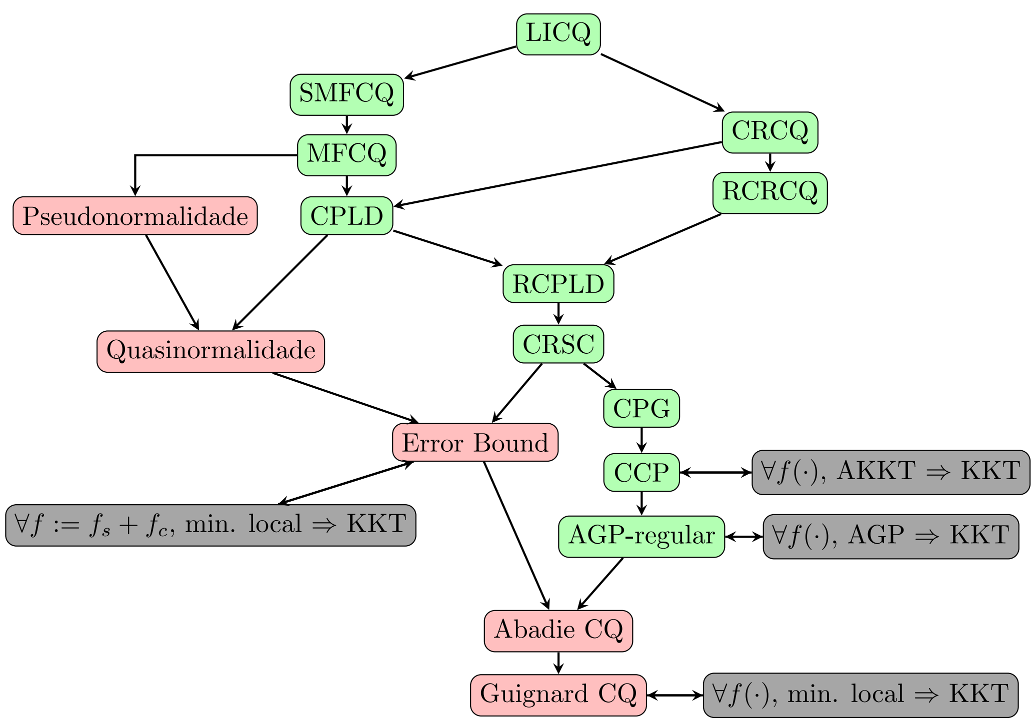



Figura 2.8: Relações entre condições de qualificação. $f_s$ é uma função $C^1$ e $f_c$ é uma função convexa (possivelmente não suave) quaisquer.

## 2.4 Exercícios

1. Suponha que $x \in \Omega$ e $-\nabla f(x) \notin T_\Omega(x)^\circ$. Mostre que existe uma direção de descida. Isto é, uma direção $d$, $f(x+td) < f(x)$ para todo $t \in (0, \varepsilon]$. Note que $x + td$ não necessariamente pertence a $\Omega$ para $t$ perto de 0.

2. Prove o Lema de Carathéodory: Se $x \in \mathbb{R}^n$ é combinação linear de $v_1, \ldots, v_m$ com escalares não-nulos $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$, então existe um subconjunto $J \subset \{1, \ldots, m\}$ e escalares $b_i, i \in J$ tais que $x = \sum_{i \in J} \beta_i v_i$, $\{v_i\}$ é linearmente independente, e $\beta_i$ tem o mesmo sinal de $\alpha_i, i \in J$.

3. Dados vetores $u_i, i = 1, \ldots, m$ e $v_j, j = 1, \ldots, p$ em $\mathbb{R}^n$, use o Lema de Carathéodory para provar que o cone

$$\{x \in \mathbb{R}^n \mid x = \sum_{i=1}^{m} \lambda_i u_i + \sum_{j=1}^{p} \mu_j v_j, \lambda_i \in \mathbb{R}, i = 1, \ldots, m, \mu_j \geq 0, j = 1, \ldots, p\}$$

   é fechado.

4. Prove que a condição de Fritz-John é equivalente a "KKT ou não-MFCQ". MFCQ é a condição de Mangasarian-Fromovitz.

5. Prove que LICQ implica existência e unicidade dos multiplicadores de Lagrange.

6. Prove que MFCQ é equivalente a dizer o conjunto dos multiplicadores de Lagrange é não vazio, fechado e limitado.

7. Mostre que MFCQ não vale em nenhum ponto de $\Omega := \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0, x_1 x_2 = 0\}$. O que acontece com a condição CPLD?

8. Escolha índices $\mathcal{I} \subset A(x)$ e $\mathcal{B} \subset \{1, \ldots, m\}$ com $x$ viável. Suponha que CPLD vale em $x$ e existe uma sequência $x^k \to x$, $x^k$ não necessariamente viável, tal que $\{\nabla h_i(x^k) : i \in \mathcal{B}\} \cup \{\nabla g_j(x^k) : j \in \mathcal{I}\}$ seja linearmente independente para todo $k \in \mathbb{N}$. Mostre que $\{\nabla h_i(x) : i \in \mathcal{B}\} \cup \{\nabla g_j(x) : j \in \mathcal{I}\}$ é linearmente independente.

9. Seja $\Omega := \{x \in \mathbb{R}^n : f_j(x) \leq 0, j \in \{m+1, \ldots, m+p\}, f_i(x) = 0, i \in \{1, \ldots, m\}\}$ e tome $\mathcal{B} \subset A_-(x) \cup \{1, \ldots, m\}$ tal que $\{\nabla f_j(x)\}_{j \in \mathcal{B}}$ é linearmente independente.

   - Mostre que MFCQ vale em $\Omega_1 := \{x \in \mathbb{R}^n : f_j(x) \leq 0, j \in A_+(x), f_j(x) = 0, i \in \mathcal{B}\}$.
   - Suponha que CRSC vale em $x \in \Omega$. Mostre que $\Omega$ pode ser localmente descrito como $\widehat{\Omega} := \{x \in \mathbb{R}^n : f_j(x) \leq 0, j \in A_+(x), f_j(x) = 0, i \in A_-(x) \cup \{1, \ldots, m\}\}$.

10. Considere as sequências $\{x^k\}$, $\{\lambda^k\}$, $\{\mu^k\}$, $\{r^k\}$ e $\{\delta^k\}$ tais que

$$\left\{ \begin{array}{l} r^k := \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) \\ g_j(x^k) \leq \delta^k, \|h_i(x^k)\| \leq \delta^k, i = 1, \ldots, m \text{ e } j = 1, \ldots, p \\ \mu_j^k \geq 0, j = 1, \ldots, p \\ \sum_{j=1}^{p} \mu_j^k (g_j(x^k) - \delta^k) = 0. \end{array} \right\} \quad (2.4.52)$$

onde $r^k \to 0$ e $\delta^k \to 0$. Seja $x^*$ qualquer ponto limite de $\{x^k\}$. Prove que $x^*$ é um ponto AKKT.

A sequência (2.4.52) foi usada por Qi e Wei como definição de aproximadamente KKT. Note que a sequência (2.4.52) não faz menção ao ponto limite $x^*$.

11. Suponha que $f$, $g_j, j = 1, \dots, p$ são funções convexas e que $h_i, i = 1, \dots, m$ são funções afins. Seja $x^*$ é um ponto AKKT e que $\{(x^k, \lambda^k, \mu^k)\}$ é uma sequência AKKT associada a $x^*$. Se $\sum_{i=1}^m \lambda_i^k h_i(x^k) \geq 0$ para todo $k \in \mathbb{N}$, prove que $x^*$ é uma solução.

12. Prove que se $x$ é um ponto AKKT então $x$ é também um ponto Fritz-John. Assim, do ponto de vista de otimização, pontos AKKT são melhores candidatos à soluções que pontos Fritz-John.

13. Prove que se $P_{\Omega(x,\gamma)}(x - \nabla f_0(x)) - x = 0$ se, e somente se as condições KKT valem em $x$.

14. Denote por $d_\gamma(x) := P_{\Omega(x,\gamma)}(x - \nabla f_0(x)) - x$, o gradiente projetado aproximado. Mostre que existem $\tau, \theta, \theta_1, \theta_2 > 0$ tal que

$$\begin{aligned} f(x + td_\gamma(x)) &\leq f(x) - \theta t \|d_\gamma(x)\|^2 \\ |h_i(x + td_\gamma(x))| &\leq |h_i(x)| + \theta_1 t \|d_\gamma(x)\|^2, i = 1, \dots, m \\ |\max\{g_j(x + td_\gamma(x)), 0\}| &\leq |\max\{g_j(x), 0\}| + \theta_2 t \|d_\gamma(x)\|^2, j = 1, \dots, p \end{aligned}$$

para todo $t \in [0, \tau]$. Assim, a direção $d_\gamma(x)$ melhora a otimalidade de $x$ enquanto não deteriora muito a viabilidade obtida.

15. Quando calculamos $d_\gamma(x) := P_{\Omega(x,\gamma)}(x - \nabla f(x)) - x$, usamos a norma euclideana para calcular a projeção. Quais propriedades de $d_\gamma(x)$ se mantêm quando em lugar da norma euclideana usamos outra norma (tal como a norma-1) para calcular a projeção?

16. Suponha que $f$, $g_j, j = 1, \dots, p$ são funções convexas e que $h_i, i = 1, \dots, m$ são funções afins. Seja $x^*$ um ponto AGP tal que $h(x^k) = 0$ para todo $k \in \mathbb{N}$. Mostre que $x^*$ é uma solução do problema.

    Conclua que na ausência de restrições de igualdade, no caso convexo, achar pontos AGP equivale a resolver o problema de otimização.

17. Prove as seguintes propriedades:

    (a) Sejam $\Phi, \Psi$ multifunções. Prove:

$$\limsup_{x \to x_0} \Phi(x) \cap \Psi(x) \subset \limsup_{x \to x_0} \Phi(x) \cap \limsup_{x \to x_0} \Psi(x)$$

$$\liminf_{x \to x_0} \Phi(x) \cup \Psi(x) \supset \liminf_{x \to x_0} \Phi(x) \cup \liminf_{x \to x_0} \Psi(x)$$

    (b) Seja $\Phi$ uma multifunção tal que $\Phi(x)$ é fechado para todo $x$. Mostre que

$$\limsup_{x \to x_0} \Phi(x) = \{w : \liminf_{x \to x_0} \operatorname{dist}(w, \Phi(x)) = 0\}$$

$$\liminf_{x \to x_0} \Phi(x) = \{w : \limsup_{x \to x_0} \operatorname{dist}(w, \Phi(x)) = 0\},$$

    onde $\operatorname{dist}(x, K)$ denota a distância de $x$ a $K$.

(c) Seja $\Psi$ uma multifunção tal que $\Psi(x)$ seja convexo e fechado para todo $x$. Prove
$$\liminf_{x \to x_0} \Psi(x) = (\limsup_{x \to x_0} \Psi(x)^\circ)^\circ.$$

18. Mostre que $T_\Omega(x) = \limsup_{t \to 0^+} (\Omega - x)/t$.

19. Seja $\{\nabla f_j(x) : j \in \mathcal{J}\}$ uma família de vetores onde as $f_j$ são continuamente diferenciáveis. Prove que $\{\nabla f_j(x) : j \in \mathcal{J}\}$ tem o mesmo posto para todo $x$ perto de $x^*$ se, e somente a multifunção $S(x) := \text{span}\ \{\nabla f_j(x) : j \in \mathcal{J}\}$ é osc em $x = x^*$.

20. Seja $x^* \in \Omega$. Prove que CCP é equivalente à semi-continuidade interior da multifunção $L_\Omega(x, x^*)$ em $x = x^*$, onde
$$L_\Omega(x, x^*) := \{d \in \mathbb{R}^n : \nabla g_j(x)^\mathtt{T} d \leq 0, \nabla h_i(x)^\mathtt{T} d = 0, j \in A(x^*), i \in \{1, \ldots, m\}\} \tag{2.4.53}$$

21. Mostre que existe uma função $f$ tal que $x^* = (0, 0)$ é um ponto AKKT para o problema de otimização min $f(x)$ s.t. $x \in \Omega$ tal que $x^*$ não é KKT, onde $\Omega = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0, x_1 x_2 = 0\}$.

22. Seja $x \in \Omega = \{x : g(x) \leq 0, h(x) = 0\}$. Denote por $r$ o número de restrições ativas de desigualdade para o ponto $x$. Prove que se a multifunção
$$x \in \mathbb{R}^n \Rightarrow L(x) := \{(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^r_+ : \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j=1}^r \mu_j \nabla g_j(x) = 0\}$$
é semi-contínua interiormente em $x = x^*$, então CCP também vale em $x^*$.

23. Prove a Proposição 2.4.

24. Considere a região $\Omega = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = 0, x_2 - x_2^2 x_1 - x_1^4 \leq 0\}$. Mostre que AGP-regular falha em $x^* = (0, 0)$.

# Capítulo 3

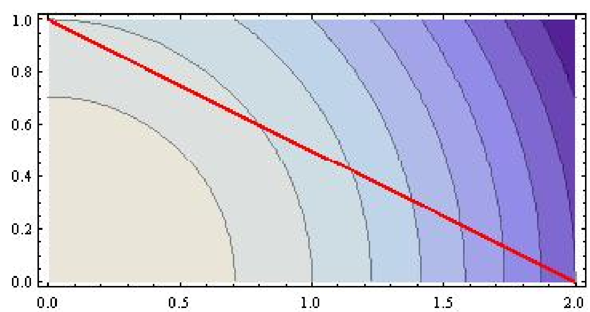
# Condições de otimalidade de segunda ordem

Neste capítulo vamos supor que as todas as funções $f$, $h_i$, $g_j : \mathbb{R}^n \to \mathbb{R}$ para $i = 1, \ldots, m$; $j = 1, \ldots, p$ são duas vezes continuamente diferenciáveis.

Considere o problema

$$\begin{array}{ll} \text{Minimizar} & f(x) := -x_1^2 - x_2^2, \\ \text{Sujeito a} & g_1(x) := x_1 + 2x_2 - 2 \leq 0, \\ & g_2(x) := -x_1 \leq 0, \\ & g_3(x) := -x_2 \leq 0. \end{array}$$

A região viável é o triângulo de vértices $(0,0), (2,0)$ e $(0,1)$ descrito na figura abaixo juntamente com as curvas de nível da função objetivo:




No ponto $x = (0.4, 0.8)$ temos $A(x) = \{1\}$ e a condição KKT

$$\nabla f(x) + \mu_1 \nabla g_1(x) = 0, \mu_1 \geq 0$$

é satisfeita com $\mu_1 = 0.8$, entretanto, observamos que $x$ não é uma solução. De fato, a partir de $x$, a função objetivo $f$ aumenta ao longo de direções que apontam para o interior da região viável (ou seja, direções $d \in \mathbb{R}^2$ tais que $\nabla g_1(x)^{\mathtt{T}} d < 0$), mas ao longo de direções viáveis ortogonais ao gradiente de $g_1$, a função objetivo decresce. O que ocorre é que a aproximação de primeira ordem não olha para direções viáveis ortogonais ao gradiente das restrições ativas em $x$, e erroneamente declara que o ponto $x$ em questão é um candidato a minimizador. Assim, informações de primeira ordem, tais como os gradientes ou direções do cone tangente, não são suficientes para caracterizar as soluções. A dificuldade é que tais direções, denominadas direções críticas, podem ou não ser direções viáveis, dependendo da curvatura das restrições.

Condições de otimalidade de segunda ordem exigem que ao longo de direções críticas a aproximação de segunda ordem da função objetivo não seja decrescente. De fato, neste exemplo, observamos que ao longo da direção crítica $d = (2, -1)$ com $\nabla g_1(x)^{\mathtt{T}} d = 0$, temos $f(x + d) \approx f(x) + \frac{1}{2} d^{\mathtt{T}} \nabla^2 f(x) d$ e como $d^{\mathtt{T}} \nabla^2 f(x) d < 0$ podemos perceber que $x$ não é um minimizador local.

No caso geral de restrições não-lineares, a curvatura das restrições desempenha um papel importante, sendo assim, a análise acima pode ser aplicada considerando o problema com função objetivo dada por $L(x) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{p} \mu_j g_j(x)$, para um multiplicador de Lagrange $(\lambda, \mu)$ fixado, que sub-estima $f$ em pontos viáveis.

**Definição 3.1.** *Dizemos que o ponto $x \in \Omega$, que satisfaz KKT, cumpre com a condição necessária (forte) de segunda ordem (SSONC) associada ao multiplicador de Lagrange $(\lambda, \mu)$ quando*

$$d^{\mathtt{T}} \left( \nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x) + \sum_{j \in A(x)} \mu_j \nabla^2 g_j(x) \right) d \geq 0, \forall d \in C^S(x),$$

*onde*

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

*é o cone crítico (forte).*

Note que, independente da escolha do multiplicador de Lagrange $\lambda$ associado a $x$, a descrição do cone crítico pode ser feita da seguinte maneira:

$$C^S(x) = \left\{ d \in \mathbb{R}^n \mid \begin{array}{l} \nabla h_i(x)^{\mathtt{T}} d = 0, i \in \{1, \ldots, m\}; \nabla g_j(x)^{\mathtt{T}} d = 0, j \in A(x), \mu_j > 0; \\ \nabla g_j(x)^{\mathtt{T}} d \leq 0, j \in A(x), \mu_j = 0 \end{array} \right\}$$

Observe que para toda direção $d$ no cone crítico $C^S(x)$ temos que $\nabla f(x)^{\mathtt{T}} d = 0$.

Note que definindo a função Lagrangiana $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p_+$ como

$$L(x, \lambda, \mu) := f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{p} \mu_j g_j(x),$$

temos que a condição de otimalidade de primeira ordem em um ponto $x \in \Omega$ é satisfeita se existe $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p$ com $\mu_j \geq 0, j = 1, \ldots, p$ e $\mu_j = 0$, se $j \notin A(x)$ tal que $\nabla_x L(x, \lambda, \mu) = 0$. Além disso, a condição de segunda ordem associada a $(\lambda, \mu)$ é satisfeita se $\nabla^2_{xx} L(x, \lambda, \mu)$ é semidefinida positiva em $C^S(x)$.

Observamos que, embora fora do escopo destas notas, uma condição suficiente de otimalidade local estrita é dada independentemente da validade de condições de qualificação, pela existência de um multiplicador de Lagrange associado a $x \in \Omega$ que cumpre a condição de segunda ordem com desigualdade estrita para $d \neq 0$, ou seja, se $\nabla^2_{xx} L(x, \lambda, \mu)$ é definida positiva em $C^S(x) \backslash \{0\}$.

Tipicamente os livros de otimização tratam de provar a validade da condição necessária de segunda ordem em um minimizador sob a hipótese de regularidade. A unicidade do multiplicador, neste caso, simplifica a análise. Neste capítulo apresentaremos outras condições de qualificação que garantem a validade da condição necessária de otimalidade e mostraremos outras condições de otimalidade de segunda ordem mais úteis para a análise de convergência global de algoritmos.

## 3.1 Condições de qualificação de segunda ordem

É interessante observar que mesmo sem assumir nenhuma condição de qualificação é possível descrever uma condição de otimalidade de segunda ordem. O teorema a seguir estende o resultado de Fritz-John (Teorema 1.2) para o caso em que as funções possuem segunda derivada contínua.

**Teorema 3.1** (Condição de Fritz John de segunda ordem)**.** *Seja $x^* \in \Omega$ uma solução local do problema (1.1.1). Então para cada $d \in C^S(x^*)$ existe um vetor não nulo $(\lambda_0, \lambda, \mu) \in \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^p_+$ com $\lambda_0 \geq 0$, $\mu_j = 0$, $j \notin A(x^*)$ tal que*

$$\lambda_0 \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*) = 0, \tag{3.1.1}$$

$$d^{\mathtt{T}} \left( \lambda_0 \nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla^2 g_j(x^*) \right) d \geq 0. \tag{3.1.2}$$

*Demonstração.* No caso em que $\{\nabla h_i(x^*)\}_{i=1}^m$ é linearmente dependente, isto é, existe $\alpha \neq 0$ com $\sum_{i=1}^m \alpha_i \nabla h_i(x^*) = 0$, temos que tanto $(\lambda_0, \lambda, \mu) := (0, \alpha, 0)$ quanto $(\lambda_0, \lambda, \mu) := (0, -\alpha, 0)$ são multiplicadores Fritz-John, ou seja, que cumprem (3.1.1). Portanto, para cada $d \in C^S(x^*)$, podemos escolher uma das duas possibilidades de modo que (3.1.2) se cumpra.

Caso contrário, fixamos $d \in C^S(x^*)$ e consideramos o seguinte problema de otimização linear nas variáveis $z$ e $w$:

$$\begin{array}{ll} \text{Minimizar} & z, \\ \text{Sujeito a} & \nabla h_i(x^*)^{\mathtt{T}} w + d^{\mathtt{T}} \nabla^2 h_i(x^*) d = 0, i = 1, \ldots, m, \\ & \nabla f(x^*)^{\mathtt{T}} w + d^{\mathtt{T}} \nabla^2 f(x^*) d \leq z, \\ & \nabla g_j(x^*)^{\mathtt{T}} w + d^{\mathtt{T}} \nabla^2 g_j(x^*) d \leq z, j \in A(x^*), \nabla g_j(x^*)^{\mathtt{T}} d = 0. \end{array}$$

Como $\{\nabla h_i(x^*)\}_{i=1}^m$ é linearmente independente, podems aplicar o Teorema da Função Implícita, para mostrar que o problema acima possui valor ótimo finito não-negativo (ver detalhes em [33], página 443).

Devido a que o problema anterior é, em realidade, um problema de programação linear, do teorema de dualidade forte, o problema dual abaixo nas variáveis $(\lambda_0, \lambda, \mu)$ com $\mu_j = 0$, $j \notin A(x^*)$, possui o mesmo valor ótimo:

$$\begin{array}{ll} \text{Maximizar} & d^{\mathtt{T}} \left( \lambda_0 \nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla^2 g_j(x^*) \right) d, \\ \text{Sujeito a} & \lambda_0 \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*) = 0, \\ & \lambda_0 + \sum_{i=1}^m \lambda_i + \sum_{j \in A(x^*)} \mu_j = 1, \\ & \lambda_0 \geq 0, \mu_j \geq 0, j \in A(x^*), \nabla g_j(x^*)^{\mathtt{T}} d = 0, \\ & \mu_j = 0, j \in A(x^*), \nabla g_i(x^*)^{\mathtt{T}} d < 0. \end{array}$$

A solução ótima do problema acima é um multiplicador do tipo Fritz John que satisfaz as condições do enunciado. A existência de solução ótima é garantida pelo teorema forte de dualidade, já que o problema primal possui valor ótimo finito. □

Embora fora do escopo destas notas, é interessante observar que se assumirmos uma versão mais forte da tese do teorema acima em um ponto $x^* \in \Omega$, a saber, que a desigualdade vale no sentido estrito quando $d \neq 0$, então obtemos uma condição suficiente do tipo Fritz John para a otimalidade local estrita de $x^*$.

A condição de Fritz-John de segunda ordem possui a propriedade de *não-gap*, isto é, a passagem da condição necessária para a condição suficiente é conseguida através de uma simples mudança do sinal da desigualdade, de maior para maior estrito, considerando o mesmo cone de direções. Por algum tempo, um dos obstáculos para a obtenção de condições de segunda ordem com a propriedade *não-gap* foi que não estava claro qual forma quadrática deveria ser considerada para indicar as condições. Acontece que para propriedades desse tipo, todos os multiplicadores de Lagrange devem ser envolvidos na formulação, justamente como a condição de Fritz-John de segunda ordem faz. Por outro lado, a condição necessária de segunda ordem do tipo Fritz John tem a desvantagem que $\lambda_0$ pode ser igual a zero, desprezando a função objetivo. Este problema pode ser resolvido assumindo certas CQs. Por exemplo, podemos considerar a condição de qualificação de Mangasarian-Fromovitz.

**Teorema 3.2** (Condição de segunda ordem sob MFCQ). *Se $x^* \in \Omega$ é uma solução local que satisfaz MFCQ, então para cada $d \in C^S(x^*)$ existe um multiplicador de Lagrange $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$, $\mu_j = 0, j \notin A(x^*)$ associado a $x^*$ tal que*

$$d^{\mathsf{T}} \left( \nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla^2 f_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla^2 g_j(x^*) \right) d \geq 0.$$

**Prova:** Basta aplicar o teorema anterior e observar que sob MFCQ os multiplicadores do tipo Fritz-John são multiplicadores de Lagrange ($\lambda_0 > 0$). □

A desvantagem da condição anterior é que ela depende do conjunto de multiplicadores de Lagrange, enquanto na prática, tipicamente, temos um candidato $x$ à solução e um candidato $(\lambda, \mu)$ ao multiplicador associado a $x$. Idealmente gostaríamos de definir condições de segunda ordem que possam ser verificadas nesta situação. Observe que sob SMFCQ, ou seja, assumindo a unicidade dos multiplicadores de Lagrange associados a uma solução $x^*$, o teorema anterior diz que $x^*$ satisfaz a condição necessária de segunda ordem da Definição 3.1.

**Teorema 3.3.** *Suponha que $x^* \in \Omega$ é um minimizador local que satisfaz SMFCQ. Logo o único multiplicador de Lagrange $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$, $\mu_j = 0, j \notin A(x^*)$ associado a $x^*$ é tal que a condição necessária (forte) de segunda ordem é satisfeita.*

Nosso objetivo é obter uma versão do teorema anterior com hipóteses menos restritivas. Uma primeira tentativa é utilizar MFCQ, mas o exemplo abaixo descrito em [16] (originalmente proposto em [17]) mostra que existem problemas que cumprem MFCQ na solução, onde a condição necessária de otimalidade de segunda ordem não é satisfeita em nenhum multiplicador:

$$\begin{array}{ll} \text{Minimizar} & x_3, \\ & x_3 \geq (x_1, x_2)^{\mathsf{T}} Q_k (x_1, x_2), \quad k = 0, \ldots, 3, \end{array}$$

onde para cada $k = 0, \ldots, 3.$, a matriz $Q_k$ é definida como

$$Q_k = \begin{pmatrix} \cos(\frac{k\pi}{4}) & -\sin(\frac{k\pi}{4}) \\ \sin(\frac{k\pi}{4}) & \cos(\frac{k\pi}{4}) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} \cos(\frac{k\pi}{4}) & \sin(\frac{k\pi}{4}) \\ -\sin(\frac{k\pi}{4}) & \cos(\frac{k\pi}{4}) \end{pmatrix}.$$

O seguinte exemplo, mais simples é apresentado em [21].

$$\begin{array}{ll} \text{Minimizar} & x_3, \\ & x_3 \geq 2\sqrt{3}x_1x_2 - 2x_2^2, \\ & x_3 \geq x_2^2 - 3x_1^2, \\ & x_3 \geq -2\sqrt{3}x_1x_2 - 2x_2^2. \end{array}$$

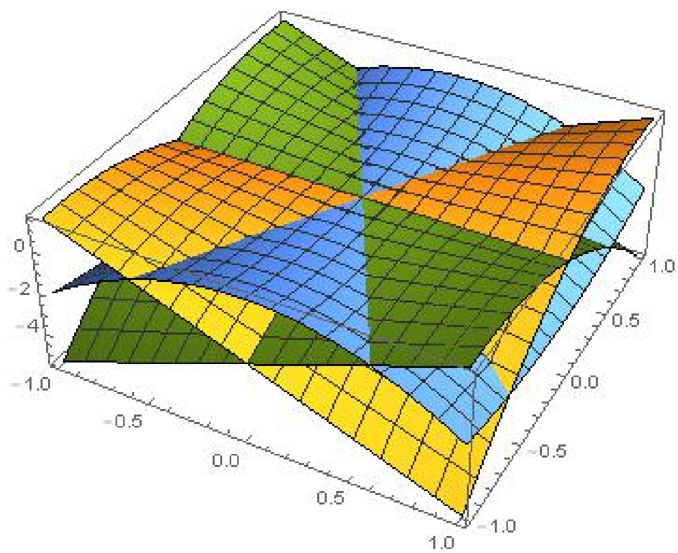



Figura 3.1: MFCQ não garante a validade da condição de otimalidade de segunda ordem

A Figura 3.1 mostra as três superfícies definidas pelas expressões do lado direito das desigualdades do problema. A região viável é formada por todos os pontos acima das três superfícies. Neste caso o gradiente da função objetivo na origem coincide com o oposto dos três gradientes das restrições de modo que os multiplicadores de Lagrange formam o simplex $\mu_1 + \mu_2 + \mu_3 = 1, \mu_1, \mu_2, \mu_3 \geq 0$. O cone crítico é o plano $xy$ e o fato que a condição de segunda ordem não é satisfeita em um mesmo multiplicador se traduz no fato que qualquer combinação convexa das funções que definem as três superfícies é decrescente ao longo de alguma direção do cone crítico.

Frente a isto e de acordo com a Figura 2.8, as únicas condições de qualificação que conhecemos que poderiam garantir a validade da condição de segunda ordem são CRCQ e RCRCQ. De fato, em [6], o resultado é provado sob CRCQ e em [9], foi observado que os resultados de [6] também garantem a validade da condição de segunda ordem sob RCRCQ. É interessante notar que o resultado de [6] garante que a condição de segunda ordem é válida independente da escolha do multiplicador de Lagrange, o que simplifica a sua verificação prática.

**Teorema 3.4.** *Suponha que $x^* \in \Omega$ é um minimizador local que satisfaz RCRCQ. Então qualquer multiplicador de Lagrange $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$, $\mu_j = 0, j \notin A(x^*)$ associado a $x^*$ é tal que a condição necessária (forte) de segunda ordem é satisfeita.*

Para provar este resultado vamos utilizar uma variação do Teorema do Posto Constante ( [76], Teorema 2.9) descrita abaixo:

**Teorema 3.5** (Posto Constante, [53])**.** *Seja $\{f_i, i \in K\}$ uma família finita de funções de classe $C^k, k \geq 1$, $f_i : \mathbb{R}^n \to \mathbb{R}^m$ tais 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 de classe $C^k$, $\phi : U \to V$, com $U$ e $V$ vizinhanças de $x$ tal que $\phi(x) = x$, $\nabla \phi(x)$ é a matriz identidade e $f_i(\phi^{-1}(x + d))$ é constante para todo $d$ com $x + d \in V$ tal que $\nabla f_i(x)^\mathtt{T} d = 0, i \in K$.*

**Lema 3.1** ( [6])**.** *Assuma que $x \in \Omega$ satisfaz RCRCQ. Então para cada $d$ tal que $\nabla h_i(x)^\mathtt{T} d = 0, i = 1, \ldots, m$, $\nabla g_j(x)^\mathtt{T} d \leq 0, j \in A(x)$ existe um função de mesma classe que as $h_i$'s e $g_j$'s, $\xi : (-\varepsilon, \varepsilon) \to \mathbb{R}^n, \varepsilon > 0$ com $\xi(0) = x, \xi'(0) = d$ e $\xi(t) \in \Omega$ para $t \in [0, \varepsilon)$. Além disso, se $\nabla g_j(x)^\mathtt{T} d = 0$ temos $g_j(\xi(t)) = 0$ para todo $t \in (-\varepsilon, \varepsilon)$.*

*Demonstração.* Tome a direção $d$ tal que $d$ satisfaz as condições do enunciado. Defina o conjunto de índices, $K = \{1, \ldots, m\} \cup \{j \in A(x) \mid \nabla g_j(x)^\mathtt{T} d = 0\}$. Como a condição RCRCQ vale podemos usar o Teorema do Posto Constante com a família de funções de classe $C^2$, $\{f_i, i \in K\} := \{h_i, i = 1, \ldots, m\} \cup \{g_j, j \in A(x), g_j(x)^\mathtt{T} d = 0\}$. Assim, existe um difeomorfismo $\phi$ com as propriedades mencionadas.

Defina $\xi(t) = \phi^{-1}(x + td)$ para todo $t \in (-\delta, \delta)$ para algum $\delta > 0$. Utilizando a expansão de Taylor para $\phi^{-1}$ temos $\xi'(0) = \lim_{t \to 0} t^{-1}(\xi(t) - x) = d$. Assim, se $\nabla f_i(x)^\mathtt{T} d = 0$ temos $f_i(\xi(t)) = f_i(\phi^{-1}(x + td)) = f_i(\phi^{-1}(x)) = 0$.

Para mostrar que $\xi(t) \in \Omega$ para $t \geq 0$ basta observar que se $\nabla g_j(x)^\mathtt{T} d < 0$ temos $g_j(\xi(t)) = t \nabla g_j(x)^\mathtt{T} d + o(t)$, onde $o(t)/t \to 0$, logo $g_j(\xi(t)) < 0$ para $t$ suficientemente pequeno. Além disso, se $j \notin A(x)$, reduzindo possivelmente o intervalo temos claramente $g_j(\xi(t)) < 0$. □

*Demonstração.* ( Prova do Teorema 3.4.)
Assuma que $x^* \in \Omega$ satisfaz RCRCQ e seja $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p$ com $\mu_j \geq 0$ qualquer multiplicador de Lagrange associado a $x^*$ (o fato de RCRCQ ser uma condição de qualificação garante que o conjunto dos multiplicadores de Lagrange é não vazio). Seja $d \in C^S(x)$. Então

$$\begin{aligned} &\nabla f(x^*)^\mathtt{T} d = 0, \\ &\nabla h_i(x^*)^\mathtt{T} d = 0, \quad i \in \{1, \ldots, m\}, \\ &\nabla g_j(x^*)^\mathtt{T} d = 0, \quad j \in A(x^*), \mu_j > 0, \\ &\nabla g_j(x^*)^\mathtt{T} d \leq 0, \quad j \in A(x^*), \mu_j = 0. \end{aligned}$$

Definimos $\phi(t) = f(\xi(t))$, onde $\xi : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ é obtida aplicando o Lema 3.1. Já que $\xi(t)$ é viável para $t \geq 0$ e $\xi(0) = x^*$ temos que $t^* = 0$ é um minimizador local de $\phi(t), t \geq 0$. Logo, $\phi''(0) \geq 0$. Como $\xi'(0) = d$ temos $\phi''(0) = d^\mathtt{T} \nabla^2 f(x^*) d + \nabla f(x)^\mathtt{T} \xi''(0) \geq 0$.

Sendo, para todo $t$, $h_i(\xi(t)) = 0, i \in \{1, \ldots, m\}$ e $g_j(\xi(t)) = 0, j \in A(x^*), \mu_j > 0$ temos

$$R(t) = \sum_{i=1}^{m} \lambda_i h_i(\xi(t)) + \sum_{j \in A(x^*)} \mu_j g_j(\xi(t)) = 0, \quad -\varepsilon < t < \varepsilon.$$

Assim, derivando duas vezes temos que $R''(0)$ é igual a

$$d^{\mathrm{T}}(\sum_{i=1}^{m}\lambda_i\nabla^2 h_i(x^*)+\sum_{j\in A(x^*)}\mu_j\nabla^2 g_j(x^*))d+(\sum_{i=1}^{m}\lambda_i\nabla h_i(x^*)+\sum_{j\in A(x^*)}\mu_j\nabla g_j(x^*))^{\mathrm{T}}z=0$$

onde $z := \xi''(0)$. Logo,

$$\phi''(0)+R''(0)=d^{\mathrm{T}}(\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i\nabla^2 h_i(x^*)+\sum_{i\in A(x^*)}\mu_j\nabla^2 g_j(x^*))d\geq 0$$

e a condição de otimalidade de segunda ordem é satisfeita. □

Podemos considerar uma versão mais fraca de RCRCQ, a saber, exigindo posto constante apenas para o conjunto de todos os gradientes de restrições ativas. Isto motiva a seguinte definição

**Definição 3.2.** *Seja $x^*$ um ponto viável. Dizemos que a condição WCR (weak constant rank) vale em $x^*$ se existe uma vizinhança $V$ de $x^*$ tal que o posto de $\{\nabla h_i(x) : i = 1,\dots,m\}\cup\{\nabla g_j(x) : j\in A(x^*)\}$ é constante para todo $x\in V$.*

WCR não é uma condição de qualificação. De fato, considere o problema de minimizar

$$\text{minimize } f(x) := -x_1 \text{ s.a. } h_1(x) := x_2-x_1^2 = 0, g_1(x) := -x_1\leq 0, g_2(x) := x_2\leq 0.$$

A solução $x^* = (0,0)$ não satisfaz as condições KKT, mas existe uma vizinhança da origem tal que o posto de $\{\nabla h_1(x),\nabla g_1(x),\nabla g_2(x)\}$ é constante. Mas, é possível mostrar que sob WCR, caso existam multiplicadores de Lagrange, eles devem satisfazer a chamada condição de otimalidade necessária fraca de segunda ordem.

**Definição 3.3.** *Dizemos que o ponto $x\in\Omega$, que satisfaz as condições KKT, cumpre com a condição necessária fraca de otimalidade de segunda ordem (WSONC) associada ao multiplicador de Lagrange $(\lambda,\mu)\in\mathbb{R}^m\times\mathbb{R}^p_+$, $\mu_j = 0, j\notin A(x)$, quando*

$$d^{\mathrm{T}}\left(\nabla^2 f(x)+\sum_{i=1}^{m}\lambda_i\nabla^2 h_i(x)+\sum_{j\in A(x)}\mu_j\nabla^2 g_j(x)\right)d\geq 0,\forall d\in C^W(x),$$

*onde*

$$C^W(x)=\{d\in\mathbb{R}^n \mid \nabla f(x)^{\mathrm{T}}d=0;\nabla h_i(x)^{\mathrm{T}}d=0, i=1,\dots,m;\nabla g_j(x)^{\mathrm{T}}d=0, j\in A(x)\}$$

*é o cone crítico fraco (subespaço crítico).*

**Teorema 3.6** ( [6]). *Se $x^*\in\Omega$ é um minimizador local que cumpre KKT e vale WCR, então para qualquer multiplicador de Lagrange, $x^*$ satisfaz a condição fraca de otimalidade de segunda ordem.*

**Prova:** Basta observar que o Teorema do Posto Constante pode ser aplicado obtendo uma curva viável $\xi(t)$ com $h_i(\xi(t)) = 0, i = 1,\dots,m$ e $g_j(\xi(t)) = 0, j\in A(x^*)$ e a prova segue como no Teorema 3.4. □

Note que o cone crítico fraco é o maior subespaço vetorial contido no cone crítico forte. Conforme comentaremos na próxima seção, a versão fraca do teorema de segunda ordem é mais interessante do ponto de vista de algoritmos numéricos. Observe que os cones críticos coincidem se vale a complementaridade estrita, isto é, se existe um multiplicador de Lagrange que não se anula em nenhum índice de restrição de desigualdade ativa.

É interessante observar que o exemplo de Anitescu/Arutyunov apresentado mostra também que apenas sob MFCQ, nem mesmo a condição de otimalidade fraca de segunda ordem é satisfeita. Sendo assim, é frequente na literatura a tentativa de assumir MFCQ e alguma outra condição com o objetivo de obter condições de otimalidade de segunda ordem (fracas ou fortes). O inconveniente desta abordagem é que, tipicamente, não é possível mostrar que a condição vale para todos os multiplicadores, mas sim que existe um multiplicador de Lagrange que cumpre a condição. Um primeiro exemplo, apresentado no Teorema 3.3, é a condição SMFCQ que garante a condição forte para o único multiplicador de Lagrange existente. Em [14] é provado que sob MFCQ+WCR a condição de otimalidade fraca de segunda ordem é satisfeita para algum multiplicador. Embora este resultado seja mais fraco que o do Teorema 3.6, em [14] é mostrado que nestas condições, um algoritmo do tipo Lagrangiano Aumentado gera uma sequência cujos pontos limites satisfazem a condição necessária fraca de segunda ordem. Em [21] é provado que existe um multiplicador de Lagrange que cumpre a condição forte de segunda ordem se vale MFCQ e o problema possui duas ou menos variáveis, ou, duas ou menos restrições de desigualdade estão ativas na solução. Em [22] é provada a existência de um multiplicador de Lagrange que cumpre a condição forte de segunda ordem sob MFCQ e se o conjunto de multiplicadores de Lagrange é um segmento de reta, além disso, assume-se que existe no máximo um índice $i_0$ de restrição de desigualdade ativa tal que a componente $i_0$ de qualquer multiplicador de Lagrange é sempre nula. Os autores conjecturam que a última condição não é necessária. Além disso, os autores mostram que se a deficiência do posto dos gradientes das restrições de igualdade e desigualdade ativas é no máximo 1, então o conjunto de multiplicadores de Lagrange é um segmento de reta (possivelmente ilimitado). Em [23] é mostrado que sob MFCQ e assumindo que o problema é convexo (condição de Slater), existe um multiplicador de Lagrange que cumpre a condição forte de segunda ordem. Em [14] conjectura-se que existe um multiplicador de Lagrange que cumpre a condição fraca de segunda ordem sob MFCQ e desde que o aumento do posto em uma vizinhança do ponto esteja limitado em no máximo 1. Outros resultados de segunda ordem (envolvendo condições sobre as derivadas segundas das restrições) são obtidos em [18, 19, 25]. Resultados de segunda ordem para problemas em dimensão infinita podem ser obtidos em [27] e resultados para problemas multi-objetivo em [59, 60].

A seguir descrevemos os resultados de [2].

Consideremos $\Omega = \{x \mid h_i(x) = 0, i = 1, \ldots, m, g_j(x) \leq 0, j = 1, \ldots, p\}$ o conjunto viável do problema original e, fixado $x^* \in \Omega$ com, definimos o conjunto viável auxiliar $\Omega' = \{x \mid h_i(x) = 0, i = 1, \ldots, m, g_j(x) = 0, j \in A(x^*)\}$. Note que $x^* \in \Omega'$ mas não existe, em geral, uma relação de inclusão entre $\Omega$ e $\Omega'$. É claro que a condição de posto constante fraco (WCR) independe se consideramos $x^* \in \Omega$ ou $x^* \in \Omega'$. Vamos mostrar que WCR implica que vale Abadie para $\Omega'$. Nas definições 1.2 e 2.1 vamos denotar o cone tangente por $\mathcal{T}_\Omega(x^*)$ e o cone linearizado por $\mathcal{L}_\Omega(x^*)$ e vamos considerar estes cones definidos em $\Omega$ e $\Omega'$.

**Teorema 3.7.** *Se $x^* \in \Omega$ satisfaz WCR, então $x^*$ satisfaz a condição de Abadie para o conjunto $\Omega'$ descrito pelas igualdades $h_i(x) = 0, i = 1, \dots, m, g_j(x) = 0, j \in A(x^*)$.*

**Prova:** Seja $0 \neq d \in \mathcal{L}_{\Omega'}$, isto é, $\nabla h_i(x^*)^{\mathtt{T}} d = 0$, $i = 1, \dots, m$ e $\nabla g_j(x^*)^{\mathtt{T}} d = 0$, $j \in A(x^*)$. Como na prova do Lema 3.1 e Teorema 3.6, podemos aplicar o Teorema do Posto Constante para construir uma curva $\xi : (-\varepsilon, \varepsilon) \to \Omega', \varepsilon > 0$ com $\xi(0) = x^*, \xi'(0) = d$. Em particular, $d = \lim_{t \to 0^+} (\xi(t) - x^*)/t$. Assim, sendo $\xi'(0) \neq 0$, podemos reduzir $\varepsilon$ se necessário para que $\xi(t) \neq x^*$. Logo

$$\frac{\xi(t) - x^*}{\|\xi(t) - x^*\|} = \frac{\xi(t) - x^*}{t} \frac{t}{\|\xi(t) - x^*\|} \to \frac{d}{\|d\|} \text{ para } t \to 0^+,$$

o que mostra que $d \in \mathcal{T}_{\Omega'}$. □

É importante observar que, como WCR não é uma condição de qualificação, o fato de valer Abadie para $\Omega'$ não é uma condição de qualificação para $x^* \in \Omega$. Vamos provar que a condição de Abadie para $\Omega'$ garante que qualquer multiplicador de Lagrange (caso existam) satisfaz a condição fraca de segunda ordem. Este resultado generaliza o Teorema 3.6. Para isto, provamos primeiramente o lema abaixo:

**Lema 3.2.** *Sejam $x^* \in \Omega$ um minimizador local e $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$, $\mu_j = 0$, $j \notin A(x^*)$ um multiplicador de Lagrange associado a $x^*$. Então*

$$d^{\mathtt{T}} \left( \nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla^2 h_i(x^*) + \sum_{j \in A(x^*)} \mu_j \nabla^2 g_j(x^*) \right) d \geq 0,$$

*para todo $d$ nulo ou tal que existe $x^k \to x^*$ satisfazendo*

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

*e as seguintes relações*

$$h_i(x^k) = o(\|x^k - x^*\|^2) \text{ para todo } i = 1, \dots, m \text{ com } \lambda_i \neq 0$$

$$g_j(x^k) = o(\|x^k - x^*\|^2) \text{ para todo } j = 1, \dots, p \text{ com } \mu_j > 0$$

*Em particular, vale quando $h_i(x^k) = 0$, $i = 1, \dots, m$ e $g_j(x^k) = 0, j \in A(x^*), \mu_j > 0$.*

**Prova:** Temos $L(x^k, \lambda, \mu) = f(x^k) + \sum_{i=1}^{m} \lambda_i h_i(x^k) + \sum_{j \in A(x^*)} \mu_j g_j(x^k) = f(x^k) + o(\|x^k - x^*\|^2)$, assim, sendo $x^*$ um minimizador local e $x^k$ viável, temos

$$\begin{aligned}
&0 \leq f(x^k) - f(x^*) \\
&\quad = L(x^k, \lambda, \mu) - L(x^*, \lambda, \mu) + o(\|x^k - x^*\|^2) \\
&\quad = \nabla_x L(x^*, \lambda, \mu)^{\mathtt{T}} (x^k - x^*) + \frac{1}{2} (x^k - x^*)^{\mathtt{T}} \nabla^2_{xx} L(\bar{x}^k, \lambda, \mu)(x^k - x^*) + o(\|x^k - x^*\|^2) \\
&\quad = \frac{1}{2} (x^k - x^*)^{\mathtt{T}} \nabla^2_{xx} L(\bar{x}^k, \lambda, \mu)(x^k - x^*) + o(\|x^k - x^*\|^2),
\end{aligned}$$

onde $\bar{x}^k$ está no segmento de reta entre $x^*$ e $x^k$. Assim, dividindo por $\|x^k - x^*\|^2$ e tomando o limite em $k$ temos $d^{\mathtt{T}} \nabla L^2_{xx}(x^*, \lambda, \mu) d \geq 0$. □

Note que o conjunto de direções do lema acima é um cone contido no cone tangente.

**Teorema 3.8.** *Seja $x^* \in \Omega$ um minimizador local e assuma que a condição de Abadie é satisfeita em $x^*$ considerando as restrições $h_i(x) = 0, i = 1, \ldots, m, g_j(x) = 0, j \in A(x^*)$. Então a condição de otimalidade fraca de segunda ordem é satisfeita para qualquer multiplicador de Lagrange associado a $x^*$ (caso existam).*

**Prova:** Note que $C^W(x^*) = \mathcal{L}_{\Omega'}(x^*)$, além disso, a hipótese garante que estes cones coincidem com $\mathcal{T}_{\Omega'}(x^*)$, que por sua vez coincide com o cone de direções dado no Lema 3.2 e o resultado segue. □

Quando o conjunto viável possui apenas igualdades, isto é, $\Omega = \Omega'$ com $A(x^*) = \emptyset$, a hipótese do Teorema 3.8 se traduz para a condição de qualificação de Abadie, assim, além de garantir a existência de multiplicadores de Lagrange, todos os multiplicadores satisfazem a condição de segunda ordem (no caso sem desigualdades o cone crítico forte coincide com o cone fraco e as condições forte e fraca são equivalentes).

**Corolário 3.1.** *Suonha que o conjunto viável possui apenas restrições de igualdade e $x^* \in \Omega$ é um minimizador local que satisfaz a condição de Abadie. Então, $x^*$ é um ponto KKT e todos os multiplicadores de Lagrange associados a $x^*$ satisfazem a condição de otimalidade (forte) de segunda ordem.*

A seguir vamos provar um resultado a respeito da condição de otimalidade forte de segunda ordem no conjunto viável $\Omega$ com igualdades e desigualdades. A hipótese será uma condição do tipo Abadie para um subconjunto fixo das restrições tratado como igualdades. Para identificar este subconjunto especial de restrições, consideramos algumas definições e lemas abaixo:

**Definição 3.4.** *Seja $x \in \Omega$ um ponto KKT. Definimos $A^0(x)$ o conjunto de índices $j$ de restrições de desigualdade ativas em $x$ tais que $\mu_j = 0$ para qualquer multiplicador de Lagrange $(\lambda, \mu)$ associado a $x$. O conjunto $A^+(x) = A(x) \backslash A^0(x)$ é o conjunto de índices $j$ de restrições de desigualdade ativas em $x$ tais que $\mu_j > 0$ para algum multiplicador de Lagrange $(\lambda, \mu)$ associado a $x$.*

Sendo $x \in \Omega$ um ponto KKT, para cada $j_0 \in A^+(x)$ existe um multiplicador de Lagrange $(\lambda^{j_0}, \mu^{j_0}) \in \mathbb{R}^m \times \mathbb{R}^p_+$ associado a $x$ tal que $\mu^{j_0}_{j_0} > 0$, assim temos que $(\lambda, \mu) := \frac{1}{|A^+(x)|} \sum_{j \in A^+(x)} (\lambda^j, \mu^j)$ é um multiplicador de Lagrange associado a $x$ tal que $\mu_j > 0$ para todo $j \in A^+(x)$. Portanto o cone crítico forte pode ser descrito como

$$C^S(x) = \left\{ d \in \mathbb{R}^n \mid \begin{array}{l} \nabla g_j(x)^{\mathrm{T}} d = 0, j \in A^+(x); \nabla g_j(x)^{\mathrm{T}} d \leq 0, j \in A^0(x) \\ \nabla h_i(x)^{\mathrm{T}} d = 0, i \in \{1, \ldots, m\} \end{array} \right\}.$$

O lema a seguir caracteriza o conjunto $A^0(x)$.

**Lema 3.3.** *Se $x \in \Omega$ é um ponto KKT, então*

$$A^0(x) = \{ j \in A(x^*) \mid \exists d \in C^S(x), \nabla g_j(x)^{\mathrm{T}} d < 0 \}.$$

**Prova:** Seja $j \in A^0(x) \subset A(x)$. Então o problema de otimização linear abaixo é viável e possui valor ótimo nulo:

$$\begin{array}{ll} \text{Maximizar}_{(\lambda,\mu)} & \mu_j, \\ \text{Sujeito a} & \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x)} \mu_i \nabla g_i(x) = -\nabla f(x), \\ & \mu_i \geq 0, i \in A(x). \end{array}$$

Segue do teorema de dualidade forte para otimização linear que o problema dual abaixo também é viável e com o mesmo valor ótimo:

$$\begin{array}{ll} \text{Minimizar}_d & \nabla f(x)^{\mathtt{T}} d, \\ \text{Sujeito a} & \nabla h_i(x)^{\mathtt{T}} d = 0, i = 1, \ldots, m, \\ & \nabla g_i(x)^{\mathtt{T}} d \leq 0, i \in A(x) \backslash \{j\}, \\ & \nabla g_j(x)^{\mathtt{T}} d \leq -1. \end{array}$$

A solução $d$ deste problema satisfaz as restrições e é tal que $\nabla f(x)^{\mathtt{T}} d = 0$, logo $d \in C^S(x)$ e $\nabla g_j(x)^{\mathtt{T}} d < 0$, o que mostra que $A^0(x) \subset \{j \in A(x^*) \mid \exists d \in C^S(x), \nabla f_j(x)^{\mathtt{T}} d < 0\}$. A inclusão recíproca é óbvia. □

**Corolário 3.2.** *Seja $x \in \Omega$ um ponto KKT. Então existe $d \in C^S(x)$ tal que*

$$\begin{array}{l} \nabla h_i(x)^{\mathtt{T}} d = 0, i \in \{1, \ldots, m\} \\ \nabla g_j(x)^{\mathtt{T}} d = 0, j \in A^+(x), \\ \nabla g_j(x)^{\mathtt{T}} d < 0, j \in A^0(x). \end{array}$$

**Prova:** Basta somar os vetores dados pelo Lema 3.3 para cada $i \in A^0(x)$. □

O teorema a seguir mostra que a condição forte de segunda ordem é satisfeita para qualquer multiplicador de Lagrange (caso existam) quando a condição de Abadie é satisfeita considerando as restrições em $A^+(x)$ como igualdades.

**Teorema 3.9.** *Seja $x^* \in \Omega$ um minimizador local e suponha que a condição de Abadie é satisfeita em $x^*$ considerando as restrições*

$$\Omega^+ = \{x \mid h_i(x) = 0, i \in \{1, \ldots, m\}; g_j(x) = 0, j \in A^+(x^*); g_j(x) \leq 0, j \in A^0(x^*)\}.$$

*Então a condição de otimalidade forte de segunda ordem é satisfeita para qualquer multiplicador de Lagrange associado a $x^*$ (caso existam).*

**Prova:** Seja $0 \neq d \in C^S(x^*)$. Então, temos que $\nabla h_i(x^*)^{\mathtt{T}} d = 0$, $i \in \{1, \ldots, m\}$, $\nabla g_j(x^*)^{\mathtt{T}} d = 0$, $j \in A^+(x^*)$ e $\nabla g_j(x^*)^{\mathtt{T}} d \leq 0$, $j \in A^0(x^*)$. Logo, por hipótese, segue que $d \in \mathcal{T}_{\Omega^+}(x^*)$. Ou seja, existe uma sequência $\{x^k\}$ tal que $x^k \to x^*$, $h_i(x^k) = 0$, $i \in \{1, \ldots, m\}$, $g_j(x^k) = 0$, $j \in A^+(x^*)$, $g_j(x^k) \leq 0, j \in A^0(x^*)$ e $(x^k - x^*)/\|x^k - x^*\| \to d/\|d\|$. Do Lema 3.2 segue que $d^{\mathtt{T}} \nabla^2_{xx} L(x^*, \lambda, \mu) d \geq 0$ para qualquer multiplicador de Lagrange $(\lambda, \mu)$ associado a $x^*$. □

A seguir mostramos que o Teorema 3.9 generaliza o Teorema 3.4.

**Teorema 3.10.** *Se $x^* \in \Omega$ satisfaz RCRCQ, então $x^*$ satisfaz a condição de Abadie para o conjunto de restrições*

$$h_i(x) = 0, i \in \{1, \ldots, m\}; g_j(x) = 0, j \in A^+(x^*); g_j(x) \leq 0, j \in A^0(x^*).$$

**Prova:** É uma adaptação da prova do Teorema 3.7. □

Note que a definição do conjunto $\Omega^+$ só faz sentido quando $x^*$ é um ponto KKT. Neste caso, a hipótese do Teorema 3.9 é uma versão mais restrita da condição de Abadie para $\Omega$, já que ela é satisfeita quando $\mathcal{L}_{\Omega^+}(x^*) \subset \mathcal{T}_{\Omega^+}(x^*)$, sendo que $C^S(x^*) = \mathcal{L}_{\Omega^+}(x^*) = \mathcal{L}_{\Omega}(x^*)$ e $\mathcal{T}_{\Omega^+}(x^*) \subset \mathcal{T}_{\Omega}(x^*)$. Vamos mostrar que a hipótese do Teorema 3.9 pode ser reformulada para desprezarmos as restrições em $A^0(x^*)$ na definição do cone tangente.

**Teorema 3.11.** *Se $x^*$ é um minimizador local, então*

$$C^S(x^*) \subset \mathcal{T}_{\Omega^+}(x^*) \Leftrightarrow C^S(x^*) \subset \mathcal{T}_{F^+}(x^*),$$

*onde* $F^+ = \{x \mid h_i(x) = 0, i \in \{1, \ldots, m\}; g_j(x) = 0, j \in A^+(x^*)\}$.

**Prova:** Sendo $\Omega^+ \subset F^+$, a implicação direta é imediata. Seja $0 \neq d \in C^S(x^*)$ e assuma sem perda de generalidade que $\|d\| = 1$.

Seja $s \in C^S(x^*)$ dado pelo Corolário 3.2 tal que $\nabla h_i(x^*)^\mathtt{T} s = 0$, $i \in \{1, \ldots, m\}$; $\nabla g_j(x^*)^\mathtt{T} s = 0$, $j \in A^+(x^*)$ e $\nabla g_j(x^*)^\mathtt{T} s < 0$, $j \in A^0(x^*)$. Agora proceda a definir $d^k := (d + \frac{1}{k}s)/\|d + \frac{1}{k}s\|$, segue que $d^k \in C^S(x^*)$ com $\nabla h_i(x^*)^\mathtt{T} d^k = 0, i \in \{1, \ldots, m\}, \nabla g_j(x^*)^\mathtt{T} d^k = 0, j \in A^+(x^*)$ e $\nabla g_j(x^*)^\mathtt{T} d^k < 0, j \in A^0(x^*)$ para todo $k$.

A hipótese garante que $d^k \in \mathcal{T}_{F^+}(x^*)$, isto é, existe $x^\ell \to_\ell x^*$ com $h_i(x^\ell) = 0, i \in \{1, \ldots, m\}, g_j(x^\ell) = 0, j \in A^+(x^*)$ tal que $(x^\ell - x^*)/\|x^\ell - x^*\| \to_\ell d^k$. Mostraremos que $g_j(x^\ell) < 0, j \in A^0(x^*)$ para $\ell$ suficientemente grande. De fato, para $j \in A^0(x^*)$ temos $g_j(x^\ell) = \nabla g_j(\bar{x}^\ell)(x^\ell - x^*)$ para algum $\bar{x}^\ell$ entre $x^*$ e $x^\ell$, já que $g_j(x^*) = 0$. Assim, como $\nabla g_j(\bar{x}^\ell) \to \nabla g_j(x^*)$, $(x^\ell - x^*)/\|x^\ell - x^*\| \to_\ell d^k$ e $\nabla g_j(x^*)^\mathtt{T} d^k < 0$ segue que $g_j(x^\ell) < 0$ para $\ell$ suficientemente grande, e portanto $d^k \in \mathcal{T}_{\Omega^+}(x^*)$. Como $d^k \to d$ e $\mathcal{T}_{\Omega^+}(x^*)$ é fechado, segue que $d \in \mathcal{T}_{\Omega^+}(x^*)$. □

Em [23] é apresentado um resultado similar ao Teorema 3.9.

**Teorema 3.12.** *Fixado um ponto $x^* \in \Omega$ que satisfaz KKT, e fixado um multiplicador de Lagrange $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p$ associado a $x^*$, definimos o conjunto de restrições*

$$\Omega_\mu = \left\{ x \mid \begin{array}{l} h_i(x) = 0, i \in \{1, \ldots, m\}; g_j(x) = 0, j \in A(x^*), \mu_j > 0; \\ g_j(x) \leq 0, j \in A(x^*), \mu_j = 0 \end{array} \right\}.$$

*Se a condição de Abadie é satisfeita em $x^*$ com relação a $\Omega_\mu$, então $(\lambda, \mu)$ satisfaz a condição de otimalidade forte de segunda ordem.*

**Prova:** A demonstração é feita exatamente como na prova do Teorema 3.9. □

Observamos que independente de $(\lambda, \mu)$, $\mathcal{L}_{\Omega_\mu}(x^*) = C^S(x^*)$ e que $\Omega^+ \subset \Omega_\mu$. O Teorema 3.12 garante que se vale Abadie para $\Omega_\mu$, então vale a condição de segunda ordem para este multiplicador $(\lambda, \mu)$ e para todos os outros $(\lambda', \mu')$ tais que $\Omega_\mu \subset \Omega_{\mu'}$, ou seja, se o conjunto de índices de multiplicadores positivos de $\mu$ contém o conjunto de índices de multiplicadores positivos de $\mu'$. Se $(\lambda, \mu)$ é tomado com $\mu_j > 0$ para todo $j \in A^+(x^*)$ e a condição do Teorema 3.12 é satisfeita, então $\Omega_\mu \subset \Omega_{\mu'}$ para qualquer multiplicador de Lagrange $(\lambda', \mu')$ associado a $x^*$ e como $\Omega_\mu = \Omega^+$, a consequência é a mesma do Teorema 3.9.

Observamos ainda que se não estamos interessados em obter uma condição de segunda ordem para todos os multiplicadores, é possível que a condição do Teorema 3.12 seja satisfeita para algum multiplicador, mas não para todos. De fato, considere o problema

$$\begin{array}{ll} \text{Minimizar} & x_2, \\ \text{Sujeito a} & g_1(x) := -x_1^2 - x_2 \leq 0, \\ & g_2(x) := -x_2 \leq 0, \\ & g_3(x) := x_1 \leq 0 \end{array}$$

na solução $x^* = (0, 0)$. Temos que os multiplicadores são da forma $\mu_1 \geq 0, \mu_2 \geq 0, \mu_1 + \mu_2 = 1, \mu_3 = 0$. Segue que $A^+(x^*) = \{1, 2\}$ e $A^0(x^*) = \{3\}$. O cone crítico é dado por $C^S(x^*) = \{d \mid d_1 \leq 0, d_2 = 0\}$. Considerando um multiplicador de Lagrange com o número máximo de entradas positivas, por exemplo,

$\mu_1 = \mu_2 = \frac{1}{2}, \mu_3 = 0$ temos $\Omega_\mu = \Omega^+ = \{(0,0)\}$, e portanto o cone tangente a este conjunto não contém $C^S(x^*)$. Entretanto, considerando $\mu = (0,1,0)$ temos $\Omega_\mu = \{x \mid x_1 \leq 0, x_2 = 0\}$ e portanto $C_S(x^*) \subset \mathcal{T}_{\Omega_\mu}(x^*)$ e podemos garantir a validade da condição de otimalidade forte de segunda ordem apenas para $\mu$.

Em [2], também foi mostrado que se um minimizador local $x^* \in \Omega$ satisfaz MFCQ e para todo $\mathcal{I} \subset \{1,\ldots,m\}$, $\mathcal{J} \subset A^+(x^*)$ com cardinalidade $|\mathcal{I}| + |\mathcal{J}|$ um a menos que $\{1,\ldots,m\} \cup A^+(x^*)$ vale que $\{\nabla h_i(x) : i \in \mathcal{I}\} \cup \{\nabla g_j(x) : j \in \mathcal{J}\}$ tem posto constante em torno de $x^*$, então a condição fraca de segunda ordem vale para pelo menos um multiplicador de Lagrange associado a $x^*$. Além disso, se $A^0(x^*)$ tem no máximo um índice, então vale a condição forte de segunda ordem para pelo menos um multiplicador de Lagrange.

## 3.2 Condições sequenciais de segunda ordem

As condições sequenciais de otimalidade têm demonstrado ser uma ferramenta útil no estudo de algoritmos práticos. Por exemplo, têm sido usadas na análise de convergência global de algoritmos, como os métodos de lagrangiano aumentado, os métodos de restauração inexata, e alguns métodos de programação quadrática sequencial sob condições de qualificação bem mais fracas que as usuais. Além disso, elas fornecem um fundamento teórico para os diferentes critérios de parada associados a algoritmos práticos e tem servido como modelo no desenho de novos algoritmos práticos, [32, 34, 35].

Sob este paradigma, vamos definir, de acordo com [8], condições sequenciais de otimalidade apropriadas para o estudo de algoritmos com convergência a pontos estacionários de segunda ordem, que desempenhem o mesmo papel de unificação que AKKT, e outras condições sequenciais de primeira ordem, o fazem na análise de convergência de muitos algoritmos.

Muitos algoritmos com convergência a pontos estacionários de segunda ordem (i.e. pontos onde WSONC vale) têm sido propostos na literatura. Andreani, Birgin, Martinez e Schuverdt [5], ver também [14], usaram um método de curvatura negativa de segunda ordem para minimização em caixas aplicados a certa classe de funções que não possuem segundas derivadas contínuas. Byrd, Schnabel e Schultz [36] empregaram um método de programação quadrática sequencial (SQP), onde a convergência a pontos estacionários de segunda ordem é obtida usando correções de segunda ordem. Coleman, Liu e Yuan [38] usaram também um método SQP com uma função de penalidade quadrática para minimização com restrições de igualdade. Conn, Gould, Orban e Toint [39] utilizaram um método de barreira logarítmica para problemas de minimização com desigualdades e com restrições de igualdade lineares. Di Pillo, Lucidi e Palagi [41] definiram um algoritmo primal-dual para problemas de minimização com restrições de desigualdade e tomaram vantagem da equivalência entre o problema original com restrições e a minimização sem restrições de uma função de Lagrangiano aumentado. Eles usaram uma técnica de busca curvilinear usando a informação sobre a não convexidade dessa função de lagrangiano aumentado. Facchinei e Lucidi [42] usaram as direções de curvatura negativa no contexto de minimização com restrições de desigualdade. Gill, Kungurtsev e Robinson em [46] utilizaram uma variante do método SQP, o método SQP regularizado, definido em [47]. Esse método é baseado em uma busca linear flexível junto com uma direção definida através de uma solução de um subproblema de programação quadrática de uma função estritamente convexa e, quando existe, uma direção de curvatura negativa para certo lagrangiano aumentado primal-dual. Em [67], Morguerza e Prieto usaram um algoritmo de pontos interiores para problemas não convexos juntamente com direções de curvatura negativa.

Nossa contribuição nesta seção é descrever a introdução da primeira condição sequencial de otimalidade que leva em consideração não só a informação de primeira ordem, mas sim, também, a informação de segunda ordem (ver detalhes em [8]). Esta condição é chamada de AKKT2. Definiremos a condição AKKT2 e mostraremos que ela é uma genuína condição de otimalidade, e que vários algoritmos com convergência a pontos estacionários de segunda ordem geram sequências cujos pontos limites cumprem esta condição. Esta condição de otimalidade é forte, no sentido, que ela implica a proposição "WSONC ou não-CQ" para CQs mais fracas que a condição de qualificação LICQ. Na seção seguinte definimos uma nova condição chamada CCP2, que é a CQ mínima associada à AKKT2 e em seguida mostramos as relações existentes entre CCP2 e outras CQs conhecidas na literatura, como por exemplo MFCQ+WCR, RCRCQ e a condição de Baccari-Trad.

Destacamos o seguinte lema de grande utilidade nas demonstrações desta seção.

**Lema 3.4.** *[28] Seja $P \in \mathbb{R}^{n\times n}$ uma matriz simétrica e vetores $a_1, \ldots, a_r \in \mathbb{R}^n$. Defina o seguinte subespaço: $\mathcal{C} := \{d \in \mathbb{R}^n : \langle a_j, d\rangle = 0$ para todo $j \in \{1, \ldots, r\}\}$. Suponha que $v^{\mathtt{T}}Pv > 0$ para todo $v \in \mathcal{C}$. Então, existem $\{c_j \in \mathbb{R}_+,\ j \in \{1, \ldots, r\}\}$ tais que $P + \sum_{j=1}^{r} c_j a_j a_j^T$ é definida positiva.*

Toda condição sequencial de otimalidade deve satisfazer as seguintes três propriedades:

1. Ela deve ser realmente uma condição de otimalidade, independente de qualquer condição de qualificação;
2. ela deve ser o mais forte possível, em nosso caso, ela deve implicar "WSONC ou não-CQ" para condições de qualificação fracas;
3. ela deve ser gerada por algoritmos práticos.

Na seção seguinte mostraremos que o método de lagrangiano aumentado de [5] e o método de regiões de confiança de [40] geram sequências onde os pontos limites satisfazem essa nova condição sequencial de otimalidade.

Procederemos a definir a condição sequencial de segunda ordem.

**Definição 3.5.** *Dizemos que um ponto viável $x^* \in \Omega$ é um ponto 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 $x^k \to x^*$, $\delta_k \to 0$,*

$$\lim_{k\to\infty} \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k) = 0 \tag{3.2.3}$$

*e*

$$\nabla^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^{m} \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^{\mathtt{T}} + \sum_{j\in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^{\mathtt{T}} \succeq -\delta_k \mathbb{I} \tag{3.2.4}$$

*para $k \in \mathbb{N}$ suficientemente grande, onde $\mathcal{A} \succeq \mathcal{B}$ significa que as matrizes simétricas $\mathcal{A}$ e $\mathcal{B}$ são tais que $\mathcal{A} - \mathcal{B}$ é semi-definida positiva e $\mathbb{I}$ denota a matriz identidade.*

A sequência $\{x^k\}$ é chamada de sequência AKKT2 e $x^*$ é chamado de ponto AKKT2.

Para provar que a condição AKKT2 é uma condição de otimalidade usaremos o seguinte lema.

**Lema 3.5** ( [5, 14]). *Seja $\bar{f} : \mathbb{R}^n \to \mathbb{R}$, $\bar{g}_j : \mathbb{R}^n \to \mathbb{R}$ para $j \in \{1, \ldots, p\}$ funções duas vezes continuamente diferenciáveis em uma vizinhança de $\bar{x}$. Defina*

$$\bar{F}(x) := \bar{f}(x) + \frac{1}{2}\sum_{j=1}^{p} \max\{0, \bar{g}_j(x)\}^2$$

*para todo $x$ em uma vizinhança aberta de $\bar{x}$. Suponha que $\bar{x}$ é uma minimizador local de $\bar{F}$. Então a matriz definida como*

$$H(x) := \nabla^2 \bar{f}(x) + \sum_{j=1}^{p} \max\{0, \bar{g}_j(x)\}\nabla^2 \bar{g}_j(x) + \sum_{j:\bar{g}(\bar{x})\geq 0} \nabla \bar{g}_j(x)\nabla \bar{g}_j(x)^{\mathtt{T}}$$

*é semidefinida positiva em $\bar{x}$.*

**Teorema 3.13.** *Seja $x^*$ um minimizador local, então $x^*$ satisfaz a condição AKKT2.*

*Demonstração.* Devido a que $x^*$ é um minimizador local, existe um escalar $\varepsilon > 0$ tal que $f(x^*) \leq f(x)$ para todo $x$ viável tal que $\|x - x^*\| \leq \varepsilon$. Assim, $x^*$ é a única solução de

$$\text{minimizar } f(x) + \frac{1}{4}\|x - x^*\|^4 \ \text{ sujeito a } h(x) = 0, \ \ g(x) \leq 0, \ \ x \in \mathbb{B}(x^*, \varepsilon). \tag{3.2.5}$$

Seja $\{\rho_k\} \subset \mathbb{R}_+$ tal que $\rho_k \uparrow \infty$. Para cada $\rho_k$, considere o seguinte problema

$$\begin{array}{ll} \min & f(x) + \frac{1}{4}\|x - x^*\|^4 + \frac{1}{2}\rho_k(\sum_{i=1}^{m} h_i(x)^2 + \sum_{j=1}^{p} \max\{g_j(x), 0\}^2) \\ \text{s.a.} & x \in \mathbb{B}(x^*, \varepsilon). \end{array} \tag{3.2.6}$$

Seja $x^k$ solução do subproblema (3.2.6). Seguindo a mesma linha de raciocínio da demonstração do Teorema 2.7, temos que $x^k$ está bem definida, $x^k \to x^*$ e para $k$ suficientemente grande $x^k$ pertence ao interior de $\mathbb{B}(x^*, \varepsilon)$. Desta forma, o gradiente da função objetivo de (3.2.6) deve ser zero em $x = x^k$:

$$\nabla f(x^k) + \sum_{i=1}^{m} \rho_k h_i(x^k)\nabla h_i(x^k) + \sum_{j=1}^{p} \rho_k g_j^+(x^k)\nabla g_j(x^k) + \xi(x^k)(x^k - x^*) = 0 \tag{3.2.7}$$

onde $g_j^+(x^k) := \max\{0, g_j(x^k)\}$ e $\xi(x^k) := \|x^k - x^*\|^3$. Aplicando o Lema 3.5 com $\bar{F}(x) = f(x) + \frac{1}{2}\rho_k \sum_{i=1}^{m} h_i(x)^2 + \frac{1}{4}\|x - x^*\|^4$, $\bar{g}_j = \sqrt{\rho_k} g_j(x)$ para $j \in \{1, \ldots, p\}$ e $\bar{x} = x^k$, obtemos que:

$$\begin{aligned} \nabla^2 f(x^k) + \sum_{i=1}^{m} \rho_k h_i(x^k)\nabla^2 h_j(x^k) + \sum_{j=1}^{p} \rho_k \max\{g_j(x^k), 0\}\nabla^2 g_j(x^k) + \\ \sum_{i=1}^{m} \rho_k \nabla h_i(x^k)\nabla h_j(x^k)^T + \sum_{j:g_j(x^k)\geq 0} \rho_k \nabla g_j(x^k)\nabla g_j(x^k)^T + \\ 2(x^k - x^*)(x^k - x^*)^{\mathtt{T}} + \|x^k - x^*\|^2 \mathbb{I} \succeq 0 \end{aligned} \tag{3.2.8}$$

Defina $\lambda_i^k := \rho_k h_i(x^k)$, $\eta_i^k := \rho_k$ para $i \in \{1, \ldots, m\}$, $\mu_j^k := \rho_k \max\{g_j(x^k), 0\}$ para $j \in \{1, .., p\}$, $\theta_j^k = \rho_k$ sempre que $g_j(x^k) \geq 0$ e $\theta_j^k = 0$ no caso contrário. Finalmente defina $\delta_k := 3\|x^k - x^*\|^2$. Obviamente $\delta_k \to 0$ e para $k$ suficientemente grande $\mu_j^k = 0$ para todo $j \notin A(x^*)$ e $\theta_j^k = 0$ para todo $j \notin A(x^*)$. Com essas escolhas e das expressões (3.2.7) e (3.2.8), concluímos que as expressões (3.2.3) e (3.2.4) são satisfeitas e portanto $x^*$ satisfaz a condição AKKT2. □

## 3.3 Condição de qualificação mínima e algoritmos

Em [49], é considerado o problema quadrático em $\mathbb{R}^4$

$$\text{minimizar } f(x) := \frac{1}{2} x^{\mathsf{T}} H x, \text{ sujeito a } x \geq 0,$$

onde $H = \mathbb{I} - \frac{3}{2}\frac{zz^{\mathsf{T}}}{\|z\|^2}$ e $z = e - 4e_1$, sendo $e$ o vetor de 1's e $e_1$ o primeiro vetor da base canônica. Para qualquer sequência de parâmetros de barreira $\mu_k \to 0^+$, definimos as funções $b_k(x) = f(x) - \mu_k \sum_{i=1}^4 \log(x_i)$ (barreira logarítmica). Observamos que $x_k = \sqrt{\mu_k} e$ é um minimizador local estrito do problema de minimizar $b_k(x)$, sujeito a $x > 0$ que satisfaz a condição suficiente de segunda ordem. Como esperado, $x_k$ converge para zero, um ponto estacionário, entretanto, a condição de otimalidade forte de segunda ordem não é satisfeita já que $e_1^{\mathsf{T}} H e_1 < 0$. Este exemplo mostra que, na prática, não se espera que um algoritmo razoável tenha garantia de gerar uma sequência cujos pontos limites satisfazem a condição forte de segunda ordem (SSONC). De fato, a mera verificação da condição forte de segunda ordem é um problema NP-difícil [68]. Por estes motivos, em se tratando de algoritmos práticos, a condição de otimalidade de segunda ordem de interesse é a condição fraca (WSONC). Há diversos algoritmos na literatura que com pequenas modificações têm convergência global a pontos estacionários de segunda ordem: Lagrangiano aumentado, programação quadrática sequencial, pontos interiores, regiões de confiança, entre outros. Nossa tarefa será unificar a teoria de convergência global de segunda ordem destas classes de algoritmos usando a condição AKKT2 como ferramenta.

A condição AKKT2 é uma condição de otimalidade forte no sentido que implica "WSONC ou não-CQ" para condições de qualificação fracas. Iniciaremos provando que, sob a condição MFCQ+WCR, todo ponto AKKT2 satisfaz a condição WSONC, ver Proposição 3.1. Posteriormente, provaremos que a condiçao CRCQ e a condição RCRCQ podem também ser usadas na expressão "WSONC ou não-CQ", ver Proposição 3.3. Como consequência a condição RCRCQ também pode ser usada na análise de convergência de algoritmos práticos.

Para provar que sob a condição MFCQ+WCR, todo ponto AKKT2 satisfaz a condição WSONC, começaremos com o seguinte lema. A seguinte propriedade de WCR é importante para as demonstrações futuras.

**Lema 3.6.** *[14] Seja $d \in C^{\mathcal{W}}(x^*)$, o cone crítico fraco da Definição 3.3. Então para toda sequência $\{x^k\} \subset \mathbb{R}^n$ com $x^k \to x^*$, existe uma sequência $\{d^k\} \subset \mathbb{R}^n$ com $d^k \to d$ tal que, $\langle \nabla h_i(x^k), d^k \rangle = 0$ para $i \in \{1, \ldots, m\}$ e $\langle \nabla g_j(x^k), d^k \rangle = 0$ para $j \in A(x^*)$.*

**Proposição 3.1.** *Seja $x^* \in \Omega$ tal que a condição AKKT2 vale. Se a condição de qualificação MFCQ+WCR se cumpre em $x^*$ então WSONC é satisfeita em $x^*$.*

*Demonstração.* Pela definição de AKKT2 existem sequências $\{x^k\}$, $\{\mu^k\}$, $\{\delta_k\}$, $\{\theta^k\}$ com $\mu_j^k = 0$ e $\theta_j^k = 0$ para todo $j \notin A(x^*)$ tais que $x^k \to x^*$, $\delta_k \to 0$ e

a) $\varepsilon^k := \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k) \to 0$;

b) $\nabla_x^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^m \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j\in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succeq -\delta_k \mathbb{I}$

Pela condição de Mangasarian-Fromovitz, a sequência $\{(\lambda^k, \mu^k)\}$ é limitada. Para provar isso, suponha por contradição que a sequência $\{(\lambda^k, \mu^k)\}$ é ilimitada. Defina a sequência $\{A_k\}$ como $A_k := \max\{|\lambda_i^k|, \mu_j^k : i \in \{1, \ldots, m\}, j \in A(x^*)\}$. Claramente, $\{A_k\}$ vai para infinito. Dividindo por $A_k$ a expressão

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

e tomando uma subsequência adequada tal que $\{(\lambda^k/A_k, \mu^k/A_k)\}$ converge, temos que $\sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j\in J(x^*)} \mu_j^* \nabla g_j(x^*) = 0$ para certos $\lambda^*$ e $\mu^* \geq 0$ não todos nulos. Uma contradição com MFCQ. Portanto, a sequência $\{(\lambda^k, \mu^k)\}$ é limitada. Assim, sem perda de generalidade podemos supor que $\{(\lambda^k, \mu^k)\}$ converge para $(\mu^*, \lambda^*)$. Defina $\mu_j^* = 0$ para todo $j \notin A(x^*)$. Tomando limite no item (a), temos que $(x^*, \mu^*, \lambda^*)$ satisfaz as condições KKT. Procederemos a provar que WSONC vale em $x^*$ com multiplicadores $(\mu^*, \lambda^*)$.

Seja $d \in C^{\mathcal{W}}(x^*)$, pelo Lema 3.6, existe uma sequência $d^k$ tal que $d^k \to d$, $\langle \nabla h_i(x^k), d^k \rangle = 0$, $i \in \{1, \ldots, m\}$ e $\langle \nabla g_j(x^k), d^k \rangle = 0$, $j \in A(x^*)$. Assim, calculando $(d^k)^{\mathtt{T}} M^k (d^k)$ e usando o item (b) obtemos que:

$$\begin{aligned}(d^k)^{\mathtt{T}} \nabla_x^2 L(x^k, \lambda^k, \mu^k)(d^k) + \sum_{i=1}^m \eta_i^k \langle \nabla h_i(x^k), d^k \rangle^2 + \\ \sum_{j\in A(x^*)} \theta_j^k \langle \nabla g_j(x^k), d^k \rangle^2 \geq -\delta_k \|d^k\|^2\end{aligned} \tag{3.3.9}$$

o que implica que

$$(d^k)^{\mathtt{T}} \nabla_x^2 L(x^k, \lambda^k, \mu^k)(d^k) \geq -\delta_k \|d^k\|^2. \tag{3.3.10}$$

Finalmente, tomando limite em (3.3.10), temos que

$$(d^*)^{\mathtt{T}} (\nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(x^*) + \sum_{j\in A(x^*)} \mu_j^* \nabla^2 g_j(x^*))(d^*) \geq 0, \tag{3.3.11}$$

como queríamos provar. □

Claramente da definição, a condição AKKT2 implica a condição sequencial AKKT. O seguinte exemplo mostra que as condições não são equivalentes.

**Exemplo 3.1.** *Considere $f(x_1, x_2) = -x_1 - x_2$, $g(x_1, x_2) = x_1^2 x_2^2 - 1$ e $x = (1, 1)$.*

Primeiro, mostraremos que AKKT vale em $x^* = (1, 1)$. Como os gradientes das funções são $\nabla f(x_1, x_2) = (-1, -1)$ e $\nabla g(x_1, x_2) = 2x_1 x_2 (x_2, x_1)$, as condições KKT valem em $x^*$ com $\mu = 1/2$ como multiplicador.

Agora, provaremos que AKKT2 falha. Suponha pelo contrário que não, assim (3.2.3) e (3.2.4) se cumprem. Defina $d^k := (x_1^k, -x_2^k)$ onde $\{(x_1^k, x_2^k)\}$ é a sequência dada pela definição de AKKT2. Como $\langle \nabla g(x_1^k, x_2^k), d^k \rangle = 0$ para todo $k \in \mathbb{N}$, segue de (3.2.4) que:

$$\mu^k \nabla^2 g(x_1^k, x_2^k)(d^k, d^k) + \delta_k \|d^k\|^2 \geq 0, \tag{3.3.12}$$

para algum $\mu^k \geq 0$, $\delta_k \geq 0$ com $\delta_k \to 0$. Do cálculo temos que $\nabla^2 g(x_1^k, x_2^k)(d^k, d^k) = -4(x_1^k x_2^k)^2$. Assim, substituindo em (3.3.12), temos $-4\mu^k(x_1^k x_2^k)^2 + \delta_k \|d^k\|^2 \geq 0$ para todo $k \in \mathbb{N}$. Mas isso é impossível porque por (3.2.3) temos que $\nabla f(x_1^k, x_2^k) + \mu^k \nabla g(x_1^k, x_2^k) \to 0$, o que implica que $-1 + 2\mu^k x_1^k (x_2^k)^2 \to 0$ e como consequência $2\mu^k(x_1^k x_2^k)^2 \to 1$ e $0 \leq -4\mu^k(x_1^k x_2^k)^2 + \delta_k \|d^k\|^2 \to -2$.

Continuando com nossa análise da condição AKKT2, mostraremos que existem algoritmos que naturalmente geram sequências cujos pontos limites satisfazem a condição AKKT2. Como exemplo, podemos mencionar o lagrangiano aumentado proposto em [5](ver também [14]), o método regularizado SQP de Gill, Kungurtsev e Robinson [46] e o método de regiões de confiança de Dennis e Vicente [40].

Mostraremos que o método de lagrangiano aumentado proposto por [5] (ver [14]) gera sequências AKKT2. Antes de começar a análise do método de lagrangiano aumentado de [14] para o problema geral de otimização (que é equivalente ao método proposto em [5] quando a restrição de caixa é todo o $\mathbb{R}^n$) vejamos algumas notações adicionais. Considere a seguinte função lagrangiana aumentada

$$L_\rho(x, \lambda, \mu) := f(x) + \frac{\rho}{2}\left(\sum_{i=1}^{m}\left[h_i(x) + \frac{\lambda_i}{\rho}\right]^2 + \sum_{j=1}^{p}\left[\max\{0, g_j(x) + \frac{\mu_j}{\rho}\}\right]^2\right), \tag{3.3.13}$$

para todo $x \in \mathbb{R}^n, \rho > 0$ e $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p_+$. A função $L_\rho$ possui primeiras derivadas contínuas com respeito à variável $x$, mas a segunda derivada não está definida nos pontos $x$ tais que $g_j(x) + \mu_j/\rho = 0$. Assim, para superar este problema, em [14], os autores definiram:

$$\begin{array}{l} \bar{\nabla}^2\left(\max\{0, g_i(x) + \frac{\mu_i}{\rho}\}\right)^2 := \nabla^2\left(g_i(x) + \frac{\mu_i}{\rho}\right)^2, \text{ se } g_i(x) + \frac{\mu_i}{\rho} = 0; \\ \bar{\nabla}^2\left(\max\{0, g_i(x) + \frac{\mu_i}{\rho}\}\right)^2 := \nabla^2\left(\max\{0, g_i(x) + \frac{\mu_i}{\rho}\}\right)^2, \text{ c.c.} \end{array} \tag{3.3.14}$$

Agora procederemos a descrever o algoritmo [14, Algoritmo 4.1]. Ver Algoritmo 3.1.

**Algoritmo 3.1 (Método de Lagrangiano aumentado de segunda ordem).** *Sejam* $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, $\gamma > 1$, $\rho_1 > 0$ *e* $\tau \in (0,1)$. *Tome* $\{\varepsilon_k\} \subset \mathbb{R}_+$ *uma sequência de escalares positivos tal que* $\lim \varepsilon_k = 0$.
*Defina* $\lambda_i^1 \in [\lambda_{\min}, \lambda_{\max}], i \in \{1, \ldots, m\}$ *e* $\mu_j^1 \in [0, \mu_{\max}], j \in \{1, \ldots, p\}$. *Escolha* $x^0 \in \mathbb{R}^n$ *como ponto inicial arbitrário. Defina* $V^0 = \max\{0, g(x^0)\}$. *Inicialize com* $k = 1$.

1. *(Minimização)*
   *Encontre um minimizador aproximado* $x^k$ *de* $L_{\rho_k}(x, \lambda^k, \mu^k)$ *até segunda ordem. Isto é, determine* $x^k$ *que satisfaça:*

$$\|\nabla L_{\rho_k}(x^k, \lambda^k, \mu^k)\| \leq \varepsilon_k$$

   *e*

$$\bar{\nabla}^2 L_{\rho_k}(x^k, \lambda^k, \mu^k) \succeq -\varepsilon_k \mathbb{I}$$

2. *(Atualização do parâmetro de penalidade)*

*Defina para todo* $j \in \{1, \ldots, p\}$

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

*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*
   $\hat{\lambda}_i^k := \lambda_i^k + \rho_k h_i(x^k)$, $i = 1, \ldots, m$ *e* $\hat{\mu}_j^k := \max\{0, \mu_j^k + \rho_k g_j(x^k)\}$, $j = 1, \ldots, p$.
   *Defina* $\lambda_i^{k+1} := proj_{[\lambda_{\min}, \lambda_{\max}]}(\hat{\lambda}_i^k) \in [\lambda_{\min}, \lambda_{\max}]$ *para todo* $i \in \{1, \ldots, m\}$ *e* $\mu_j^{k+1} := proj_{[0, \mu_{\max}]}(\hat{\mu}_j^k) \in [0, \mu_{\max}]$*, para todo* $j \in \{1, \ldots, p\}$*. Faça* $k \leftarrow k + 1$ *e vá para o Passo 1.*

Do passo 1 do Algoritmo 3.1 temos que qualquer ponto limite $\{x^k\}$ satisfaz a condição AKKT2. De fato, seja $x^*$ qualquer ponto limite da sequência $\{x^k\}$. Pela demonstração de [14, Teorema 4.1], temos que para $k$ suficientemente grande o Passo 1 do Algoritmo 3.1 é equivalente a:

$$\|\nabla L(x^k, \hat{\lambda}^k, \hat{\mu}^k)\| \leq \varepsilon_k \tag{3.3.15}$$

e

$$\nabla^2 L(x^k, \hat{\lambda}^k, \hat{\mu}^k) + \rho_k \sum_{i=1}^{m} \nabla h_i(x) \nabla h_i(x)^T + \rho_k \sum_{j \in A(x^*)} \nabla g_j(x) \nabla g_j(x)^T \succeq -\varepsilon_k \mathbb{I}, \tag{3.3.16}$$

onde $\hat{\lambda}_i^k := \lambda_i^k + \rho_k h_i(x^k)$ para $i \in \{1, \ldots, m\}$ e $\hat{\mu}_j^k := \max\{0, \mu_j^k + \rho_k g_j(x^k)\}$ para $j \in \{1, \ldots, p\}$.

Não é difícil ver que para $k$ suficientemente grande, $\hat{\mu}_j^k = 0$ para todo $j \notin A(x^*)$. Assim, das expressões (3.3.15) e (3.3.16) temos que a condição AKKT2 vale em $x^*$.

O método de região de confiança proposto por Dennis e Vicente [40] também gera sequências AKKT2. Veja [72] para detalhes.

A análise da convergência global para pontos estacionários de segunda ordem do método de lagrangiana aumentado [14] e do método SQP regularizado [46] estão baseados na condição conjunta MFCQ e WCR. Desde que ambos métodos geram sequências AKKT2, a pergunta natural é se a Proposição 3.1 pode ser demonstrada usando condições de qualificação mais fracas. Como veremos nesta seção, a resposta é afirmativa. Isto implica uma melhora na teoria da convergência global de cada algoritmo que gera sequências AKKT2. Ainda mais, nesta seção, forneceremos a condição de qualificação mínima para a qual todo ponto AKKT2 satisfaz WSONC.

Defina para cada $x \in \mathbb{R}^n$, o cone

$$C^{\mathcal{W}}(x, x^*) := \left\{ d \in \mathbb{R}^n \mid \begin{array}{l} \langle \nabla h_i(x), d \rangle = 0 \text{ para } i \in \{1, \ldots, m\}; \\ \langle \nabla g_j(x), d \rangle = 0 \text{ para } j \in A(x^*)\}. \end{array} \right\} \tag{3.3.17}$$

O cone $C^{\mathcal{W}}(x, x^*)$ é uma pertubação do cone crítico fraco $C^{\mathcal{W}}(x^*)$ ao redor do ponto viável $x^* \in \Omega$. Claramente, $C^{\mathcal{W}}(x, x^*)$ é um subespaço e $C^{\mathcal{W}}(x^*, x^*)$ coincide com o cone crítico fraco $C^{\mathcal{W}}(x^*)$. Usando a linguagem da análise variacional, podemos reescrever o Lema 3.6 como: WCR implica a semi-continuidade interior da multifunção $C^{\mathcal{W}}(x, x^*)$ em $x = x^*$, isto é, $C^{\mathcal{W}}(x^*, x^*) \subset \liminf_{x \to x^*} C^{\mathcal{W}}(x, x^*)$. De fato, a semi-continuidade interior de $C^{\mathcal{W}}(x, x^*)$ em $x = x^*$ é equivalente a WCR. Ver [72].

Agora vamos definir o objeto principal para a definição da condição mínima. Para cada $x \in \mathbb{R}^n$ considere o seguinte cone

$$K_2^{\mathcal{W}}(x) := \bigcup_{\substack{(\lambda,\mu)\in\mathbb{R}^m\times\mathbb{R}^p_+,\\ \mu_j=0 \text{ para } j\notin A(x^*)}} S^{\mathcal{W}}(\lambda, \mu, x), \tag{3.3.18}$$

onde

$$S^{\mathcal{W}}(\lambda, \mu, x) := \left\{ \begin{array}{l} (\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j\in A(x^*)} \mu_j \nabla g_j(x), H), \text{ tal que} \\ H \preceq \sum_{i=1}^m \lambda_i \nabla^2 h_i(x) + \sum_{j\in A(x^*)} \mu_j \nabla^2 g_j(x) \text{ sobre } C^W(x, x^*) \end{array} \right\}.$$

O conjunto $K_2^{\mathcal{W}}(x)$ é um cone convexo contido em $\mathbb{R}^n \times \mathrm{Sym(n)}$, que nos permite escrever a condição fraca de segunda ordem WSONC em uma forma mais compacta, isto é,

$$(-\nabla f(x^*), -\nabla^2 f(x^*)) \in K_2^{\mathcal{W}}(x^*). \tag{3.3.19}$$

A seguinte definição é a nossa nova condição de qualificação associada à condição sequencial AKKT2.

**Definição 3.6.** *Dizemos que $x^* \in \Omega$ satisfaz a Propriedade de Cone Contínuo de Segunda Ordem (CCP2) se a multifunção $x \mapsto K_2^{\mathcal{W}}(x)$, definida em* (3.3.18)*, é semicontínua exteriormente em $x^*$, isto é,*

$$\limsup_{x\to x^*} K_2^{\mathcal{W}}(x) \subset K_2^{\mathcal{W}}(x^*). \tag{3.3.20}$$

Como mostra o teorema a seguir, a condição CCP2 é a condição de qualificação mais fraca que pode ser usada para generalizar a Proposição 3.1.

**Teorema 3.14.** *Seja $x^* \in \Omega$ um ponto viável. As seguintes condições são equivalentes:*

*(1) CCP2 vale em $x^*$;*

*(2) Para toda função objetivo $f : \mathbb{R}^n \to \mathbb{R}$ tal que a condição AKKT2 vale em $x^*$, temos que a condiçao WSONC é satisfeita em $x^*$.*

*Demonstração.* Primeiro, suponha que a condição CCP2 vale em $x^*$ e existe uma função objetivo $f$ tal que a condição AKKT2 é satisfeita. Da definição de AKKT2, existem sequências $\{x^k\}$, $\{\lambda^k\}$, $\{\mu^k\}$, $\{\eta^k\}$, $\{\theta^k\}$, $\{\delta_k\}$, com $\mu^k \geq 0$, $\mu_j^k = 0$ para $j \notin A(x^*)$ e $\theta^k \geq 0$, $\theta_j^k = 0$ para $j \notin A(x^*)$ tais que $x^k \to x^*$, $\delta_k \to 0$ e

a) $\varepsilon^k := \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k) \to 0$;

b) $\nabla_x^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^m \eta_i^k \nabla h_i(x^k)\nabla h_i(x^k)^T + \sum_{j\in A(x^*)} \theta_j^k \nabla g_j(x^k)\nabla g_j(x^k)^T \succeq -\delta_k \mathbb{I}$.

Usando o item (a) e o item (b), deduzimos que

$$(-\nabla f(x^k) + \varepsilon^k, -\nabla^2 f(x^k) - \delta_k \mathbb{I}) \in K_2^{\mathcal{W}}(x^k) \text{ para todo } k \in \mathbb{N}.$$

Usando, agora, a continuidade de $\nabla f(x)$ e de $\nabla^2 f(x)$ junto com a semi-continuidade exterior de $K_2^{\mathcal{W}}(x)$ em $x^*$, obtemos que $(-\nabla f(x^*), -\nabla^2 f(x^*)) \in K_2^{\mathcal{W}}(x^*)$ e como consequência WSONC vale.

Provemos a outra implicação. Seja $(w, W)$ um elemento de $\limsup K_2^{\mathcal{W}}(x)$ quando $x \to x^*$. Mostraremos que $(w, W)$ pertence a $K_2^{\mathcal{W}}(x^*)$. De 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 $x^k \to x^*$,

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

e

$$H^k \preceq \sum_{i=1}^{m} \lambda_i^k \nabla^2 h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla^2 g_j(x^k) \text{ sobre } C^{\mathcal{W}}(x^k, x^*).$$

Defina a função

$$f(x) := -\langle w, x - x^* \rangle - \frac{1}{2} W(x - x^*, x - x^*).$$

Mostraremos que a condição AKKT2 vale em $x^*$ com $f(x)$ como função objetivo. Do cálculo temos que $\nabla f(x) = -w - W(x - x^*)$ e $\nabla^2 f(x) = -W$. Para provar que (3.2.3) vale, é suficiente notar que:

$$\lim_{k \to \infty} \nabla_x L(x^k, \lambda^k, \mu^k) = \lim_{k \to \infty} \left( \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k g_j(x^k) - w \right) - \lim_{k \to \infty} W(x^k - x^*) = 0.$$

Agora, para provar a validade de (3.2.4), usamos o Lema 3.4 com

$$P^k := \sum_{i=1}^{m} \lambda_i^k \nabla^2 h_i(x^k) + \sum_{j \in A(x^*)} \mu_j^k \nabla^2 g_j(x^k) - H^k + \frac{1}{k} \mathbb{I}, \tag{3.3.21}$$

e $a_i$ como as colunas da matriz $[\nabla h_i(x^k), i = 1, \ldots, m; \nabla g_j(x^k), j \in A(x^*)]$. Assim, pelo Lema 3.4 existem sequências positivas $\{\theta^k\}$ e $\{\eta^k\}$ tais que:

$$S^k := P^k + \sum_{i=1}^{m} \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T + \sum_{j \in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T \succ 0. \tag{3.3.22}$$

Coloque $\theta_j^k = 0$ para $j \notin A(x^*)$. Usando (3.3.21), (3.3.22) e $\nabla^2 f(x) = -W$, obtemos que

$$\begin{aligned} &\nabla_x^2 L(x^k, \lambda^k, \mu^k) + \sum_{i=1}^{m} \eta_i^k \nabla h_i(x^k) \nabla h_i(x^k)^T \\ &+ \sum_{j \in A(x^*)} \theta_j^k \nabla g_j(x^k) \nabla g_j(x^k)^T = -W + H^k + S^k - \frac{1}{k} \mathbb{I}. \end{aligned} \tag{3.3.23}$$

Agora, procederemos a achar um limitante inferior para a matriz simétrica do lado direito da expressão (3.3.23). Usando a definição de norma:

$$-W + H^k \succeq -|\lambda_1(W - H^k)| \mathbb{I},$$

onde $\lambda_1(W-H^k)$ denota o menor autovalor de $W-H^k$. De (3.3.22), $S^k \succ 0$, assim temos que

$$-W+H^k+S^k-\frac{1}{k}\mathbb{I} \succeq -|\lambda_1(W-H^k)|\mathbb{I}+S^k-\frac{1}{k}\mathbb{I} \succeq -|\lambda_1(W-H^k)|\mathbb{I}-\frac{1}{k}\mathbb{I} = -\delta_k\mathbb{I}, \tag{3.3.24}$$

onde $\delta_k := |\lambda_1(W-H^k)|+1/k$. Como $H^k \to W$ e $\delta_k \to 0$, substituindo (3.3.24) em (3.3.23), temos que a condição (3.2.4) vale. Assim, $x^*$ é um ponto AKKT2 e pela hipótese, WSONC vale em $x^*$. Usando (3.3.19), $(w,W) = (-\nabla f(x^*), -\nabla^2 f(x^*))$ pertence a $K_2^{\mathcal{W}}(x^*)$, terminando com a demonstração. □

Como AKKT2 é uma condição de otimalidade, pelo Teorema 3.14, temos que:

**Corolário 3.3.** *Se $x^*$ é um minimizador local tal que CCP2 vale, então WSONC vale.*

Diretamente da Proposição 3.1 e do Teorema 3.14 temos que:

**Proposição 3.2.** *A condição CCP2 é implicada por MFCQ+WCR.*

De fato CCP2 é estritamente mais fraca que MFCQ+WCR como o seguinte exemplo mostra.

**Exemplo 3.2.** *(A condição CCP2 não implica MFCQ+WCR.)*

Em $\mathbb{R}$, considere $x^* = 0$ e as restrições de desigualdade definidas pelas funções

$$g_1(x) = x \quad \text{e} \quad g_2(x) = -x. \tag{3.3.25}$$

Em $x^*$, temos que CPP2 vale mas MFCQ falha (como consequência MFCQ+WCR falha). Para ver isso, notemos que desde que os gradientes das funções $g_1$ e $g_2$ são positivamente dependentes, MFCQ não vale em $x^* = 0$. Calculemos o cone $K_2^{\mathcal{W}}(x)$ para todo $x \in \mathbb{R}$. Como $\nabla g_1(x) = 1, \nabla^2 g_1(x) = 0, \nabla g_2(x) = -1$ e $\nabla^2 g_2(x) = 0$ temos que $C^{\mathcal{W}}(x,x^*) = 0$. Desta forma, qualquer matriz $H \in \mathrm{Sym}(1) = \mathbb{R}$ satisfaz a desigualdade matricial:

$$H \preceq \mu_1 \nabla^2 g_1(x) + \mu_2 \nabla^2 g_2(x) = 0 \text{ sobre } C^{\mathcal{W}}(x,x^*) = \{0\} \text{ para todo } \mu_1, \mu_2 \geq 0.$$

Por conseguinte, $K_2^{\mathcal{W}}(x) = \mathbb{R} \times \mathbb{R}$ para todo $x \in \mathbb{R}$ e $K_2^{\mathcal{W}}(x)$ é semicontínua exteriormente em $\mathbb{R}$.

Provaremos a seguir que RCRCQ é estritamente mais forte que a condição CCP2. Para continuar, precisaremos dos seguintes lemas preparatórios. O primeiro lema é uma re-escrita do Teorema do Posto Constante (Teorema 3.5).

**Lema 3.7.** *Considere que os gradientes $\{\nabla h_i(x), \nabla g_j(x) : i \in \mathcal{I}, j \in \mathcal{J}\}$ têm posto constante em uma vizinhança de $x \in \mathbb{R}^n$. Então para cada $d \in \mathbb{R}^n$ tal que*

$$\langle \nabla h_i(x), d\rangle = 0 \;\; \text{para } i \in \mathcal{I} \text{ e } \langle \nabla g_j(x), d\rangle = 0 \;\; \text{para } j \in \mathcal{J} \tag{3.3.26}$$

*existe uma curva $t \to \phi(t)$, $t \in (-T,T)$, $T > 0$ duas vezes diferenciável tal que $\phi(0) = x$, $\phi'(0) = d$ e para todo $i \in \mathcal{I}$ e $j \in \mathcal{J}$ temos que $h_i(\phi(t)) = h_i(x)$ e $g_j(\phi(t)) = g_j(x)$ $\forall t \in (-T,T)$.*

O seguinte lema é uma variação do Lema de Caratheódory útil para a condição RCRCQ.

**Lema 3.8.** *[9, Lema 1] Suponha que $v = \sum_{i\in\mathcal{I}} \alpha_i p_i + \sum_{j\in\mathcal{J}} \beta_j q_j$ com $p_i, q_j \in \mathbb{R}^n$ para todo $i \in \mathcal{I}, j \in \mathcal{J}$, $\{p_i\}_{i\in\mathcal{I}}$ é um conjunto linearmente independente e $\alpha_i, \beta_j$ são diferentes de zero para todo $i \in \mathcal{I}, j \in \mathcal{J}$. Então existe um subconjunto $\mathcal{J}' \subset \mathcal{J}$ e escalares $\hat{\alpha}_i, \hat{\beta}_j$ para todo $i \in \mathcal{I}$, $j \in \mathcal{J}'$ tais que*

- $v = \sum_{i\in\mathcal{I}} \hat{\alpha}_i p_i + \sum_{j\in\mathcal{J}'} \hat{\beta}_j q_j$;
- *Para todo $j \in \mathcal{J}'$ temos $\beta_j \hat{\beta}_j > 0$;*
- *$\{p_i, q_j\}_{i\in\mathcal{I}, j\in\mathcal{J}'}$ é um conjunto linearmente independente.*

Uma caracterização útil de RCRCQ.

**Teorema 3.15.** *[9, Teorema 1] Seja $I \subset \{1, \dots m\}$ um conjunto de índices tal que $\{\nabla h_i(x) : i \in I\}$ é uma base de span$\{\nabla h_i(x) : i \in \{1, \dots, m\}\}$. Um ponto viável $x \in \Omega$ satisfaz RCRCQ se, e somente se, existe uma vizinhança $V$ de $x$ tal que:*

*a) $\{\nabla h_i(y) : i \in \{1, \dots, m\}\}$ tem o mesmo posto para todo $y \in V$;*

*b) Para cada $J \subset A(x)$, se $\{\nabla h_i(x), \nabla g_j(x) : i \in \{1, \dots, m\}, j \in J\}$ é linearmente dependente então $\{\nabla h_i(y), \nabla g_j(y) : i \in \{1, \dots, m\}, j \in J\}$ é linearmente dependente para todo $y \in V$.*

Provemos a próxima proposição.

**Proposição 3.3.** *A condição CCP2 é implicada por RCRCQ.*

*Demonstração.* Seja $(w, W)$ um elemento de $\limsup_{x\to x^*} K_2^{\mathcal{W}}(x)$. 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 $x^k \to x^*$,

$$w^k := \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k) \to w \text{ e } H^k \to W, \tag{3.3.27}$$

onde $H^k \preceq \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla^2 g_j(x^k)$ sobre $C^{\mathcal{W}}(x^k, x^*)$.
Seja $\mathcal{I} \subset \{1, \dots, m\}$ um subconjunto de índices tal que os gradientes $\{\nabla h_i(x^*)\}_{i\in\mathcal{I}}$ formam uma base do subespaço vetorial gerado por $\{\nabla h_i(x^*) : i \in \{1, \dots, m\}\}$. Da continuidade, $\{\nabla h_i(x^k) : i \in \mathcal{I}\}$ é linearmente independente para $k$ suficientemente grande. Pelo Teorema 3.15, item (a), temos que $\{\nabla h_i(x^k) : i \in \mathcal{I}\}$ é uma base do subespaço vetorial gerado por $\{\nabla h_i(x^k) : i \in \{1, \dots, m\}\}$ para $k$ suficientemente grande. Então, existe uma sequência $\bar{\lambda}_i^k \in \mathbb{R}$ para $i \in \mathcal{I}$ tal que $\sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) = \sum_{i\in\mathcal{I}} \bar{\lambda}_i^k \nabla h_i(x^k)$. Assim, podemos escrever:

$$w^k = \sum_{i\in\mathcal{I}} \bar{\lambda}_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \mu_j^k \nabla g_j(x^k). \tag{3.3.28}$$

Usando o Lema 3.8 para a expressão acima, encontramos um subconjunto de índices $\mathcal{J}_k \subset A(x^*)$ e multiplicadores $\hat{\lambda}_i^k \in \mathbb{R}, i \in \mathcal{I}$ e $\hat{\mu}_j^k \in \mathbb{R}_+, j \in \mathcal{J}_k$ para $k$ suficientemente grande tal que

$$w^k = \sum_{i\in\mathcal{I}} \hat{\lambda}_i^k \nabla h_i(x^k) + \sum_{j\in\mathcal{J}_k} \hat{\mu}_j^k \nabla g_j(x^k) \tag{3.3.29}$$

e $\{\nabla h_i(x^k), \nabla g_j(x^k) : i \in \mathcal{I}, j \in \mathcal{J}_k\}$ é um conjunto linearmente independente. Como $A(x^*)$ é um conjunto finito de índices, sem perda de generalidade, podemos supor que $\mathcal{J} := \mathcal{J}_k$ para todo $k$ suficientemente grande. Do Teorema 3.15, item (b), temos que $\{\nabla h_i(x^*), \nabla g_j(x^*) : i \in \mathcal{I}, j \in \mathcal{J}\}$ é um conjunto linearmente independente e $\{\hat{\lambda}_i^k, \hat{\mu}_j^k : i \in \mathcal{I}, j \in \mathcal{J}\}_{k\in\mathbb{N}}$ é uma sequência limitada, assim, sem perda de generalidade podemos supor que $\hat{\lambda}_i^k \to \lambda_i$ e $\hat{\mu}_j^k \to \mu_j$. Tomando limites em (3.3.29) obtemos que $w = \sum_{i\in\mathcal{I}} \lambda_i \nabla h_i(x^*) + \sum_{j\in\mathcal{J}} \mu_j \nabla g_j(x^*)$. Para cada $k \in \mathbb{N}$, defina $\hat{\lambda}_i^k = 0$ para $i \notin \mathcal{I}$ e $\hat{\mu}_j^k = 0$ para $j \notin \mathcal{J}$, defina também $\lambda_i = 0$ para $i \notin \mathcal{I}$ e $\mu_j = 0$ para $j \notin \mathcal{J}$. Agora, mostraremos que para todo $d \in C^{\mathcal{W}}(x^*)$ vale que $d^{\mathtt{T}} H d \leq d^{\mathtt{T}} \left(\sum_{i=1}^m \lambda_i \nabla^2 h_i(x^*) + \sum_{j\in A(x^*)} \mu_j \nabla^2 g_j(x^*)\right) d$.

Defina para cada $k \in \mathbb{N}$, $\Lambda_i^k = \lambda_i^k - \hat{\lambda}_i^k$, $\Upsilon_j^k = \mu_j^k - \hat{\mu}_j^k$. De (3.3.29) e (3.3.27) temos que

$$\sum_{i=1}^m \Lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \Upsilon_j^k \nabla g_j(x^k) = 0 \quad \text{para } k \in \mathbb{N} \text{ suficientemente grande.} \tag{3.3.30}$$

Seja $d \in C^{\mathcal{W}}(x^*)$. Uma vez que RCRCQ implica WCR, temos que existe uma sequência $d^k \to d$ tal que $d^k \in C^{\mathcal{W}}(x^k, x^*)$, ver Lema 3.6, portanto

$$(d^k)^{\mathtt{T}} H^k (d^k) \leq \sum_{i=1}^m \lambda_i^k (d^k)^{\mathtt{T}} \nabla^2 h_i(x^k)(d^k) + \sum_{j\in A(x^*)} \mu_j^k (d^k)^{\mathtt{T}} \nabla^2 g_j(x^k)(d^k) \tag{3.3.31}$$

$$\leq \sum_{i=1}^m \hat{\lambda}_i^k (d^k)^{\mathtt{T}} \nabla^2 h_i(x^k)(d^k) + \sum_{j\in A(x^*)} \hat{\mu}_j^k (d^k)^{\mathtt{T}} \nabla^2 g_j(x^k)(d^k) + \Xi^k, \tag{3.3.32}$$

em que

$$\Xi^k := \sum_{i=1}^m \Lambda_i^k (d^k)^{\mathtt{T}} \nabla^2 h_i(x^k)(d^k) + \sum_{j\in A(x^*)} \Upsilon_j^k (d^k)^{\mathtt{T}} \nabla^2 g_j(x^k)(d^k). \tag{3.3.33}$$

De RCRCQ, temos que para $k$ suficientemente grande, $x^k$ tem uma vizinhança onde o posto $\{\nabla h_i(x^k), \nabla g_j(x^k) : i \in \{1, \ldots, m\}, j \in A(x^*)\}$ é constante, assim pelo Lema 3.7, existe uma curva $t \to \phi_k(t)$ para $t \in (-T_k, T_k)$, $T_k > 0$ com $\phi_k(0) = x^k$, $\phi_k'(0) = d^k$ tal que $h_i(\phi_k(t)) = h_i(x^k)$ para todo $i \in \{1, \ldots, m\}$ e $g_j(\phi_k(t)) = g_j(x^k)$ para todo $j \in A(x^*)$. Denote $v^k := \phi_k''(0)$. Derivando $h_i(\phi_k(t)) = 0$, $i \in \{1, \ldots, m\}$ e $g_j(\phi_k(t)) = 0$, $j \in A(x^*)$ duas vezes em $t = 0$, obtemos:

$$\langle \nabla h_i(x^k), v^k \rangle + \nabla^2 h_i(x^k)(d^k, d^k) = 0, \text{ para todo } i \in \{1, \ldots, m\}; \tag{3.3.34}$$

$$\langle \nabla g_j(x^k), v^k \rangle + \nabla^2 g_j(x^k)(d^k, d^k) = 0, \text{ para todo } j \in A(x^*). \tag{3.3.35}$$

Assim, substituindo as expressões (3.3.34) e (3.3.35) em (3.3.33)

$$\Xi^k = -\sum_{i=1}^m \Lambda_i^k \langle \nabla h_i(x^k), v^k \rangle - \sum_{j\in A(x^*)} \Upsilon_j^k \langle \nabla g_j(x^k), v^k \rangle \tag{3.3.36}$$

$$= -\langle \sum_{i=1}^m \Lambda_i^k \nabla h_i(x^k) + \sum_{j\in A(x^*)} \Upsilon_j^k \nabla g_j(x^k), v^k \rangle = 0, \tag{3.3.37}$$

onde na última igualdade usamos (3.3.30). Como $\Xi^k = 0$ para todo $k$ suficientemente grande, temos que (3.3.32) torna-se

$$(d^k)^{\mathtt{T}} H^k(d^k) \leq \sum_{i=1}^{m} \hat{\lambda}_i^k (d^k)^{\mathtt{T}} \nabla^2 h_i(x^k)(d^k) + \sum_{j \in A(x^*)} \hat{\mu}_j^k (d^k)^{\mathtt{T}} \nabla^2 g_j(x^k)(d^k). \quad (3.3.38)$$

Tomando limite em (3.3.38) o resultado fica demonstrado. □

Como consequência CRCQ também implica CCP2.

**Exemplo 3.3.** *A condição CCP2 é estritamente mais fraca que RCRCQ.*

Em $\mathbb{R}^2$, considere $x^* = (0,0)$ e o seguinte conjunto de restrições definido pelas funções

$$\begin{aligned} h_1(x_1,x_2) &= x_1; \\ g_1(x_1,x_2) &= -x_1^2 + x_2; \\ g_2(x_1,x_2) &= -x_1^2 + x_2^3. \end{aligned}$$

Do cálculo, temos que $\nabla h_1(x_1,x_2) = (1,0)$, $\nabla g_2(x_1,x_2) = (-2x_1, 1)$ e $\nabla g_3(x_1,x_2) = (-2x_1, 3x_2^2)$. Desta forma, RCRCQ falha em $x^* = (0,0)$. Agora, desde que $C^{\mathcal{W}}(x,x^*) = \{0\}$, temos que $K_2^{\mathcal{W}}(x) = \mathbb{R} \times \mathbb{R}_+ \times \mathrm{Sym}(2)$. Claramente, $K_2^{\mathcal{W}}(x)$ é semicontínua exteriormente.

Outra condição de qualificação de segunda ordem conhecida na literatura relacionada com MFCQ, foi introduzida por Baccari e Trad, [22]:

**Definição 3.7.** *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.*

Apesar de a condição de Baccari-Trad garantir a validade de WSONC em um minimizador local, da mesma forma que CCP2, ambas condições não estão relacionadas.

**Exemplo 3.4** (A condição de Baccari-Trad não implica CCP2)**.**

Considere em $\mathbb{R}^2$ o ponto $x^* = (0,0)$ e as restrições de desigualdades definidas pelas funções

$$\begin{aligned} g_1(x_1,x_2) &= -x_2; \\ g_2(x_1,x_2) &= x_1^2 - x_2. \end{aligned}$$

Então, a condição de Baccari-Trad vale em $x^*$ mas CCP2 falha.

*A condição de Baccari-Trad vale em $x^*$.*
Do cálculo, $\nabla g_1(x_1,x_2) = (0,-1)$ e $\nabla g_2(x_1,x_2) = (2x_1,-1)$. Assim, em $x^* = (0,0)$, temos que MFCQ vale e que o posto dos gradientes das restrições ativa é exatamente um a menos que o número das restrições ativa (nosso caso, 2).

*A condição CCP2 não é satisfeita em $x^*$.*
Da forma dos gradientes, temos que $C^{\mathcal{W}}(x^*,x^*) = \mathbb{R} \times \{0\}$ e $C^{\mathcal{W}}(x,x^*) = \{(0,0)\}$, se $x_1 \neq x_1^*$. Assim, para qualquer sequência $\{x^k\}$ com $x^k \to x^*$ e $x_1^k > 0$, o cone $K_2^{\mathcal{W}}(x^k) = \mathbb{R}_+ \times \mathbb{R}_- \times \mathrm{Sym}(2)$ mas $K_2^{\mathcal{W}}(x^*)$ é um subconjunto próprio de $\{0\} \times \mathbb{R}_- \times \mathrm{Sym}(2)$. Por conseguinte, CCP2 falha.

Para ver que a condição CCP2 não implica a condição de Baccari-Trad, é suficiente notar que MFCQ+WCR implica CCP2, Proposição 3.2, enquanto MFCQ+WCR não implica a condição de Baccari-Trad, [14, contra-exemplo 5.2].

A independência entre CCP2 e a condição Baccari-Trad tem implicações práticas. Devido ao Teorema 3.14, a condição de Baccari-Trad não é suficiente para garantir que um ponto limite de uma sequência AKKT2 cumpra com WSONC, como mostra o seguinte exemplo.

**Exemplo 3.5** (AKKT2, sob a condição de Baccari-Trad, não implica WSONC)**.**

Considere o problema de otimização não linear:

$$\text{Minimizar } f(x_1, x_2) = -2x_1^2 \text{ sujeito a } g_1(x_1, x_2) = -x_2 \leq 0, g_2(x_1, x_2) = x_1^2 - x_2 \leq 0. \tag{3.3.39}$$

Devido ao Exemplo 3.4, a condição de Baccari-Trad vale em $x^* = (0, 0)$. Para mostrar que $x^*$ é um ponto AKKT2, escolha $x_1^k := 1/k$, $x_2^k := x_1^k$, $\mu_1^k := 0$, $\mu_2^k := 0$, $\theta_2^k := 2(x_1^k)^{-2}$, $\theta_1^k := 2\theta_2^k$ e $\delta_k := 0$. Com essas escolhas temos que $\nabla f(x^k) + \mu_1^k \nabla g_1(x^k) + \mu_2^k \nabla g_2(x^k) = (-4x_1^k, 0) \to (0, 0)$ e

$$\nabla^2 L(x^k, \mu^k) + \theta_1^k \nabla g_1(x^k) \nabla g_1(x^k)^T + \theta_2^k \nabla g_2(x^k) \nabla g_2(x^k)^T = \begin{pmatrix} 4 & -2\theta_2^k x_2^k \\ -2\theta_2^k x_2^k & 3\theta_2^k \end{pmatrix},$$

onde a última matriz é semi-definida positiva. Observe que WSONC falha em $x^*$ e que $x^*$ não é uma solução ótima para o problema (3.3.39).

Assim, neste exemplo, temos um ponto $x^* = (0, 0)$ que não é uma solução ótima e que não satisfaz WSONC, mas ele pode ser atingido por uma sequência AKKT2 (sequência, talvez, gerada pelo método de lagrangiano aumentado ou pelo método SQP regularizado) e como consequência aceito como possível solução. Isto não acontece, se em lugar da condição de Baccari-Trad, considerarmos qualquer outra condição de qualificação que implique CCP2, como LICQ, MFCQ+WCR, CRCQ, RCRCQ.

Assim, definimos uma nova condição de qualificação CCP2 e provamos que é uma condição de qualificação para o qual WSONC vale em um minimizador local. A condição CCP2 é estritamente mais fraca que a condição MFCQ+WCR e que a condição RCRCQ, ainda mais, esta condição é a mínima condição para assegurar que qualquer ponto AKKT2 (ponto limite de uma sequência AKKT2) satisfaz a condição WSONC.

Isto tem consequências práticas, porque melhora os resultados de convergência para pontos estacionários de segunda ordem de qualquer algoritmo que gere sequências AKKT2, como é o caso do Lagrangiano aumentado de [5] ou o método de regiões de confiança de [40]. Além disso, esse tipo de resultado não pode ser melhorado, em princípio, usando condições de qualificação mais fracas que CCP2.

A seguir, analisaremos a condição WSONC e forneceremos outra razão pela qual a condição WSONC é a condição natural quando analisamos algoritmos práticos com convergência a pontos estacionários de segunda ordem.

Suponha que desejamos uma condição que garanta que todo ponto limite de qualquer sequência AKKT2 satisfaz não somente a condição WSONC mas também a condição forte de segunda ordem, SSONC. Observe que segundo nosso conhecimento, não há algoritmos com convergência a pontos estacionários de segunda ordem onde SSONC vale. Com esse objetivo em mente, definimos a seguinte condição de qualificação no espirito do Teorema 3.14, substituindo WSONC por SSONC. Nosso principal objetivo é tentar entender porque não é natural esperar que algoritmos práticos gerem sequências cujo pontos limites satisfazem SSONC.

**Definição 3.8.** *Dizemos que a propriedade forte do cone contínuo de segunda ordem, SCCP2, vale em $x^* \in \Omega$ se*

$$\limsup_{x \to x^*} K_2^{\mathcal{W}}(x) \subset K_2^{\mathcal{S}}(x^*),$$

*onde $K_2^{\mathcal{S}}(x^*)$ é o cone:*

$$K_2^{\mathcal{S}}(x^*) := \bigcup_{\substack{(\lambda,\mu)\in\mathbb{R}^m\times\mathbb{R}^p_+,\\ \mu_j=0 \text{ para } j\notin A(x^*)}} S^{\mathcal{S}}(\lambda,\mu,x) \tag{3.3.40}$$

*onde*

$$S^{\mathcal{S}}(\lambda,\mu,x) := \left\{ \begin{array}{l} (\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j\in A(x^*)} \mu_j \nabla g_j(x^*), H), \text{ tal que} \\ H \preceq \sum_{i=1}^m \lambda_i \nabla^2 h_i(x^*) + \sum_{j\in A(x^*)} \mu_j \nabla^2 g_j(x^*) \text{ sobre } C^S(x^*,\mu) \end{array} \right\},$$

*e $C^{\mathcal{S}}(x^*,\mu)$ é o cone crítico forte.*

Usando o cone $K_2^{\mathcal{S}}(x^*)$ podemos escrever SSONC de uma forma mais geométrica: SSONC vale em $x^*$ se, e somente se o par $(-\nabla f(x^*), -\nabla^2 f(x^*))$ pertence a $K_2^{\mathcal{S}}(x^*)$.

Notamos que a condição SCCP2 é mais forte que CCP2, já que o cone $K_2^{\mathcal{S}}(x^*)$ é um subconjunto de $K_2^{\mathcal{W}}(x^*)$. Para ver isto, é suficiente observar que para todo $\mu \geq 0$, o cone crítico fraco $C^{\mathcal{W}}(x^*)$ está incluso em $C^{\mathcal{S}}(x^*,\mu)$.

Seguindo o mesmo raciocínio do Teorema 3.14 temos:

**Teorema 3.16.** *Seja $x^* \in \Omega$. Então as condições abaixo são equivalentes:*

*(1) SCCP2 vale em $x^*$;*

*(2) Para qualquer função objetivo $f : \mathbb{R}^n \to \mathbb{R}$ tal que AKKT2 é satisfeita em $x^*$, temos que SSONC vale em $x^*$.*

Os seguintes exemplos mostram que a condição SCCP2 é tão forte que pode falhar mesmo em problemas bem comportados, com estrutura simples e onde a condição LICQ é satisfeita.

**Exemplo 3.6** (SCCP2 falha para restrições de caixa simples)**.** *Considere em $\mathbb{R}^n$ $(n \geq 1)$ a restrição de caixa simples $\Omega = \{x \in \mathbb{R}^n : x \geq 0\}$. Então SCCP2 falha em $x^* = 0$ mas CCP2 vale.*

Claramente, a caixa $\Omega$ está definida pelas restrições de desigualdade dadas pelas funções lineares $g_j(x) = -x_j$, para todo $j \in \{1,\dots,n\}$. Procederemos com o cálculo do cone $K_2^{\mathcal{W}}(x)$. Para $x^* = 0$, o conjunto das restrições ativas é $A(x^*) = \{1,\dots,n\}$. Além disso, para todo $x \in \mathbb{R}^n$ temos que $\nabla g_j(x) = -e_j$ e $\nabla^2 g_j(x) = 0$ independente de $j$. Então temos que

$$C^{\mathcal{W}}(x,x^*) = \{d \in \mathbb{R}^n : \ \langle \nabla g_j(x), d\rangle = 0 \text{ para todo } j \in \{1,\dots,n\}\} = \{0\},$$

e como consequência

$$K_2^{\mathcal{W}}(x) = \{(\sum_{j\in A(x^*)} -\mu_j e_j, H) : H \preceq 0 \text{ sobre } C^{\mathcal{W}}(x,x^*) = \{0\},\ \mu_j \geq 0\},$$

se reduz a $K_2^{\mathcal{W}}(x) = \mathbb{R}^n_- \times \mathrm{Sym(n)}$ independentemente de $x$. Assim $K_2^{\mathcal{W}}$ é semicontínua exteriormente em $x^*$ e CCP2 vale. Ainda mais, $\limsup_{x\to x^*} K_2^{\mathcal{W}}(x) = \mathbb{R}^m_- \times \mathrm{Sym(n)}$.

Agora mostraremos que SCPP2 falha em $x^* = 0$. Para provar que SCCP2 não vale em $x^*$, será suficiente encontrar um vetor $\hat{\mu} \in \mathbb{R}^n_+$ e uma matriz simétrica $H$ tal que $w^{\mathsf{T}} H w > 0$ para algum $w \in C^{\mathcal{S}}(x^*, \hat{\mu})$. Pois, nesse caso o par $(-\hat{\mu}, H) \in K_2^{\mathcal{W}}(x) = \mathbb{R}^m_- \times \mathrm{Sym(n)}$ e $(-\hat{\mu}, H)$ não pertence a $K_2^{\mathcal{S}}(x^*)$. Defina $e := \sum_{j=1}^n e_j$, $\hat{\mu} := e - e_1$ e $H := e_1 e_1^T$. Da definição de cone crítico forte $C^{\mathcal{S}}(x^*, \hat{\mu})$, temos que $e_1 \in C^{\mathcal{S}}(x^*, \hat{\mu})$ e da definição da matriz $H$, $e_1^{\mathsf{T}} H e_1 = \|\langle e_1, e_1 \rangle\|^2 > 0$. Desta forma temos que o par $(-\hat{\mu}, H) \in \limsup_{x \to x^*} K_2^{\mathcal{W}}(x)$ e não pertence a $K_2^{\mathcal{S}}(x^*)$.

Notemos que no exemplo anterior, a condição de independência linear vale em $x^* = 0$. Já o próximo exemplo mostra que a condição SCPP2 pode valer em problemas onde a independência linear falha.

**Exemplo 3.7** (SCPP2 não implica LICQ)**.**

Considere em $\mathbb{R}$, o ponto $x^* = 0$ e o conjunto de restrições de desigualdade definidas por

$$\begin{aligned} g_1(x) &= -\exp(x) + 1; \\ g_2(x) &= x. \end{aligned}$$

Calculando os gradientes e as hessianas das funções em cada ponto, temos que: $\nabla g_1(x) = -\exp(x)$, $\nabla^2 g_1(x) = -\exp(x)$, $\nabla g_2(x) = 1$ e $\nabla^2 g_2(x) = 0$, $\forall x \in \mathbb{R}$.

*LICQ não vale em $x^*$.*
Claramente, ambas restrições são ativas em $x^*$, além disso, temos que a condição de independência linear não vale em $x^*$, de fato não vale em nenhum ponto de $\mathbb{R}$.

*A condição SCPP2 vale em $x^*$.*
Do cálculo, temos que os dois cones $C^{\mathcal{W}}(x, x^*)$ e $C^{\mathcal{S}}(x^*, \mu)$ com $\mu \in \mathbb{R}^2_+$ são subespaços triviais, isto é, $C^{\mathcal{W}}(x, x^*) = \{0\}$ e $C^{\mathcal{S}}(x^*, \mu) = \{0\}$ para todo $\mu \in \mathbb{R}^2_+$. Assim, $K^{\mathcal{W}}(x) = \mathbb{R} \times \mathrm{Sym}(1) = K^{\mathcal{S}}(x^*)$ para todo $x \in \mathbb{R}$, o que implica que SCCP2 vale. A Figura 3.2 mostra as relações existentes entre as condições de qualificação discutidas nesta seção.

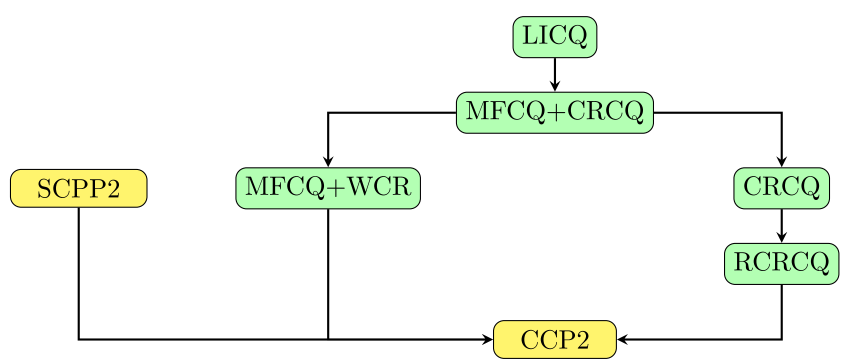



Figura 3.2: Relações das CQs associadas com a convergência a pontos estacionários de segunda ordem.

## 3.4 Exercícios

1. Mostre que o cone crítico forte $C^S(x)$ pode ser escrito como:

$$\left\{ d \in \mathbb{R}^n \mid \begin{array}{ll} \nabla h_i(x)^{\mathtt{T}} d = 0, & i \in \{1, \ldots, m\} \\ \nabla g_j(x)^{\mathtt{T}} d = 0, & j \in A(x) \text{ se } \lambda_i > 0\} \\ \nabla g_j(x)^{\mathtt{T}} d \leq 0, & j \in A(x) \text{ se } \lambda_i = 0\} \end{array} \right\} \tag{3.4.41}$$

2. Demonstre o Lema 3.4.

3. Toda condição sequencial de otimalidade tem associada um critério de parada. Qual é o critério de parada associado à condição AKKT2? Ou seja, como você verificaria AKKT2 na prática?

4. Sabemos que AKKT2 implica AKKT e que CCP2 e CCP estão relacionados com AKKT2 e AKKT, respectivamente. Então, que relações de implicação existem entre CCP2 e CCP?

5. Complete a demonstração do Teorema 3.1.

6. Considere o problema

$$\begin{array}{ll} \text{Minimizar} & x_3, \\ & x_3 \geq 2\sqrt{3}x_1x_2 - 2x_2^2, \\ & x_3 \geq x_2^2 - 3x_1^2, \\ & x_3 \geq -2\sqrt{3}x_1x_2 - 2x_2^2, \end{array}$$

   cuja solução é a origem (veja a Figura 3.1). Mostre que vale MFCQ, mas não se cumpre nem mesmo a condição de otimalidade fraca de segunda ordem (Definição 3.3). Verifique com uma conta que a tese do Teorema 3.2 se cumpre, e que as hipóteses do Teorema 3.6, não.

# Bibliografia

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.; BEHLING, R.; HAESER, G.; SILVA, P. On second order optimality conditions for nonlinear optimization. *to appear in Optimization Methods & Software*, 2016.

3 ANDREANI, R.; BIRGIN, E. G.; MARTÍNEZ, J. M.; SCHUVERDT, M. L. Augmented lagrangian methods under the constant positive linear dependence constraint qualification. *Mathematical Programming*, v. 111, p. 5–32, 2007.

4 ANDREANI, R.; BIRGIN, E. G.; MARTÍNEZ, J. M.; SCHUVERDT, M. L. On augmented lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, v. 18, p. 1286–1309, 2008.

5 ANDREANI, R.; BIRGIN, E. G.; MARTÍNEZ, J. M.; SCHUVERDT, M. L. Second-order negative-curvature methods for box-constrained and general constrained optimization. *Computational Optimization and Applications*, v. 45, p. 209–236, 2010.

6 ANDREANI, R.; ECHAGüE, C. E.; SCHUVERDT, M. L. Constant-Rank Condition and Second-Order Constraint Qualification. *Journal of Optimization Theory and Applications*, v. 146, n. 2, p. 255–266, fev. 2010. ISSN 0022-3239.

7 ANDREANI, R.; HAESER, G.; MARTÍNEZ, J. On sequential optimality conditions for smooth constrained optimization. *Optimization*, v. 60, n. 5, p. 627–641, 2011.

8 ANDREANI, R.; HAESER, G.; RAMOS, A.; SILVA, P. A second-order sequential optimality condition associated to the convergence of optimization algorithms. 2015. Optimization Online.

9 ANDREANI, R.; HAESER, G.; SCHUVERDT, M.; SILVA, P. A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programming*, v. 135, p. 255–273, 2012.

10 ANDREANI, R.; HAESER, G.; SCHUVERDT, M.; SILVA, P. Two new weak constraint qualifications and applications. *SIAM Journal on Optimization*, v. 22, n. 3, p. 1109–1135, 2012.

11 ANDREANI, R.; MARTÍNEZ, J. M.; RAMOS, A.; SILVA, P. A cone-continuity constraint qualification and algorithmic consequences. *to appear in SIAM Journal on Optimization*.

12 ANDREANI, R.; MARTÍNEZ, J. M.; RAMOS, A.; SILVA, P. Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Optimization Online*.

13 ANDREANI, R.; MARTÍNEZ, J. M.; SANTOS, L. T.; SVAITER, B. F. On the behaviour of constrained optimization methods when lagrange multipliers do not exist. *Optimization Methods and Software*, v. 29, p. 646–657, 2014.

14 ANDREANI, R.; MARTÍNEZ, J. M.; SCHUVERDT, M. L. On second-order optimality conditions for nonlinear programming. *Optimization*, v. 56, n. 5-6, p. 529–542, 2007.

15 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, p. 473–485, 2005.

16 ANITESCU, M. Degenerate nonlinear programming with a quadratic growth condition. *SIAM Journal on Optimization*, v. 10, n. 4, p. 1116–1135, 2000.

17 ARUTYUNOV, A. Perturbations of extremal problems with constraints and necessary optimality conditions. *Journal of Soviet Mathematics*, Kluwer Academic Publishers-Plenum Publishers, v. 54, n. 6, p. 1342–1400, 1991. ISSN 0090-4104.

18 ARUTYUNOV, A. V. Smooth abnormal problems in extremum theory and analysis. *Russian Math. Surveys*, v. 67, n. 3, p. 403–457, 2012.

19 AVAKOV, E.; ARUTYUNOV, A.; IZMAILOV, A. Necessary conditions for an extremum in a mathematical programming problem. *Proceedings of the Steklov Institute of Mathematics*, Nauka/Interperiodica, v. 256, n. 1, p. 2–25, 2007. ISSN 0081-5438.

20 AZÉ, D. A characterization of the Lagrange-Karush-Kuhn-Tucker property. *Optimization Online*, December, 2014.

21 BACCARI, A. On the Classical Necessary Second-Order Optimality Conditions. *Journal of Optimization Theory and Applications*, v. 123, n. 1, p. 213–221, out. 2004. ISSN 0022-3239.

22 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, jan. 2005. ISSN 1052-6234.

23 BAZARAA, M.; SHERALI, H.; SHETTY, C. *Nonlinear Programming: Theory and Algorithms*. Third. [S.l.]: Wiley, 2006.

24 BAZARAA, M. S.; SHERALI, H. D.; SHETTY, C. M. *Practical methods of Optimization: theory and algorithms*. [S.l.]: John Wiley and Sons, 2006.

25 BEN-TAL, A. Second-order and related extremality conditions in nonlinear programming. *Journal of Optimization Theory and Applications*, Kluwer Academic Publishers-Plenum Publishers, v. 31, n. 2, p. 143–165, 1980. ISSN 0022-3239.

26 BEN-TAL, A.; NEMIROVSKI, A. *Lectures on Modern Convex Optimization: Analysis, algorithms and engineering applications*. [S.l.]: MPS/SIAM series on Optimization, 2001.

27 BEN-TAL, A.; ZOWE, J. A unified theory of first and second order conditions for extremum problems in topological vector spaces. In: GUIGNARD, M. (Ed.). *Optimality and Stability in Mathematical Programming*. [S.l.]: Springer Berlin Heidelberg, 1982, (Mathematical Programming Studies, v. 19). p. 39–76. ISBN 978-3-642-00849-8.

28 BERTSEKAS, D. P. *Constrained Optimization and Lagrange Multiplier Methods*. [S.l.]: Academic Press, 1982.

29 BERTSEKAS, D. P. *Nonlinear programming*. [S.l.]: Athenas Scientific, 1999.

30 BIEGLER, L. *NonLinear Programming: Concepts, Algorithms, and Applications to Chemical Processes*. [S.l.]: MOS-SIAM Series on Optimization, 2010.

31 BIRGIN, E.; MARTÍNEZ, J. M. *Practical Augmented Lagrangian Methods for Constrained Optimization*. [S.l.]: SIAM Publications, 2014.

32 BIRGIN, E. G.; BUENO, L. F.; MARTÍNEZ, J. M. Assessing the reliability of general purpose inexact restoration methods. *Journal of Computational and Applied Mathematics*, v. 282, p. 1–16, 2015.

33 BONNANS, J.; SHAPIRO, A. *Perturbation Analysis of Optimization Problems*. [S.l.]: Springer, 2000.

34 BUENO, L.; HAESER, G.; MARTÍNEZ, J. An inexact restoration approach to optimization problems with multiobjective constraints under weighted-sum scalarization. *to appear in Optimization Letters*, 2016.

35 BUENO, L. F.; HAESER, G.; MARTÍNEZ, J. M. A flexible inexact-restoration method for constrained optimization. *Journal of Optimization Theory and Applications*, p. 1–21, 2014.

36 BYRD, R. H.; SCHNABEL, R. B.; SHULTZ, G. A. A trust region algorithm for nonlinearly constrained optimization. *SIAM Journal of Numerical Analysis*, v. 24, p. 1152–1170, 1987.

37 CHEN, L.; GOLDFARB, D. Interior point $\ell_2$-penalty methods for nonlinear programming with strong global convergence properties. *Mathematical Programming*, v. 108, p. 1–26, 2006.

38 COLEMAN, T. F.; LIU, J.; YUAN, W. A new trust-region algorithm for equality constrained optimization. *Computacional Optimization and Applications*, v. 21, p. 177–199, 2002.

39 CONN, A. R.; GOULD, N. I. M.; ORBAN, D.; TOINT, P. L. A primal-dual trust-region algorithm for non-convex functions subject to general inequality and linear equality constraints. *Nonlinear optimization and related topics*, p. 15–49, 1998.

40 DENNIS, J. E.; VICENTE, L. N. On the convergence theory of trust-region-based algorithms for equality-constrained optimization. *SIAM Journal of Optimization*, v. 7(4), p. 927–950, 1997.

41 DIPILLO, G.; LUCIDI, S.; PALAGI, L. Convergence to second-order stationary points of a primal-dual algorithm model for nonlinear programming. *Mathematics of Operations Research*, v. 30, p. 897–915, 2005.

42 FACCHINEI, F.; LUCIDI, S. Convergence to second-order stationary points in inequality constrained optimization. *Mathematics of Operations Research*, v. 23, p. 746–766, 1998.

43 FISCHER, A.; FRIEDLANDER, A. A new line search inexact restoration approach for nonlinear programming. *Computational Optimization and Applications*, v. 46, p. 333–346, 2010.

44 FLETCHER, R. *Practical methods of Optimization: Constrained Optimization.* [S.l.]: Wiley, 1981.

45 GAUVIN, J. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Mathematical Programming*, v. 12, p. 136–138, 1977.

46 GILL, P. E.; KUNGURTSEV, V.; ROBINSON, D. P. A regularized SQP method with convergence to second-order optimal points. *Optimization Online*, 2013.

47 GILL, P. E.; ROBINSON, D. P. A globally convergent stabilized SQP method. *SIAM Journal of Optimization*, v. 4, p. 1983–2013, 2013.

48 GOULD, F.; TOLLE, J. A necessary and sufficient qualification for constrained optimization. *SIAM Journal on Applied Mathematics*, v. 20, p. 164–172, 1969.

49 GOULD, N. I. M.; TOINT, P. L. A note on the convergence of barrier algorithms to second-order necessary points. *Mathematical Programming*, v. 85, n. 2, p. 433–438, jun. 1999. ISSN 0025-5610.

50 GUIGNARD, M. Generalized Kuhn-Tucker conditions for mathematical programming problems in Banach space. *SIAM Journal on Control*, v. 7, p. 232–241, 1969.

51 HAESER, G. *Condições Sequenciais de Otimalidade*. Tese (Doutorado) — Instituto de Matemática Estatística e Computação Científica, Universidade de Campinas, Brasil, Setembro 2009.

52 HESTENES, M. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, v. 4, p. 303–320, 1969.

53 JANIN, R. Directional derivative of the marginal function in nonlinear programming. *Mathematical Programming Study*, v. 21, p. 110–126, 1984.

54 JOHN, F. Extremum problems with inequalities as subsidiary conditions. In: AL., K. F. et (Ed.). *Studies and Essays, Courant Anniversary Volume.* [S.l.]: Wiley/Interscience, 1948. p. 187–204.

55 KARUSH, W. *Minima of Functions of Several Variables with Inequalities as Side Constraints.* Dissertação (Mestrado) — Dept. of Mathematics, Univ. of Chicago, 1939.

56 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.

57 KYPARISIS, J. On uniqueness of Kuhn-Tucker multipliers in nonlinear programming. *Mathematical Programming*, v. 32, p. 242–246, 1985.

58 LU, S. Implications of the constant rank constraint qualification. *Mathematical Programming*, Springer-Verlag, v. 126, n. 2, p. 365–392, 2011. ISSN 0025-5610.

59 MACIEL, M. C.; SANTOS, S. A.; SOTTOSANTO, G. N. On second-order optimality conditions for vector optimization. *Journal of Optimization Theory and Applications*, Springer US, v. 149, n. 2, p. 332–351, 2011. ISSN 0022-3239.

60 MACIEL, M. C.; SANTOS, S. A.; SOTTOSANTO, G. N. On second-order optimality conditions for vector optimization: Addendum. *Journal of Optimization Theory and Applications*, Springer US, p. 1–6, 2012. ISSN 0022-3239.

61 MANGASARIAN, O.; 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.

62 MARTÍNEZ, J. M. Inexact restoration method with lagrangian tangent decrease and new merit function for nonlinear programing. *Journal of Optimization Theory and Applications*, v. 11, p. 39–58, 2001.

63 MARTÍNEZ, J. M.; PILOTTA, E. A. Inexact restoration algorithm for constrained optimization. *Journal of Optimization Theory and Applications*, v. 104, p. 135–163, 2000.

64 MARTÍNEZ, J. M.; SVAITER, B. F. A practical optimality condition without constraint qualifications for nonlinear programming. *Journal of Optimization Theory and Applications*, v. 118 No 1, p. 117–133, 2003.

65 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.

66 MORDUKHOVICH, B. S. *Variational Analysis and Generalized Differentiation, I: Basis Theory, II: Applications*. [S.l.]: Springer, 2006.

67 MORGUERZA, J. M.; PRIETO, F. J. An augmented lagrangian interior-point method using directions of negative curvature. *Mathematical Programming*, v. 95 (3), p. 573–616, 2003.

68 MURTY, K. G.; KABADI, S. N. Some np-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, Springer-Verlag, v. 39, n. 2, p. 117–129, 1987. ISSN 0025-5610.

69 NOCEDAL, J.; WRIGHT, S. J. *Numerical optimization*. [S.l.]: Springer, 2006.

70 POWELL, M. J. D. A method for nonlinear constraints in minimization problems. *Optimization ed. by R. Fletcher*, Academic Press, New York, NY, v. 19, p. 283–298, 1969.

71 QI, L.; WEI, Z. On the constant positive linear dependence conditions and its application to SQP methods. *SIAM Journal on Optimization*, v. 10, p. 963–981, 2000.

72 RAMOS, A. *Tópicos em condições de otimalidade para otimização não linear*. Tese (Doutorado) — IME-USP, Departamento de Matemática Aplicada, 2016.

73 ROCKAFELLAR, R. Lagrange multipliers and optimality. *SIAM Review*, v. 35, n. 2, p. 183–238, 1993.

74 ROCKAFELLAR, R. T.; WETS, R. *Variational Analysis*. [S.l.]: Series: Grundlehren der mathematischen Wissenschaften, Vol. 317, 2009.

75 RUSZCZYNSKI, A. *Nonlinear Optimization*. [S.l.]: Princento University Press, 2006.

76 SPIVAK, M. *A Comprehensive Introduction to Differential Geometry*. third. [S.l.]: Publish or Perish, 1999.

77 WACHSMUTH, G. On LICQ and the uniqueness of Lagrange multipliers. *Operations Research Letters*, v. 41, n. 1, p. 78 – 80, 2013.

# Índice

Abadie
condição de, 14, 56, 59, 60
análise variacional, 34
Aproximadamente-KKT, 23, 38, 39
segunda-ordem, 62

Carathéodory
lema de, 45, 70
condição de qualificação, 13
relações, 21, 43, 76
condição de segunda ordem
forte, 50, 58
fraca, 55, 64
condição sequencial de otimalidade, 22
cone contínuo
propriedade de segunda-ordem, 68
propriedade do, 33, 38, 39
cone crítico, 50, 58
fraco, 55
cone linearizado, 11
cone normal, 35, 41
cone polar, 5
cone tangente, 4, 7, 46
constant positive generators, 33, 36, 39
constant rank of the subspace component, 19, 33
critério de parada, 23, 28

dependência linear positiva constante, 30
condição de, 17
condição relaxada de, 18

error bound
propriedade de, 20, 33

Farkas
lema de, 12
Fritz John
condição de, 5, 15, 25, 45
condição de segunda ordem de, 51

gradiente projetado aproximado, 27, 42, 43
condição de qualificação, 41
Guignard
condição de, 13, 21, 33, 38

independência linear
condição de, 14

Karush-Kuhn-Tucker
condição de, 13

Lagrangiano aumentado, 26
segunda-ordem, 66

Mangasarian-Fromovitz
condição de, 15, 25, 52, 60, 64
condição estrita de, 16, 52
multifunção, 34, 46
domínio, 34
fechada, 34
gráfico, 34
imagem, 34
limite exterior, 34
semi-continuidade, 34
multiplicadores de Lagrange, 50
unicidade, 16, 45

penalidade externa, 24
posto constante
condição de, 17
condição relaxada de, 18, 53, 59, 70
teorema do, 53, 70

weak constant rank, 55, 56, 64