# Condições de otimalidade de primeira e segunda ordem em otimização não linear

Gabriel Haeser
Departamento de Matemática Aplicada,
Instituto de Matemática e Estatística, Universidade de São Paulo

16 de Março de 2015

# Resumo

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 tem convergência global para pontos estacionários sob hipóteses mais fracas.

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

# Abstract

In this work we are interested in identifying first and second order properties satisfied by a local minimizer of a nonlinear optimization problem. Our main goal is to find conditions that can be verified by practical algorithms. We define first and second order optimality conditions stronger than the usual ones, requiring less restrictive assumptions on the problem, and as a consequence, we show global convergence to stationarity of several classes of first and second order algorithms under less restrictive assumptions.

**Key words:** nonlinear optimization; optimality conditions; constraint qualifications; practical algorithms.

# Conteúdo

# Capítulo 1

# Introdução

Considere o problema geral de otimização não linear

$$
\begin{array}{ll}
\text{Minimizar} & f_0(x), \\
\text{Sujeito a} & f_i(x) = 0, i = 1, \dots, m, \\
 & f_i(x) \leq 0, i = m+1, \dots, m+p,
\end{array}
$$

onde $f_i : \mathbb{R}^n \to \mathbb{R}, i = 0, \dots, m+p$ são funções continuamente diferenciáveis.

Denotamos por $\Omega = \{x \in \mathbb{R}^n \mid f_i(x) = 0, i = 1, \dots, m, f_i(x) \leq 0, i = m+1, \dots, m+p\}$ o chamado conjunto viável. Para $x \in \Omega$, denotamos por $A(x) = \{i \in \{m+1, \dots, m+p\} \mid f_i(x) = 0\}$ o conjunto de índices de restrições de desigualdade ativas em $x$. Por simplicidade assumiremos que $A(x) = \{m+1, \dots, m+r\}$ em um ponto de interesse $x$. A rigor, $r := r(x)$, entretanto, o ponto $x$ associado a $r$ estará claro do contexto. Um ponto viável $x^* \in \Omega$ é uma solução (ou minimizador) local do problema se existe uma vizinhança $\mathcal{B}$ de $x^*$ tal que $f_0(x^*) \leq f_0(x)$ para todo $x \in \Omega \cap \mathcal{B}$. Se $x^*$ satisfaz a propriedade acima com $\mathcal{B} = \mathbb{R}^n$, dizemos que $x^*$ é uma solução global.

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. Tais condições são denominadas *condições de otimalidade*.

Formalmente, uma condição de otimalidade é qualquer condição satisfeita por uma solução. 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. Neste trabalho 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.

Assumindo que $x^*$ é uma solução, uma primeira condição de otimalidade útil, chamada *condição geométrica*, é obtida analisando o ângulo de $\nabla f_0(x^*)$ com direções que apontam para o interior do conjunto viável. Ou seja, considerando uma sequência de pontos viáveis $x^k \in \Omega$ convergindo para $x^*$ e sendo $d \in \mathbb{R}^n$ a direção unitária pela qual $x^k$ converge para $x^*$, isto é, $d = \lim \frac{x^k - x^*}{\|x^k - x^*\|}$, assumindo que o limite exista, temos

$$f_0(x^k) = f_0(x^*) + \nabla f_0(x^*)^{\mathtt{T}}(x^k - x^*) + o(x^k - x^*),$$

onde $\frac{o(x^k - x^*)}{\|x^k - x^*\|} \to 0, k \to +\infty$. Dividindo por $\|x^k - x^*\|$ e tomando o limite em $k$, observando que $f_0(x^*) \leq f_0(x^k)$ temos que $\nabla f_0(x^*)^{\mathtt{T}} d \geq 0$. Formalizamos este resultado a seguir.

**Definição 1.1 (Cone tangente)** *Dado $x \in \Omega$, denotamos por $\mathcal{T}(x)$ o cone tangente a $\Omega$ em $x$, dado por*

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

**Definição 1.2 (Cone polar)** *Seja $\mathcal{C} \subset \mathbb{R}^n$ um cone. Denotamos por $\mathcal{C}^\circ$ o cone polar de $\mathcal{C}$, dado por*

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

**Teorema 1.3 (Condição de otimalidade geométrica)** *Se $x^* \in \Omega$ é uma solução local, então $-\nabla f_0(x^*) \in \mathcal{T}(x^*)^\circ$.*

Observe que a verificação da condição geométrica na prática é muito difícil, pois 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 $f_i$ que o definem. (Note que um mesmo conjunto $\Omega$ pode possuir descrições analíticas distintas).

Uma primeira condição analítica aparece nos trabalhos Kuhn-Tucker [35] e Fritz John [33], onde temos o seguinte resultado:

**Teorema 1.4 (Condição de Fritz John)** *Se $x^*$ é uma solução local, então existe $0 \neq (\lambda_0, \lambda) \in \mathbb{R} \times \mathbb{R}^{m+r}$ tal que*

$$\sum_{i=0}^{m+r} \lambda_i \nabla f_i(x^*) = 0, \lambda_0 \geq 0, \lambda_i \geq 0, i \in A(x^*).$$

**Prova:** Veja a observação após a prova do Teorema 2.6. □

Note que 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 $f_i(x^*) < 0$, então a restrição $f_i(x) \leq 0$ 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 $\mathcal{T}(x)$ permanece inalterado independente se incluímos ou não as restrições inativas no problema.

Fritz John desenvolveu esta condição 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 [33], 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 $\leq n$ 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 [34], 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 $L(x^*,\lambda) \leq L(x^*,\lambda^*) \leq L(x,\lambda^*), \forall x, \lambda$, onde $x^*$ é uma solução e $\lambda^*$ multiplicador de Lagrange associado, onde $L(x,\lambda) = f_0(x) + \sum_{i=1}^{m+p} \lambda_i f_i(x), \lambda_i \geq 0, i = m+1, \ldots, m+p$ é a função lagrangiana. 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. No Capítulo 4 apresentamos algumas conclusões e trabalhos futuros.

Este trabalho é baseado em alguns artigos do autor na área e seus colaboradores. A definição de condição sequencial é apresentada em [7] e novas condições de qualificação associadas a algoritmos são apresentadas em [9, 10]. Novas condições de segunda ordem teóricas são apresentadas em [2] e condições práticas para algoritmos de segunda ordem são apresentadas em [8].

# Capítulo 2

# Condições de otimalidade de primeira ordem

## 2.1 Condições de qualificação

Sendo o cone tangente $\mathcal{T}(x)$ um objeto geométrico difícil de ser simulado numericamente, podemos definir um objeto analítico que captura o mesmo espírito do cone tangente. Em torno de $x \in \Omega$, as restrições não lineares $f_i(x+d) = 0, i = 1, \ldots, m$ serão substituídas pelo hiperplano tangente $x + d, \nabla f_i(x)^{\mathtt{T}} d = 0$ e de forma análoga, as restrições de desigualdade $f_i(x+d) \leq 0, i \in A(x)$ serão substituídas por $\nabla f_i(x)^{\mathtt{T}} d \leq 0$. Desta maneira, podemos definir um cone que coincide com o cone tangente em muitos casos, chamado cone linearizado conforme a definição abaixo:

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

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

Com um argumento análogo à prova da condição geométrica, é facil ver que $\mathcal{T}(x) \subset \mathcal{L}(x)$, embora a igualdade possa não acontecer. De fato, basta considerar $\Omega \subset \mathbb{R}^2$ definido por $x_1 x_2 = 0, x_1 \geq 0, x_2 \geq 0$ em torno da origem, onde o cone linearizado é o ortante positivo, enquanto o cone tangente é formado pela união dos semi-eixos positivos. Embora os cones não coincidam, é fácil verificar que o polar de ambos os cones é o ortante negativo. O conjunto em $\mathbb{R}^2$ definido por $x_2 \leq -x_1^3, x_2 \geq x_1^3$ em torno da origem mostra que, em geral, $\mathcal{T}(x)^\circ \neq \mathcal{L}(x)^\circ$, embora, é claro, $\mathcal{L}(x)^\circ \subset \mathcal{T}(x)^\circ$.

Assumindo que $\mathcal{T}(x)^\circ = \mathcal{L}(x)^\circ$, a condição de otimalidade geométrica se reduz a $-\nabla f_0(x) \in \mathcal{L}(x)^\circ$, e ocorre que o polar do cone linearizado pode ser facilmente computado utilizando o Lema de Farkas, obtendo:

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

Com efeito, seja $S(x)$ o cone do lado direito da igualdade acima. É fácil observer que $S(x)^\circ = \mathcal{L}(x)$ e sendo $S(x)$ um cone convexo fechado, $(S(x)^\circ)^\circ = S(x)$.

Note que a condição $-\nabla f_0(x) \in \mathcal{L}(x)^\circ$ é precisamente a condição KKT.

**Teorema 2.2 (Karush-Kuhn-Tucker)** *Seja $x$ uma solução e assuma que $\mathcal{T}(x)^\circ = \mathcal{L}(x)^\circ$, então existem multiplicadores de Lagrange $\lambda_i, i = 1, \ldots, m+r$ tais que*

$$\nabla f_0(x) + \sum_{i=1}^{m+r} \lambda_i \nabla f_i(x) = 0, \lambda_i \geq 0, i \in A(x). \tag{2.1}$$

Os pontos viáveis $x \in \Omega$ que cumprem (2.1) são chamado pontos estacionários (ou pontos KKT) do problema, e são os principais candidatos à solução quando o problema cumpre uma condição de qualificação.

Sendo assim, a igualdade dos cones polares é uma condição de qualificação, chamada condição de Guignard [31]. Mais do que isso, a condição de Guignard é a condição de qualificação mais fraca possível [29], no sentido de que, fixado $x \in \Omega$, se para qualquer função objetivo $f_0$ tal que $x$ é solução do problema de minimizar $f_0(x)$ 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

$$\mathcal{T}(x)^\circ \subset \{-\nabla f_0(x) \mid f_0 \text{ assume um minimizador local restrito em } x\}$$

e será feita no Teorema 2.3 no final desta sessão, onde corrigimos uma pequena imprecisão de [29]. Note que a inclusão recíproca

$$\mathcal{T}(x)^\circ \supset \{-\nabla f_0(x) \mid f_0 \text{ assume um minimizador local restrito em } x\}$$

tambem é satisfeita, e é precisamente uma re-escrita da condição de otimalidade geométrica do Teorema 1.3.

Formalmente, a condição KKT puramente não é uma condição de otimalidade, já que para o problema de minimizar $x$, sujeito a $x^2 = 0$, a solução $x = 0$ não é um ponto estacionário. As condições de otimalidade clássicas associadas à condição KKT são da forma

$$\text{KKT ou não-CQ},$$

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. Embora conheçamos a condição de qualificação mais fraca possível, não se conhece um algoritmo capaz de encontrar pontos estacionários, a menos que se exija mais regularidade do problema. 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*.

A condição de igualdade entre os cones $\mathcal{T}(x) = \mathcal{L}(x)$, 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. Vamos listar condições de qualificação associadas à algoritmos, iniciando das mais restritivas, embora a conexão com algoritmos seja postergada para a próxima seção.

Uma das condições de qualificação mais frequentes na literatura de otimização é a *linear independence constraint qualification (LICQ)* - ou *regularidade* - que afirma que os gradientes das restrições de igualdade e de desigualdade ativas são linearmente independentes, isto é,

$$\{\nabla f_i(x)\}_{i=1}^{m+r} \quad \text{é linearmente independente.}$$

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). Sob LICQ, é fácil ver que os multiplicadores de Lagrange são únicos.

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, [40]). Vale MFCQ em $x \in \Omega$ quando

$$\sum_{i=1}^{m+r} \alpha_i \nabla f_i(x) = 0, \quad \alpha_i \geq 0, \forall i \in A(x) \Rightarrow \alpha_i = 0, \forall i = 1, \ldots, m+r.$$

Neste caso diremos que os gradientes $\nabla f_i(x), i = 1, \ldots, m$ e $\nabla f_i(x), i = m+1, \ldots, m+r$ são positivo-linearmente independentes. Utilizando um teorema de alternativa, podemos ver que a condição MFCQ é equivalente ao fato que existe uma direção estritamente viável, no seguinte sentido: $\nabla f_i(x), i = 1, \ldots, m$ é linearmente independente e existe uma direção $d \in \mathbb{R}^n$ tal que $\nabla f_i(x)^{\mathtt{T}} d =$

$0, i = 1, \ldots, m$ e $\nabla f_i(x)^\intercal d < 0, i \in A(x)$. Ver [47]. 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 compacto não vazio [27]. 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.

A condição equivalente ao fato dos multiplicadores de Lagrange serem únicos é denominada *strict MFCQ, SMFCQ [36]* e é definida a seguir. Suponha que $\lambda \in \mathbb{R}^{m+r}$ seja um multiplicador de Lagrange associado a $x \in \Omega$. Vale SMFCQ quando os gradientes $\nabla f_i(x), i \in \{1, \ldots, m\} \cup \{i \in A(x) | \lambda_i > 0\}$ e $\nabla f_i(x), i \in \{i \in A(x) | \lambda_i = 0\}$ são positivo-linearmente independentes, isto é, as restrições de desigualdade ativas associadas a um multiplicador positivo são tratadas como restrições de igualdade na definição de MFCQ.

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

$$\sum_{i \in \{1,\ldots,m\} \cup \{i \in A(x) | \lambda_i > 0\}} (\bar{\lambda}_i - \lambda_i) \nabla f_i(x) + \sum_{i \in A(x) | \lambda_i = 0} \bar{\lambda}_i \nabla f_i(x) = 0,$$

o que contradiz SMFCQ. Reciprocamente, se não vale SMFCQ, existe $0 \neq \alpha \in \mathbb{R}^{m+r}$ com $\alpha_i \geq 0, i \in \{i \in A(x) \mid \lambda_i = 0\}$ tal que $\sum_{i=1}^{m+r} \alpha_i \nabla f_i(x) = 0$. Temos então $\nabla f_0(x) + \sum_{i=1}^{m+r} (\lambda_i + \alpha_i) \nabla f_i(x) = 0$. Tomando $\alpha$ pequeno o suficiente de modo que $\lambda_i + \alpha_i \geq 0$ para $i \in A(x)$ temos que o multiplicador de Lagrange $\lambda$ não é único.

Notemos que a definição de SMFCQ depende previamente da existência de multiplicadores de Lagrange, 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 [50], isto é, se assumimos que para qualquer função objetivo, existem multiplicadores de Lagrange que cumprem SMFCQ, então vale LICQ. Com efeito, sendo $x^* \in \Omega$ e definindo $f_0(x) := -\sum_{i \in A(x^*)} f_i(x)$ temos que $x^*$ é um minimizador local de $f_0$ em $\Omega$. Supondo que existem únicos multiplicadores de Lagrange, a saber $\lambda_i^* = 1, i \in A(x^*)$ e $\lambda_i^* = 0$ caso contrário, se $\alpha$ é tal que $\sum_{i=1}^{m+r} \alpha_i \nabla f_i(x^*) = 0$, é fácil ver que para $C > 0$ grande o suficiente, $\bar{\lambda} := \lambda + \frac{\alpha}{C}$ também são multiplicadores de Lagrange, donde concluímos que $\alpha = 0$ e vale LICQ.

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 necessario exigir tal propriedade para todos os subconjuntos de gradientes, da maneira a seguir. Dizemos que vale a condição de posto constante (*constant rank constraint qualification*, CRCQ, [32]) em

$x \in \Omega$ se existe uma vizinhança $\mathcal{B}$ de $x$ tal que para todo $\mathcal{I} \subset \{1, \ldots, m+r\}$, se $\nabla f_i(x), i \in \mathcal{I}$ é linearmente dependente, então $\nabla f_i(y), i \in \mathcal{I}$ 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 $\mathcal{I} \subset \{1, \ldots, m+r\}$ o posto de $\nabla f_i(y), i \in \mathcal{I}$ não se altera, para cada $y \in \mathcal{B}$. Note que $A(x) = \{m+1, \ldots, m+r\}$, e quando analisamos restrições avaliadas em $y$ na vizinhança de $x$, algumas restrições podem deixar de ser ativas, alterando o conjunto de restrições ativas ou, possivelmente, tornando o ponto inviável (embora sempre possamos escolher a vizinhança pequena o suficiente de modo que as restrições inativas em $x$ permaneçam inativas em $y$, isto é, $A(y) \subset A(x)$). De qualquer maneira, o símbolo $r$ utilizado será sempre correspondente às restrições ativas em $x$.

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 $\mathcal{I} \subset \{1, \ldots, m+r\}$, o conjunto $\nabla f_i(y), i \in \mathcal{I}$ é linearmente independente, para todo $y \in \mathcal{B}$, o que mostra que CRCQ é mais fraca que LICQ.

Um caso frequente em modelagem é quando uma restrição de igualdade $h(x) = 0$ é descrita como duas restrições de desigualdade $f_1(x) := h(x) \leq 0$ e $f_2(x) := -h(x) \leq 0$. 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 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 [37]).

Da mesma maneira que MFCQ enfraquece LICQ olhando para o sinal dos multiplicadores, a condição de dependência linear positiva constante (CPLD, [45]) enfraquece CRCQ no mesmo sentido. Vale CPLD em $x \in \Omega$ se existe uma vizinhança $\mathcal{B}$ de $x$ tal que para todo $\mathcal{I} \subset \{1, \ldots, m+r\}$, se $\nabla f_i(x), i \in \mathcal{I}$ é positivo-linearmente dependente, então $\nabla f_i(y), i \in \mathcal{I}$ 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. Rigorosamente, deveríamos dizer que os gradientes $\nabla f_i(x), i \in \mathcal{I} \cap \{1, \ldots, m\}$ e $\nabla f_i(x), i \in \mathcal{I} \cap \{m+1, \ldots, m+r\}$ são positivo-linearmente dependentes, mas isto fica claro do contexto.

Mais recentemente, outro tipo de enfraquecimento destas condições surgiu na literatura. A ideia é não considerar todos os subconjuntos de restrições $\mathcal{I} \subset \{1, \ldots, m+r\}$. Dizemos que vale Relaxed-CRCQ (RCRCQ, [42]) se existe uma vizinhança $\mathcal{B}$ de $x \in \Omega$ tal que para todo $\{1, \ldots, m\} \subset \mathcal{I} \subset \{1, \ldots, m+r\}$, o posto de $\nabla f_i(y), i \in \mathcal{I}$ é constante para todo $y \in \mathcal{B}$. Ou seja, só é necessário verificar a propriedade de posto constante para subconjuntos que incluem todas as restrições de igualdade.

Para definirmos uma relaxação análoga para CPLD (Relaxed-CPLD, RCPLD, [9]), vamos interpretar a condição RCRCQ em termos de dependência linear constante. Seja $B \subset \{1, \ldots, m\}$ tal que $\nabla f_i(x), i \in B$ forma uma base para $\text{span}\{\nabla f_i(x)\}_{i=1}^m$. A condição RCRCQ pode ser equivalentemente definida como a existência de uma vizinhança $\mathcal{B}$ de $x$ tal que:

- $\{\nabla f_i(y)\}_{i=1}^m$ tem posto constante para todo $y \in B$
- para todo $\mathcal{I} \subset A(x)$, se $\{\nabla f_i(x)\}_{i \in B \cup \mathcal{I}}$ é linearmente dependente, então $\{\nabla f_i(y)\}_{i \in B \cup \mathcal{I}}$ 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, isto é, fixado $B \subset \{1, \ldots, m\}$ tal que $\nabla f_i(x), i \in B$ forma uma base para $\text{span}\{\nabla f_i(x)\}_{i=1}^m$, vale RCPLD em $x$ se

- $\{\nabla f_i(y)\}_{i=1}^m$ tem posto constante para todo $y \in B$
- para todo $\mathcal{I} \subset A(x)$, se $\{\nabla f_i(x)\}_{i \in B \cup \mathcal{I}}$ é positivo-linearmente dependente, então $\{\nabla f_i(y)\}_{i \in B \cup \mathcal{I}}$ é positivo-linearmente dependente $\forall y \in \mathcal{B}$

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 ser considerado para garantir a existência de multiplicadores de Lagrange. Considere o polar do cone linearizado

$$\mathcal{L}(x)^\circ = \left\{ v \in \mathbb{R}^n \mid v = \sum_{i=1}^{m+r} \lambda_i \nabla f_i(x), \lambda_i \geq 0, i \in A(x) \right\},$$

e considere o maior subespaço contido em $\mathcal{L}(x)^\circ$, isto é, $\text{span}\{\nabla f_i(x)\}_{i \in J_-(x)}$, onde $J_-(x) = \{i \in \{1, \ldots, m+r\} \mid -\nabla f_i(x) \in \mathcal{L}(x)^\circ\}$. É claro que $J_-(x)$ contém todos os índices de restrições de igualdade, mas contém também, possivelmente, alguns índices de restrições de desigualdade. Tais restrições de desigualdade correspondem a uma 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.

Dizemos que $x \in \Omega$ satisfaz CRSC quando existe uma vizinhança $\mathcal{B}$ de $x$ tal que os gradientes $\nabla f_i(y), i \in J_-(x)$ tem 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. O conceito de posto constante é adequado para lidar com restrições de igualdade, enquanto o conceito de dependência linear positiva se adequa 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 $\mathcal{L}(x)^\circ$ na soma direta entre seu maior subespaço e um cone pontudo, seja tal que a componente de subespaço mantenha localmente a dimensão.

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 de um *error bound*, isto é, a distância ao conjunto viável pode ser medida, em uma vizinhança de $x$, utilizando uma medida analítica de inviabilidade, isto é, 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:

$$\min_{z\in\Omega} \|z-y\| \leq \alpha \max\{|f_1(y)|, \dots, |f_m(y)|, \max\{0, f_{m+1}(y)\}, \dots, \max\{0, f_{m+p}(y)\}\}.$$

Esta propriedade é uma condição de qualificação, já que dado $d \in \mathcal{L}(x)$ facilmente verificamos por Taylor que $f_i(x+td) = o(t), i = 1, \dots, m, f_i(x+td) \leq o(t), i \in A(x)$ e $f_i(x+td) < 0, i = m+r+1, \dots, m+p$. Assim, $\text{dist}(x+td, \Omega) = o(t)$, com $\frac{o(t)}{t} \to 0$, o que não é difícil ver, equivale ao fato de $d \in \mathcal{T}(x)$, e portanto a condição de Abadie é satisfeita. A Figura 2.1 mostra a relação entre as condições 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 $f_0 := g + h$, onde $g$ é continuamente diferenciável e $h$ é convexa possivelmente não suave [17].

No caso suave, ou seja, quando $h = 0$, a condição mais fraca possível que garante KKT é a condição de Guignard. Terminamos esta sessão com a prova deste resultado de [29], onde corrigimos algumas pequenas imprecisões do trabalho original.

**Teorema 2.3** *Seja $x \in \Omega$. Então $\mathcal{L}(x)^\circ = \mathcal{T}(x)^\circ$ (condição de Guignard) é equivalente ao fato que para qualquer função $f_0$ que assume mínimo restrito a $\Omega$ em $x$, é tal que $x$ é um ponto KKT.*

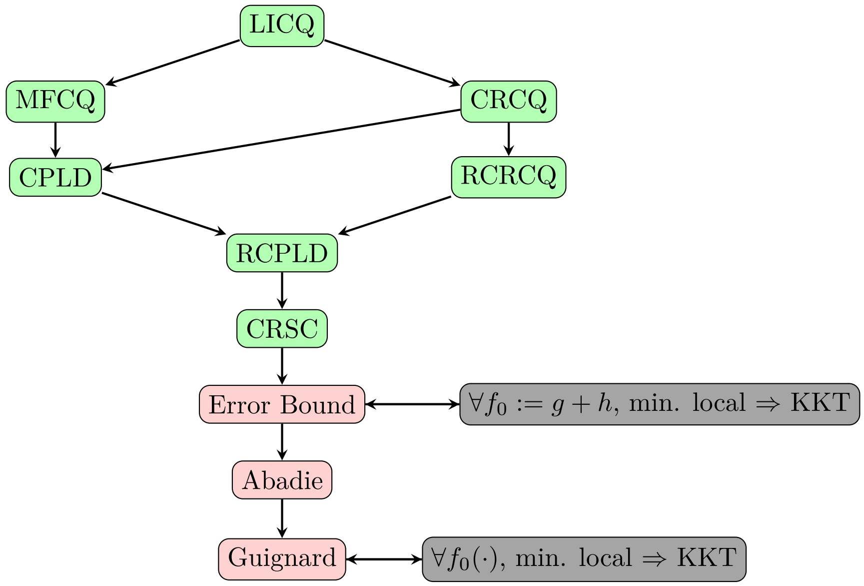



Figura 2.1: Relação entre as condições de qualificação apresentadas. $g$ é uma função $C^1$ e $h$ é uma função convexa (possivelmente não suave) quaisquer.

**Prova:** O fato que a condição de Guignard é uma condição de qualificação já foi mostrado ao longo do texto. Sabemos também que $\mathcal{L}(x)^\circ \subset \mathcal{T}(x)^\circ$. Agora, seja $0 \neq d \in \mathcal{T}(x)^\circ$ e vamos construir uma função suave $f_0$ que assume um mínimo local em $x$ (restrito a $\Omega$), de modo que $d = -\nabla f_0(x)$. A hipótese de $x$ ser estacionário para este problema resulta em $d \in \mathcal{L}(x)^\circ$. 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 $y \in T(x)$, o que contradiz o fato de $d \in T(x)^\circ$. 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 $d = (0, \dots, 0, 1)$, assim, definimos $P : \mathbb{R}^{n-1} \to \mathbb{R}$ no subespaço ortogonal a $d$ da seguinte maneira:

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

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 $f_0 : \mathbb{R}^n \to \mathbb{R}$ definida por $f_0(y) = P(z) - d^{\mathsf{T}} y$, onde $y = (z, w) \in \mathbb{R}^{n-1} \times \mathbb{R}$ e vamos mostrar que $f_0$ restrita a $\Omega$ assume um mínimo local em 0. Com efeito, seja $0 \neq y = (z, w) \in \mathbb{R}^{n-1} \times \mathbb{R}, y \in \Omega \cap \mathcal{B}(0, \varepsilon_3)$, 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 $d \in \mathcal{T}(0)^{\circ}$, 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 $f_0(0) = 0$ e $f_0(y) > 0$ para $0 \neq y \in S$ suficientemente pequeno, segue da hipótese que $x = 0$ é um ponto KKT, ou seja $d = -\nabla f_0(0) \in \mathcal{L}(x)^{\circ}$. $\square$

Observamos que a construção pode ser feita com $f_0$ diferenciável em todo $\mathbb{R}^n$ e de modo que $x$ seja o único minimizador global de $f_0$ restrito a $\Omega$. Ver [48], Teorema 6.11.

## 2.2 Condições sequenciais de primeira ordem

Na prática, um algoritmo iterativo não verifica condições do tipo "KKT ou não-CQ" onde CQ é alguma condição de qualificação. Isso porque um algoritmo gera uma sequência $\{x^k\}$, onde apenas no limite esta condição é verificada. Na prática, 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 candidado 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 $\varepsilon_k \to 0$. 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. A condição sequencial de otimalidade mais frequentemente utilizada na prática é a condição Aproximadamente-KKT que definimos a seguir:

**Definição 2.4 (Aproximadamente-KKT, [7])** *Dizemos que $x \in \Omega$ satisfaz a condição Aproximadamente-KKT (AKKT) se existem sequências $x^k \to x$, $\{\lambda^k\} \subset \mathbb{R}^{m+r}$ com $\lambda_i^k \geq 0, i \in A(x)$ tais que $\nabla f_0(x^k) + \sum_{i=1}^{m+r} \lambda_i^k \nabla f_i(x^k) \to 0$.*

Quando $\{x^k\}$ é uma sequência como na definição acima dizemos que $\{x^k\}$ é uma sequência AKKT.

O teorema a seguir mostra como verificar AKKT na prática

**Teorema 2.5** *O ponto $x$ satisfaz AKKT se, e só se, existem sequências $x^k \to x$, $\{\lambda^k\} \subset \mathbb{R}^{m+p}$ com $\lambda_i^k \geq 0, i = m+1, \dots, m+p$ e $\{\varepsilon_k\} \subset \mathbb{R}$, $\varepsilon_k \geq 0$ tais que*

$$\|(f_i(x^k))_{i=1}^m\| \leq \varepsilon_k, \|(\max\{0, f_i(x^k)\})_{i=m+1}^{m+p}\| \leq \varepsilon_k,$$
$$\|\nabla f_0(x^k) + \sum_{i=1}^{m+p} \lambda_i^k \nabla f_i(x^k)\| \leq \varepsilon_k,$$
$$\forall i = m+1, \dots, m+p, \text{ se } f_i(x^k) < -\varepsilon_k \text{ então } \lambda_i^k = 0,$$
$$\varepsilon_k \to 0^+.$$

**Prova:** Assuma que $x$ satisfaz AKKT e defina

$$\varepsilon_k = \max\{\|(f_i(x^k))_{i=1}^m\|, \|(\max\{0, f_i(x^k)\})_{i=m+1}^{m+p}\|, \|\nabla f_0(x^k) + \sum_{i=1}^{m+r} \lambda_i^k \nabla f_i(x^k)\|, -f_j(x^k), j \in A(x)\}.$$

Temos $\varepsilon_k \to 0^+$. Note que se $i \in \{m+1, \dots, m+p\}$ é tal que $f_i(x^k) < -\varepsilon_k$, então $-f_i(x^k) > \varepsilon_k \geq -g_j(x^k), \forall j \in A(x)$. Em particular, $i \notin A(x)$. Portanto, basta definir $\lambda_i^k = 0$ e temos o resultado.

Assumindo as condições do teorema, a continuidade das funções garante que $x \in \Omega$, e basta observar que se $i \notin A(x)$, então $f_i(x) < 0$, e portanto, para $k$ suficientemente grande vale $f_i(x^k) < -\varepsilon_k$, logo $\lambda_i^k = 0$ e vale AKKT. $\square$

Assim, se um algoritmo gera uma sequência que cumpre as condições acima com $\varepsilon_k \leq \varepsilon$, com $\varepsilon > 0$ uma tolerância pequena previamente estabelecida, é seguro dizer que a sequência $\{x^k\}$ gerada é, provavelmente, uma sequência AKKT, logo, o iterando $x^k$ está próximo de um ponto AKKT, um candidato à solução do problema.

Resta mostrar que AKKT é uma condição de otimalidade. De fato, é interessante observar que AKKT é satisfeita em uma solução do problema independente da validade de condições de qualificação.

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

$$\begin{array}{ll} \text{Minimizar} & f_0(x), \\ \text{Sujeito a} & f_i(x) = 0, i = 1, \dots, m, \\ & f_i(x) \leq 0, i = m+1, \dots, m+p, \\ & x \in \bar{\Omega}, \end{array}$$

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

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

$$\begin{array}{ll} \text{Minimizar} & f_0(x) + \rho_k P(x), \\ \text{Sujeito a} & x \in \bar{\Omega}, \end{array}$$

onde $P(x)$ é a função $P(x) = \|(f_i(x^k))_{i=1}^m\|^2 + \|(\max\{0, f_i(x^k)\})_{i=m+1}^{m+p}\|^2$.

Agora, seja $x^*$ um ponto limite arbitrário de $\{x^k\}$, caso exista. Temos $x^* \in \bar{\Omega}$. Para mostrar que $x^*$ é um minimiador 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 $f(x^k) + \rho_k P(x^k) > f(z) + \rho_k P(z) + (\rho_k c + f(x^k) - f(z))$. 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 que $x^*$ é uma solução, já que $f(x^k) \leq f(x^k) + \rho_k P(x^k) \leq f(z) + \rho_k P(z), \forall z \in \bar{\Omega}$, ou seja, 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.

Para mostrarmos que AKKT é uma condição de otimalidade, considere $x^*$ uma solução e vamos considerar a aplicação do algoritmo de penalidade externa para o problema

$$\begin{array}{ll} \text{Minimizar} & f_0(x) + \frac{1}{2}\|x - x^*\|^2, \\ \text{Sujeito a} & f_i(x) = 0, i = 1, \dots, m, \\ & f_i(x) \leq 0, i = m+1, \dots, m+p, \\ & \|x - x^*\|^2 \leq \delta, \end{array}$$

onde $\delta > 0$ é suficientemente pequeno e $\bar{\Omega} = \{x \mid \|x - x^*\|^2 \leq \delta\}$ é não vazio 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, todos os 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, $x^*$, isto é, $x^k \to x^*$.

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 $f_0(x) + \frac{1}{2}\|x - x^*\|^2 + \rho_k P(x)$ se anula em $x^k$, temos

$$\nabla f_0(x^k) + x^k - x^* + \sum_{i=1}^{m} (2\rho_k f_i(x^k)) \nabla f_i(x^k) + \sum_{i=m+1}^{m+p} (2\rho_k \max\{0, f_i(x^k)\}) \nabla f_i(x^k) = 0.$$

Definindo $\lambda_i^k = 2\rho_k f_i(x^k), i = 1, \dots, m$ e $\lambda_i^k = 2\rho_k \max\{0, f_i(x^k)\} \geq 0, i = m+1, \dots, m+p$ e observando que se $f_i(x^*) < 0$, então $\lambda_i^k = 0$ para $k$ suficientemente grande temos

$$\nabla f_0(x^k) + x^k - x^* + \sum_{i=1}^{m+r} \lambda_i^k \nabla f_i(x^k) = 0.$$

Tomando o limite temos que $x^*$ satisfaz AKKT. Temos então o seguinte resultado.

**Teorema 2.6** *Se $x^*$ é uma solução local, então $x^*$ satisfaz AKKT.*

Observação: A existência de multiplicadores Fritz John (Teorema 1.4) pode ser provada como uma consequência do Teorema 2.6. De fato, sendo $x^*$ uma solução local, considere sequências $\{x^k\}, \{\lambda^k\}, \lambda_i^k \geq 0, i \in A(x^*)$ como na Definição 2.4 de AKKT, de modo que $\varepsilon_k := \nabla f_0(x^k) + \sum_{i=1}^{m+r} \lambda_i^k \nabla f_i(x^k) \to 0$. Dividindo $\varepsilon_k$ por $\|(1, \lambda^k)\|$, temos $\alpha_0^k \nabla f_0(x^k) + \sum_{i=1}^{m+r} \alpha_i^k \nabla f_i(x^k) \to 0$ com $\|(\alpha_0^k, \alpha^k)\| = 1$ e $\alpha_i^k \geq 0, i \in \{0\} \cup A(x^*)$. Tomando o limite em uma subsequência tal que $\{(\alpha_0^k, \alpha^k)\}$ é convergente, seu limite é um multiplicador que satisfaz as condições do Teorema 1.4. 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 2.9 e 2.12 generalizam este resultado.

Em [10] foi mostrado que algoritmos do tipo Lagrangiano aumentado [4, 5], Programação Quadrática Sequencial [45, 44], Restauração Inexata [41, 26] e Pontos Interiores [24] geram sequências AKKT. Vamos mostrar este resultado apenas para o algoritmo de Lagrangiano aumentado definido abaixo:

Considere a função Lagrangiano aumentado:

$$L_\rho(x, \lambda) := f_0(x) + \frac{\rho}{2}\left(\sum_{i=1}^{m}\left(f_i(x) + \frac{\lambda_i}{\rho}\right)^2 + \sum_{i=m+1}^{m+p}\left(\max\left\{0, f_i(x) + \frac{\lambda_i}{\rho}\right\}\right)^2\right),$$

$\rho > 0, \lambda_i \geq 0, i = m+1, \ldots, m+p$. Definimos abaixo o algoritmo:

Passo 0 (inicialização): Sejam $\lambda_i^{\min} < \lambda_i^{\max}, i = 1, \ldots, m+p$, com $0 = \lambda_i^{\min}, i = m+1, \ldots, m+p; \gamma > 1, \tau \in (0,1), \varepsilon_k \geq 0, \varepsilon_k \to 0^+$. Sejam $\lambda^1 \in \mathbb{R}^{m+p}$ com $\lambda^{\min} \leq \lambda^1 \leq \lambda^{\max}, \rho_1 > 0$ e $x^0 \in \mathbb{R}^n$. Definimos $V_i^0 := f_i(x^0), i = 1, \ldots, m; V_i^0 = \max\{0, f_i(x^0)\}, i = m+1, \ldots, m+p$. Inicialize $k := 1$.

Passo 1 (minimização): Encontre um minimizador aproximado $x^k$ de $L_{\rho_k}(x, \lambda^k)$, isto é,

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

Passo 2 (atualização do parâmetro de penalidade): Defina $V_i^k := f_i(x^k), i = 1, \ldots, m; V_i^k := \max\{f_i(x^k), -\frac{\lambda_i^k}{\rho_k}\}, i = m+1, \ldots, m+p$. Se $\|V^k\| \leq \tau \|V^{k-1}\|$ definimos $\rho_{k+1} := \rho_k$, senão $\rho_{k+1} := \gamma \rho_k$.

Passo 3 (atualização do multiplicador de Lagrange): Calcule $\lambda^k \in \mathbb{R}^{m+p}$ com $\lambda^{\min} \leq \lambda^k \leq \lambda^{\max}$. Faça $k := k+1$ e vá para o Passo 1.

**Teorema 2.7** *Seja $x^*$ um ponto limite da sequência $\{x^k\}$ gerada pelo algoritmos. Então $x^*$ é um ponto estacionário para o problema de minimizar $\frac{1}{2}\sum_{i=1}^{m} f_i(x)^2 + \sum_{i=m+1}^{m+p} \max\{0, f_i(x)\}^2$.*

**Prova:** Se $\{\rho_k\}$ é limitada, $V^k \to 0$ e portanto $f_i(x^*) = 0, i = 1, \ldots, m; f_i(x^*) \leq 0, i = m+1, \ldots, m+p$. Senão, temos $\delta_k := \nabla f_0(x^k) + \sum_{i=1}^{m+p} \hat{\lambda}_i^k \nabla f_i(x^k)$ com $\|\delta_k\| \leq \varepsilon_k$, onde $\hat{\lambda}_i^k = \lambda_i^k + \rho_k f_i(x^k), i = 1, \ldots, m; \hat{\lambda}_i^k = \max\{0, \lambda_i^k + \rho_k f_i(x^k)\}, i = m+1, \ldots, m+p$. Dividindo $\delta_k$ por $\rho_k$, como $\{\lambda^k\}$ é limitada, observamos que $\frac{\hat{\lambda}_i^k}{\rho_k}$ converge para $f_i(x^*)$ se $i = 1, \ldots, m$ e para $\max\{0, f_i(x^*)\}$ se $i = m+1, \ldots, m+p$ e o resultado segue. $\square$

**Teorema 2.8** *Assuma que a sequência $\{x^k\}$ gerada pelo algoritmo de Lagrangiano aumentado admite um ponto limite viável $x$. Então $x$ é um ponto AKKT.*

**Prova:** Pela definição de $L_\rho$ e pelo Passo 1 do algoritmo temos

$$\|\nabla f_0(x^k) + \sum_{i=1}^{m+p} \hat{\lambda}_i^k \nabla f_i(x^k)\| \leq \varepsilon_k,$$

onde $\hat{\lambda}_i^k = \lambda_i^k + \rho_k f_i(x^k), i = 1, \ldots, m; \hat{\lambda}_i^k = \max\{0, \lambda_i^k + \rho_k f_i(x^k)\}, i = m+1, \ldots, m+p$. Se $f_i(x^*) < 0, i = m+1, \ldots, m+p$, para o caso $\rho_k \to +\infty$, como $\{\lambda_i^k\}$ é limitada, temos $\hat{\lambda}_i^k = 0$. Caso contrário, $V^k \to 0$ e portanto $-\frac{\lambda_i^k}{\rho_k} \to 0$. Logo $\hat{\lambda}_i^k = 0$. Segue que $x^*$ satisfaz AKKT. $\square$

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, CPLD, MFCQ, etc...

Para definir a condição CPG, considere vetores $\mathcal{V} = \{v_1, \ldots, v_{m+r}\} \in \mathbb{R}^n$ 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 $\mathcal{I} = \{1, \ldots, m\}, \mathcal{J} = \{m+1, \ldots, m+r\}$ e os vetores $v_1, \ldots, v_{m+r}$ estarão claros do contexto. Note que quando $v_i = \nabla f_i(x), i = 1, \ldots, m+r$, temos $\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V}) = \mathcal{L}(x)^\circ$.

Para cones desta forma (cones positivamente gerados), podemos definir uma noção de base. Diremos que o par ordenado $(\mathcal{I}', \mathcal{J}')$, $\mathcal{I}', \mathcal{J}' \subset \{1, \ldots, m+r\}$ 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 $\mathcal{I}' = \mathcal{I} \cup \{j\}$ $\mathcal{J}' = \mathcal{J} - \{j\}$ 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 $\text{span}_+(\mathcal{I}, \mathcal{J}; \mathcal{V})$.

Considerando $\mathcal{V}(x) = \{\nabla f_i(x)\}_{i=1}^{m+r}$, $\mathcal{I} = \{1, \dots, m\}$ e $\mathcal{J} = A(x)$, dizemos que vale CPG 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,

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

para todo $y$ em alguma vizinhança de $x$.

Considerando a decomposição de $\mathcal{L}(x)^\circ$ na soma direta de seu maior subespaço $\{v \mid v = \sum_{i \in J_-(x)} \lambda_i \nabla f_i(x)\}$ com o cone pontudo $\{v \mid v = \sum_{i \notin J_-(x)} \lambda_i \nabla f_i(x), \lambda_i \geq 0, \forall i\}$, podemos considerar apenas $\mathcal{I}' \subset J_-(x)$ tal que $v_i, i \in \mathcal{I}'$ forma uma base para a componente do subespaço, enquanto $\mathcal{J}' = \{i \mid i \notin J_-(x)\}$.

Considere o conjunto viável definido por $f_1(x_1, x_2) := x_1^3 - x_2 \leq 0, f_2(x_1, x_2) := x_1^3 + x_2 \leq 0, f_3(x_1, x_2) := x_1 \leq 0$. Em $x = (0, 0)$, $J_-(x) = \{1, 2\}$. Tomando $\mathcal{I}' = \{1\}$ e $\mathcal{J}' = \{3\}$ temos $\nabla f_1(x)$ e $\nabla f_3(x)$ 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, em $y \neq 0$, $\text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}(y))$ 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 $\{\nabla f_1(0), \nabla f_2(0)\}$ 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 2.9** *Se $x \in \Omega$ satisfaz CRSC, então $x$ satisfaz CPG*

**Prova:** 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 $\text{span}\{\nabla f_i(x), i \in J_-(x)\}$, sendo o posto de $\{\nabla f_i(y), i \in J_-(x)\}$ constante, esta base deve permanecer uma base em uma vizinhança. $\square$

Incluindo a restrição $f_4(x_1, x_2) := x_2^3 \leq 0$ 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. Com isso concluímos a conexão entre condições pontuais e sequenciais.

**Teorema 2.10** *Se $x \in \Omega$ é AKKT e satisfaz CPG, então $x$ cumpre KKT.*

**Prova:** Sejam $\{x^k\} \subset \mathbb{R}^n$, $x^k \to x$ e $\{\lambda^k\} \subset \mathbb{R}^{m+r}$ com $\lambda_i^k \geq 0, i \in A(x)$ e $\nabla f_0(x^k) + \sum_{i=1}^{m+r} \lambda_i^k \nabla f_i(x^k) \to 0$.

Sendo $(\mathcal{I}', \mathcal{J}')$ uma base para $\mathcal{L}(x)^\circ$ tal que $\text{span}_+(\{1, \ldots, m\}, A(x); \mathcal{V}(y)) \subset \text{span}_+(\mathcal{I}', \mathcal{J}'; \mathcal{V}(y))$ para $y$ em uma vizinhança de $x$, temos para $k$ suficientemente grande que 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 $\nabla f_0(x^k) + \sum_{i \in \mathcal{I}' \cup \mathcal{J}'} \lambda_i^k \nabla f_i(x^k) \to 0$.

Sendo $M_k = \max\{|\bar{\lambda}_i^k|, i \in \mathcal{I}' \cup \mathcal{J}'\}$, se $\{M_k\}$ é ilimitada, dividindo por $M_k$ podemos tomar uma subsequência tal que $\frac{\bar{\lambda}_i^k}{M_k} \to \alpha_i$ 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.

Se $\{M_k\}$ é limitada podemos tomar uma subsequência tal que $\bar{\lambda}_i^k \to \lambda_i$, $\lambda_i \geq 0, i \in \mathcal{J}'$ e tomando o limite temos $\nabla f_0(x) + \sum_{i \in \mathcal{I}' \cup \mathcal{J}'} \lambda_i \nabla f_i(x) = 0$. Sendo $(\mathcal{I}', \mathcal{J}')$ base para $\mathcal{L}(x)^\circ$ temos que $x$ é KKT. □

Acreditava-se que CPG era a condição de qualificação mais fraca possível que garante que AKKT implica KKT. Entretanto, em [46, 11], é definida a condição de qualificação CCP (*cone continuity property*) com tal propriedade. CCP é estritamente mais fraca que CPG, embora CCP implique Abadie.

**Definição 2.11 (CCP, [46])** *Diremos que $x \in \Omega$ satisfaz a propriedade do cone contínuo (CCP) quando a multifunção $K(\cdot)$ é semi-contínua exteriormente em $x$, isto é,* $\operatorname{limsup}_{y \to x} K(y) \subset K(x)$*, onde*

$$K(y) = \{v \in \mathbb{R}^n \mid v = \sum_{i=1}^{m+r} \lambda_i \nabla f_i(y), \lambda_i \geq 0, i \in A(x)\}$$

*e*

$$\operatorname*{limsup}_{y \to x} K(y) = \{w \in \mathbb{R}^n \mid \exists y^k \to x, \exists w^k \to w, w^k \in K(y^k)\}.$$

**Teorema 2.12 ([46])** *$x \in \Omega$ satisfaz CCP, se e somente se, para qualquer função objetivo $f_0$ tal que $x$ é AKKT, vale que $x$ é KKT.*

**Prova:** Seja $x \in \Omega$ tal que vale CCP e $f_0$ uma função objetivo tal que $x$ é AKKT. Logo existem sequências $x^k \to x$, $\{\lambda^k\} \subset \mathbb{R}^{m+r}, \lambda_i^k \geq 0, i \in A(x)$ tais que $\nabla f_0(x^k) + \omega^k \to 0$, onde $\omega^k = \sum_{i=1}^{m+r} \lambda_i^k \nabla f_i(x^k) \in K(x^k)$.

Como $\omega^k \to -\nabla f_0(x)$, temos que $-\nabla f_0(x) \in \operatorname{limsup}_{y \to x} K(y)$. Da condição CCP temos $-\nabla f_0(x) \in K(x) = \mathcal{L}(x)^\circ$ e $x$ satisfaz KKT.

Reciprocamente, seja $\omega \in \operatorname{limsup}_{y \to x} K(y)$. Logo existem sequências $x^k \to x$ e $\omega^k \to \omega$ com $\omega^k \in K(x^k)$. Para provar que $\omega \in K(x)$, definimos a função objetivo $f_0(y) = -\sum_{i=1}^n \omega_i y_i$. Do fato de $\omega^k \in K(x^k)$, existe $\{\lambda^k\} \subset \mathbb{R}^{m+r}, \lambda_i^k \geq 0, i \in A(x)$ tal que $\omega^k = \sum_{i=1}^{m+r} \lambda_i^k \nabla f_i(x^k)$, assim, como $\nabla f_0(x^k) = -\omega$ e $\omega^k \to \omega$ temos $\nabla f_0(x^k) + \sum_{i=1}^{m+r} \lambda_i^k \nabla f_i(x^k) \to 0$ e $x$ cumpre AKKT. Da hipótese, $x$ é KKT, ou seja, $-\nabla f_0(x) = \omega \in K(x)$ e CCP é satisfeita. □

Na Figura 2.2 apresentamos um diagrama completo de relações entre diversas condições de qualificação.

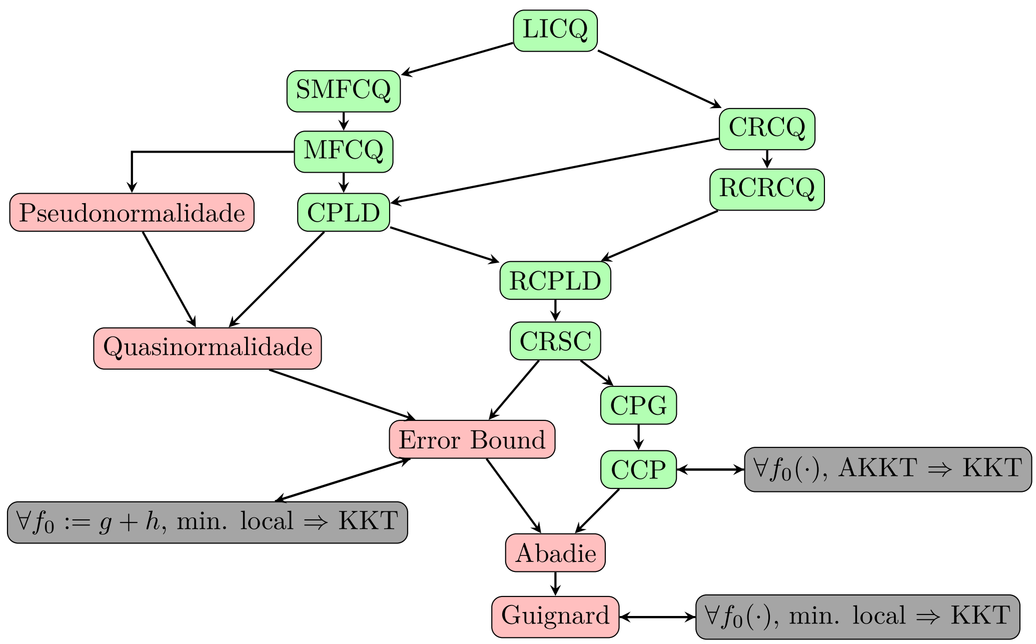



Figura 2.2: Relações entre condições de qualificação. $g$ é uma função $C^1$ e $h$ é uma função convexa (possivelmente não suave) quaisquer.

# Capítulo 3

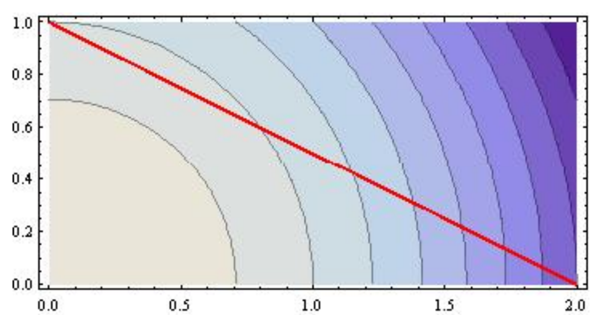
# Condições de otimalidade de segunda ordem

Neste capítulo vamos assumir que as funções $f_i : \mathbb{R}^n \to \mathbb{R}, i = 0, \ldots, m+p$ são duas vezes continuamente diferenciáveis.

Considere o problema

$$\begin{array}{ll} \text{Minimizar} & f_0(x) := -x_1^2 - x_2^2, \\ \text{Sujeito a} & f_1(x) := x_1 + 2x_2 - 2 \leq 0, \\ & f_2(x) := -x_1 \leq 0, \\ & f_3(x) := -x_2 \leq 0. \end{array}$$

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

$$\nabla f_0(x) + \lambda_1 \nabla f_1(x) = 0, \lambda_1 \geq 0$$

é satisfeita com $\lambda_1 = 0.8$, entretanto, observamos que $x$ não é uma solução. De fato, a partir de $x$, a função objetivo $f_0$ aumenta ao longo de direções que apontam para o interior da região viável (ou seja, direções $d \in \mathbb{R}^2$ tais que $\nabla f_1(x)^\mathtt{T} d < 0$), mas ao longo de direções viáveis ortogonais ao gradiente de $f_1$, 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. 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 $\nabla f_1(x)^\mathtt{T} d = 0$, temos $f_0(x + d) \approx f_0(x) + \frac{1}{2} d^\mathtt{T} \nabla^2 f_0(x) d$ e como $d^\mathtt{T} \nabla^2 f_0(x) d < 0$ 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 $L(x) = f_0(x) + \sum_{i=1}^{m+p} \lambda_i f_i(x)$, para um multiplicador de Lagrange $\lambda$ fixado, que subestima $f_0$ em pontos viáveis.

**Definição 3.1** *Dizemos que o ponto KKT $x \in \Omega, A(x) = \{m+1, \dots, m+r\}$ satisfaz a condição necessária (forte) de segunda ordem associada ao multiplicador de Lagrange $\lambda$ (associado a $x$) quando*

$$d^\mathtt{T} \left( \nabla^2 f_0(x) + \sum_{i=1}^{m+r} \lambda_i \nabla^2 f_i(x) \right) d \geq 0, \forall d \in C^S(x),$$

*onde*

$$C^S(x) = \{d \in \mathbb{R}^n \mid \nabla f_i(x)^\mathtt{T} d = 0, i = 1, \dots, m, \nabla f_i(x)^\mathtt{T} d \leq 0, i \in A(x) \cup \{0\}\}$$

*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:

$$C^S(x) = \left\{ d \in \mathbb{R}^n \mid \begin{array}{l} \nabla f_i(x)^\mathtt{T} d = 0, i \in \{1, \dots, m\} \cup \{i \in \{m+1, \dots, m+r\} \mid \lambda_i > 0\} \\ \nabla f_i(x)^\mathtt{T} d \leq 0, i \in \{m+1, \dots, m+r \mid \lambda_i = 0\} \end{array} \right\}$$

Assim, para todo $d \in C^S(x)$ temos $\nabla f_0(x)^\mathtt{T} d = 0$.

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

$$L(x, \lambda) := L(x, (\lambda_i)_{i=1}^m, (\lambda_i)_{i=m+1}^{m+p}) = f_0(x) + \sum_{i=1}^{m+p} \lambda_i f_i(x),$$

temos que a condição de otimalidade de primeira ordem em um ponto $x \in \Omega$ é satisfeita se existe $\lambda \in \mathbb{R}^{m+p}$ com $\lambda_i \geq 0, i = m+1, \dots, m+p$ e $\lambda_i = 0, i =$

$m+1,\dots,m+p$ com $i \notin A(x)$ tal que $\nabla_x L(x,\lambda)=0$. Além disso, a condição de segunda ordem associada a $\lambda$ é satisfeita se $\nabla^2_{xx}L(x,\lambda)$ é semidefinida positiva em $C^S(x)$. Quando temos um multiplicador de Lagrange $\lambda \in \mathbb{R}^{m+r}$ associado a $x^*$ com $A(x^*)=\{m+1,\dots,m+r\}$ faremos o abuso de notação $L(x,\lambda)$ significando $L(x,\bar{\lambda})$, onde $\bar{\lambda} \in \mathbb{R}^{m+p}$ é o vetor $\lambda$ aumentado de zeros nas posições $m+r+1,\dots,m+p$.

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 $\nabla^2_{xx}L(x,\lambda)$ 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

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 1.4) para o caso em que as funções possuem segunda derivada contínua.

**Teorema 3.2 (Condição de Fritz John de segunda ordem)** *Se $x^* \in \Omega$ é uma solução local, então para cada $d \in C^S(x^*)$ existe $0 \neq (\lambda_0,\lambda) \in \mathbb{R}\times\mathbb{R}^{m+r}$ com $\sum_{i=0}^{m+r}\lambda_i\nabla f_i(x^*)=0, \lambda_0 \geq 0, \lambda_i \geq 0, i \in A(x^*)$ tal que*

$$d^{\mathtt{T}}\left(\sum_{i=0}^{m+r}\lambda_i\nabla^2 f_i(x)\right)d \geq 0.$$

**Prova:** No caso em que $\{\nabla f_i(x^*)\}_{i=1}^m$ é linearmente dependente com $\sum_{i=1}^m\lambda_i\nabla f_i(x^*)=0$ e $\lambda \neq 0$, tomamos um multiplicador Fritz John satisfazendo as condições considerando $\lambda$ ou $-\lambda$ e os demais $\lambda_i$'s nulos. Caso contrário, fixamos $d \in C^S(x^*)$ e consideramos o seguinte problema de otimização linear nas variáveis $z$ e $w$:

$$\begin{array}{ll}
\text{Minimizar} & z, \\
\text{Sujeito a} & \nabla f_i(x^*)^{\mathtt{T}}w + d^{\mathtt{T}}\nabla^2 f_i(x^*)d = 0, i=1,\dots,m, \\
 & \nabla f_i(x^*)^{\mathtt{T}}w + d^{\mathtt{T}}\nabla^2 f_i(x^*)d \leq z, i \in \{0\}\cup\{i \in A(x^*) \mid \nabla f_i(x^*)^{\mathtt{T}}d=0\}
\end{array}$$

Aplicando o Teorema da Função Implícita, é possível mostrar que o problema acima possui valor ótimo finito não-negativo (ver detalhes em [23], página 443).

Assim, o problema dual abaixo nas variáveis $\lambda_i, i = 0, \ldots, m + r$ possui o mesmo valor ótimo:

$$
\begin{array}{ll}
\text{Maximizar} & d^{\mathtt{T}} \left( \sum_{i=0}^{m+r} \lambda_i \nabla^2 f_i(x) \right) d, \\
\text{Sujeito a} & \sum_{i=0}^{m+r} \lambda_i \nabla f_i(x^*) = 0, \\
& \sum_{i=0}^{m+r} \lambda_i = 1, \\
& \lambda_i \geq 0, i \in \{0\} \cup \{i \in A(x^*) \mid \nabla f_i(x^*)^{\mathtt{T}} d = 0\}, \\
& \lambda_i = 0, i \in A(x^*), \nabla f_i(x^*)^{\mathtt{T}} d < 0.
\end{array}
$$

A solução ótima do problema acima é um multiplicador do tipo Fritz John que satisfaz as condições do enunciado. □

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 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 é resolvido assumindo a condição de qualificação de Mangasarian-Fromovitz.

**Teorema 3.3 (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 $\lambda \in \mathbb{R}^{m+r}, \lambda_i \geq 0, i \in A(x^*)$ associado a $x^*$ tal que*

$$
d^{\mathtt{T}} \left( \nabla^2 f_0(x) + \sum_{i=1}^{m+r} \lambda_i \nabla^2 f_i(x) \right) d \geq 0.
$$

**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 $\lambda$ 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 3.4** *Suponha que $x^* \in \Omega$ é um minimizador local que satisfaz SMFCQ. Logo o único multiplicador de Lagrange $\lambda \in \mathbb{R}^{m+r}, \lambda_i \geq 0, i \in A(x^*)$ 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 [13] (originalmente proposto em [15]) mostra que existem problemas que cumprem MFCQ na solução, onde a condição necessária de otimalidade de segunda ordem não é satisfeita em nenhum multiplicador:

$$\begin{array}{ll} \text{Minimizar} & x_3, \\ & x_3 \geq (x_1, x_2)^{\mathtt{T}} Q_k (x_1, x_2), k = 0, \ldots, 4, \end{array}$$

onde $Q_k = \begin{pmatrix} \cos(\frac{k\pi}{4}) & -\sin(\frac{k\pi}{4}) \\ \sin(\frac{k\pi}{4}) & \cos(\frac{k\pi}{4}) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} \cos(\frac{k\pi}{4}) & \sin(\frac{k\pi}{4}) \\ -\sin(\frac{k\pi}{4}) & \cos(\frac{k\pi}{4}) \end{pmatrix}, k = 0, \ldots, 3.$

O seguinte exemplo, mais simples, analisado na origem, é apresentado em [18].

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

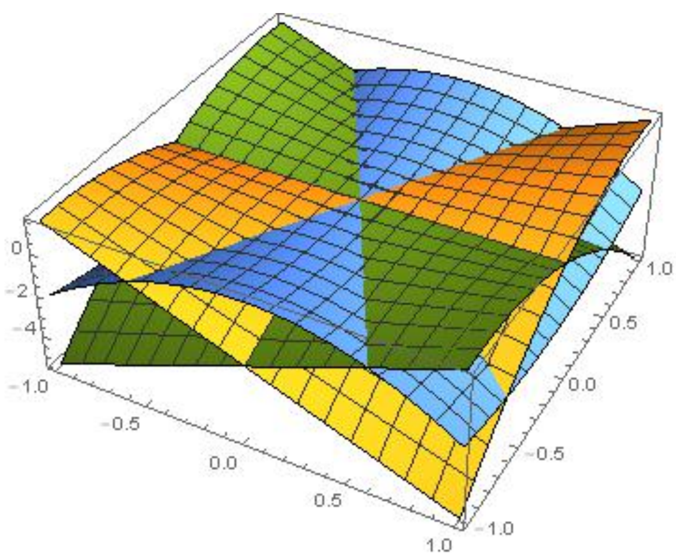



Figura 3.1: MFCQ não garante a validade da condição de otimalidade de segunda ordem

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 $\lambda_1+\lambda_2+\lambda_3 = 1, \lambda_1, \lambda_2, \lambda_3 \geq 0$. 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 2.1, 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], observamos 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 3.5** *Suponha que $x^* \in \Omega$ é um minimizador local que satisfaz RCRCQ. Então qualquer multiplicador de Lagrange $\lambda \in \mathbb{R}^{m+r}, \lambda_i \geq 0, i \in A(x^*)$ 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 ([49], Teorema 2.9) descrita abaixo:

**Teorema 3.6 (Posto Constante, [32])** *Seja $\{h_i(x), i \in K\}$ uma família finita de funções de classe $C^k, k \geq 1$, $h_i : \mathbb{R}^n \to \mathbb{R}^m$ tais que o posto de $\{\nabla h_i(y), i \in K\}$ 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 $h_i(\phi^{-1}(x + d))$ é constante para todo $d$ com $x + d \in V$ tal que $\nabla h_i(x)^\mathtt{T} d = 0, i \in K$.*

**Lema 3.7 ([6])** *Assuma que $x \in \Omega$ satisfaz RCRCQ. Então para cada $d$ tal que $\nabla f_i(x)^\mathtt{T} d = 0, i = 1, \ldots, m$, $\nabla f_i(x)^\mathtt{T} d \leq 0, i \in A(x)$ existe um função de mesma classe que as $f_i$'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 $\nabla f_i(x)^\mathtt{T} d = 0$ temos $f_i(\xi(t)) = 0$ para todo $t \in (-\varepsilon, \varepsilon)$.*

**Prova:** Fixado $d$ nas condições do enunciado, defina $K = \{i \in \{1, \ldots, m + r\} \mid \nabla f_i(x)^\mathtt{T} d = 0\}$. Como vale RCRCQ podemos usar o Teorema do Posto Constante com a família $h_i := f_i, i \in K$, obtendo um difeomorfismo $\phi$ com as propriedades mencionadas. Definimos $\xi(t) = \phi^{-1}(x + td)$ para todo $t$ em algum intervalo aberto contendo a origem. Utilizando a expansão de Taylor para $\phi^{-1}$ temos $\xi'(0) = \lim_{t\to 0} \frac{\xi(t)-x}{t} = d$.

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 $\nabla f_i(x)^\mathtt{T} d < 0$ temos $f_i(\xi(t)) = t\nabla f_i(x)^\mathtt{T} d + o(t)$, $\frac{o(t)}{t} \to 0$, logo $f_i(\xi(t)) < 0$ para $t$ suficientemente pequeno. Além disso, se $i \notin A(x)$, reduzindo possivelmente o intervalo temos claramente $f_i(\xi(t)) < 0$. $\square$

**Prova (Teorema 3.5)**: Assuma que $x^* \in \Omega$ satisfaz RCRCQ e seja $\lambda \in \mathbb{R}^{m+r}$ com $\lambda_i \geq 0, i \in A(x^*)$ 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

$$\begin{aligned} \nabla f_i(x^*)^{\mathsf{T}} d = 0, &\quad i \in \{0, 1, \dots, m\} \cup \{i \in \{m+1, \dots, m+r\} \mid \lambda_i > 0\}, \\ \nabla f_i(x^*)^{\mathsf{T}} d \le 0, &\quad i \in \{i \in \{m+1, \dots, m+r\} \mid \lambda_i = 0\}. \end{aligned}$$

Definimos $\phi(t) = f_0(\xi(t))$, onde $\xi : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ é obtida aplicando o Lema 3.7. Já que $\xi(t)$ é viável para $t \ge 0$ e $\xi(0) = x^*$ temos que $t^* = 0$ é um minimizador local de $\phi(t), t \ge 0$. Logo, $\phi''(0) \ge 0$. Como $\xi'(0) = d$ temos $\phi''(0) = d^{\mathsf{T}} \nabla^2 f_0(x^*) d + \nabla f_0(x)^{\mathsf{T}} \xi''(0) \ge 0$.

Sendo $f_i(\xi(t)) = 0$ para todo $t$ e $i \in \{1, \dots, m\} \cup \{i \in \{m+1, \dots, m+r\} \mid \lambda_i > 0\}$ temos

$$R(t) = \sum_{i=1}^{m+r} \lambda_i f_i(\xi(t)) = 0, -\varepsilon < t < \varepsilon.$$

Assim,

$$R''(0) = d^{\mathsf{T}} (\sum_{i=1}^{m+r} \lambda_i \nabla^2 f_i(x^*)) d + (\sum_{i=1}^{m+r} \lambda_i \nabla f_i(x^*))^{\mathsf{T}} \xi''(0) = 0.$$

Logo, $\phi''(0) + R''(0) = d^{\mathsf{T}} (\nabla f_0(x^*) + \sum_{i=1}^{m+r} \lambda_i \nabla f_i(x^*)) d \ge 0$ e a condição de otimalidade de segunda ordem é satisfeita. $\square$

Podemos considerar uma versão mais fraca de RCRCQ, a saber, exigindo posto constante apenas para o conjunto de todos os gradientes de restrições ativas $\{\nabla f_i(x)\}_{i=1}^{m+r}$. Esta condição é denominada condição de posto constante fraca (*weak constant rank*, WCR [12]) e embora não seja uma condição de qualificação (de fato, considere o problema de minimizar $f_0(x) := -x_1$, sujeito a $f_1(x) := x_2 - x_1^2 = 0, f_2(x) := -x_1 \le 0, f_3(x) = x_2 \le 0$, a solução $x^* = (0,0)$ não satisfaz KKT, mas o posto de $\{\nabla f_i(x)\}_{i=1}^3$ é constante em uma vizinhança da origem), 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 3.8** *Dizemos que o ponto KKT $x \in \Omega, A(x) = \{m+1, \dots, m+r\}$ satisfaz a condição necessária fraca de otimalidade de segunda ordem associada ao multiplicador de Lagrange $\lambda$ (associado a $x$) quando*

$$d^{\mathsf{T}} \left( \nabla^2 f_0(x) + \sum_{i=1}^{m+r} \lambda_i \nabla^2 f_i(x) \right) d \ge 0, \forall d \in C^W(x),$$

*onde*

$$C^W(x) = \{d \in \mathbb{R}^n \mid \nabla f_i(x)^{\mathsf{T}} d = 0, i = 0, \dots, m+r\}$$

*é o cone crítico fraco.*

**Teorema 3.9 ([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 $f_i(\xi(t)) = 0, i = 1, \ldots, m+r$ e a prova segue como no Teorema 3.5. □

Note que o cone crítico fraco é um 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 aplicações. 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 3.4, é a condição SMFCQ que garante a condição forte para o único multiplicador de Lagrange existente. Em [12] 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 3.9, em [12] 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 [18] 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 [19] 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 [20] 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 [12] 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 [21, 14, 16]. Resultados de segunda ordem para problemas em dimensão infinita podem ser obtidos em [22] e resultado para problemas multi-objetivo em [38, 39].

A seguir descrevemos os resultados de [2].

Consideremos $\Omega = \{x \mid f_i(x) = 0, i = 1, \ldots, m, f_i(x) \le 0, i = m+1, \ldots, m+p\}$ o conjunto viável do problema original e, fixado $x^* \in \Omega$ com $A(x^*) = \{m+1, \ldots, m+r\}$, definimos o conjunto viável auxiliar $\Omega' = \{x \mid f_i(x) = 0, i = 1, \ldots, m+r\}$. 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 1.1 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 3.10** *Se $x^* \in \Omega$ satisfaz WCR, então $x^*$ satisfaz a condição de Abadie para as restrições de igualdade $f_i(x) = 0, i = 1, \ldots, m+r$.*

**Prova:** Seja $0 \neq d \in \mathcal{L}_{\Omega'}$, isto é, $\nabla f_i(x^*)^{\mathsf{T}} d = 0, i = 1, \ldots, m+r$. Como na prova do Lema 3.7 e Teorema 3.9, 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, $d = \lim_{t \to 0^+} \frac{\xi(t) - x^*}{t}$. Assim, sendo $\xi'(0) \neq 0$, podemos reduzir $\varepsilon$ se necessário para que $\xi(t) \neq x^*$. Logo $\frac{\xi(t)-x^*}{\|\xi(t)-x^*\|} = \frac{\xi(t)-x^*}{t} \frac{t}{\|\xi(t)-x^*\|} \to \frac{d}{\|d\|}$ 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 3.9. Para isto, provamos primeiramente o lema abaixo:

**Lema 3.11** *Seja $x^* \in \Omega$ um minimizador local, $\lambda \in \mathbb{R}^{m+r}$ um multiplicador de Lagrange associado a $x^*$. Então*

$$d^{\mathsf{T}} \left( \nabla^2 f_0(x^*) + \sum_{i=1}^{m+r} \lambda_i \nabla^2 f_i(x^*) \right) d \ge 0,$$

*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, para todo $i = 1, \ldots, m+p$ com $\lambda_i > 0$ vale $f_i(x^k) = o(\|x^k - x^*\|^2)$. (Em particular, a condição é satisfeita quando $f_i(x^k) = 0, i \in \{1, \ldots, m\} \cup \{i \in A(x^*) \mid \lambda_i > 0\}$.)*

**Prova:** Temos $L(x^k,\lambda) = f_0(x^k) + \sum_{i=1}^{m+r} \lambda_i f_i(x^k) = f_0(x^k) + o(\|x^k - x^*\|^2)$, assim, sendo $x^*$ um minimizador local e $x^k$ viável, temos

$$\begin{aligned}
0 &\leq f_0(x^k) - f_0(x^*) \\
&= L(x^k,\lambda) - L(x^*,\lambda) + o(\|x^k - x^*\|^2) \\
&= \nabla_x L(x^*,\lambda)^{\mathtt{T}}(x^k - x^*) + \frac{1}{2}(x^k - x^*)^{\mathtt{T}}\nabla^2_{xx}L(\bar{x}^k,\lambda)(x^k - x^*) + o(\|x^k - x^*\|^2) \\
&= \frac{1}{2}(x^k - x^*)^{\mathtt{T}}\nabla^2_{xx}L(\bar{x}^k,\lambda,\mu)(x^k - x^*) + o(\|x^k - x^*\|^2),
\end{aligned}$$

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 $d^{\mathtt{T}}\nabla L^2_{xx}(x^*,\lambda)d \geq 0$. $\square$

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

**Teorema 3.12** *Seja $x^* \in \Omega$ um minimizador local e assuma que a condição de Abadie é satisfeita em $x^*$ considerando as restrições $f_i(x) = 0, i = 1, \ldots, m + r$. 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 3.11 e o resultado segue. $\square$

Quando o conjunto viável possui apenas igualdades, isto é, $\Omega = \Omega'$ com $A(x^*) = \emptyset$, a hipótese do Teorema 3.12 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 3.13** *Assuma 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 3.14** *Seja $x \in \Omega$ um ponto KKT. Definimos $A^0(x)$ o conjunto de índices $i$ de restrições de desigualdade ativas em $x$ tais que $\lambda_i = 0$ para qualquer multiplicador de Lagrange associado a $x$. O conjunto $A^+(x) = A(x) \backslash A^0(x)$ é o conjunto de índices $i$ de restrições de desigualdade ativas em $x$ tais que $\lambda_i > 0$ para algum multiplicador de Lagrange associado a $x$.*

Sendo $x \in \Omega$ um ponto KKT, para cada $i \in A^+(x)$ existe um multiplicador de Lagrange $\lambda^i \in \mathbb{R}^{m+r}$ associado a $x$ tal que $\lambda^i_i > 0$, assim, $\lambda = \frac{\sum_{i \in A^+(x)} \lambda^i}{|A^+(x)|}$ é um multiplicador de Lagrange associado a $x$ tal que $\lambda_i > 0$ para todo $i \in A^+(x)$. Portanto o cone crítico forte pode ser descrito como

$$C^S(x) = \{d \in \mathbb{R}^n \mid \nabla f_i(x)^\mathtt{T} d = 0, i \in \{1, \ldots, m\} \cup A^+(x); \nabla f_i(x)^\mathtt{T} d \leq 0, i \in A^0(x)\}.$$

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

**Lema 3.15** *Se $x \in \Omega$ é um ponto KKT, então*

$$A^0(x) = \{i \in A(x^*) \mid \exists d \in C^S(x), \nabla f_i(x)^\mathtt{T} d < 0\}.$$

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

$$\begin{array}{ll} \text{Maximizar}_\lambda & \lambda_j, \\ \text{Sujeito a} & \sum_{i=1}^{m+r} \lambda_i \nabla f_i(x) = -\nabla f_0(x), \\ & \lambda_i \geq 0, i \in A(x). \end{array}$$

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

$$\begin{array}{ll} \text{Minimizar}_d & \nabla f_0(x)^\mathtt{T} d, \\ \text{Sujeito a} & \nabla f_i(x)^\mathtt{T} d = 0, i = 1, \ldots, m, \\ & \nabla f_i(x)^\mathtt{T} d \leq 0, i \in A(x) \backslash \{j\}, \\ & \nabla f_j(x)^\mathtt{T} d \leq -1. \end{array}$$

A solução $d$ deste problema satisfaz as restrições e é tal que $\nabla f_0(x)^\mathtt{T} d = 0$, logo $d \in C^S(x)$ e $\nabla f_j(x)^\mathtt{T} d < 0$, o que mostra que $A^0(x) \subset \{i \in A(x^*) \mid \exists d \in C^S(x), \nabla f_i(x)^\mathtt{T} d < 0\}$. A inclusão recíproca é óbvia. □

**Corolário 3.16** *Seja $x \in \Omega$ um ponto KKT. Então existe $d \in C^S(x)$ tal que*

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

**Prova:** Basta somar os vetores dados pelo Lema 3.15 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 3.17** *Seja $x^* \in \Omega$ um minimizador local e assuma que a condição de Abadie é satisfeita em $x^*$ considerando as restrições*

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

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

**Prova:** Seja $0 \neq d \in C^S(x^*)$, isto é, $\nabla f_i(x^*)^{\mathtt{T}} d = 0, i \in \{1,\ldots,m\} \cup A^+(x^*)$ e $\nabla f_i(x^*)^{\mathtt{T}} d \leq 0, i \in A^0(x^*)$. Logo, por hipótese, segue que $d \in \mathcal{T}_{\Omega^+}(x^*)$. Ou seja, existe $x^k \to x^*$, $f_i(x^k) = 0, i \in \{1,\ldots,m\} \cup A^+(x^*)$, $f_i(x^k) \leq 0, i \in A^0(x^*)$ tal que $\frac{x^k - x^*}{\|x^k - x^*\|} \to \frac{d}{\|d\|}$. Do Lema 3.11 segue que $d^{\mathtt{T}} \nabla^2_{xx} L(x^*, \lambda) d \geq 0$ para qualquer multiplicador de Lagrange $\lambda$ associado a $x^*$. $\square$

A seguir mostramos que o Teorema 3.17 generaliza o Teorema 3.5.

**Teorema 3.18** *Se $x^* \in \Omega$ satisfaz RCRCQ, então $x^*$ satisfaz a condição de Abadie para o conjunto de restrições $f_i(x) = 0, i \in \{1,\ldots,m\} \cup A^+(x^*); f_i(x) \leq 0, i \in A^0(x^*)$.*

**Prova:** É uma adaptação da prova do Teorema 3.10. $\square$

Note que a definição do conjunto $\Omega^+$ só faz sentido quando $x^*$ é um ponto KKT. Neste caso, a hipótese do Teorema 3.17 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 3.17 pode ser reformulada para desprezarmos as restrições em $A^0(x^*)$ na definição do cone tangente.

**Teorema 3.19** *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 $F^+ = \{x \mid f_i(x) = 0, i \in \{1,\ldots,m\} \cup A^+(x^*)\}$.*

**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 $h \in C^S(x^*)$ dado pelo Corolário 3.16 tal que $\nabla f_i(x^*)^{\mathtt{T}} h = 0, i \in \{1,\ldots,m\} \cup A^+(x^*)$ e $\nabla f_i(x^*)^{\mathtt{T}} h < 0, i \in A^0(x^*)$ e defina $d^k = \frac{d + \frac{1}{k} d^k}{\|d + \frac{1}{k} d^k\|}$, segue que $d^k \in C^S(x^*)$ com $\nabla f_i(x^*)^{\mathtt{T}} d^k = 0, i \in \{1,\ldots,m\} \cup A^+(x^*)$ e $\nabla f_i(x^*)^{\mathtt{T}} d^k < 0, i \in A^0(x^*)$ para todo $k$.

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

Em [20] é apresentado um resultado similar ao Teorema 3.17.

**Teorema 3.20** *Fixado um ponto KKT $x^* \in \Omega$, e fixado um multiplicador de Lagrange $\lambda \in \mathbb{R}^{m+r}$ associado a $x^*$, definimos o conjunto de restrições*

$$\Omega_\lambda = \left\{ x \mid \begin{array}{l} f_i(x) = 0, i \in \{1,\ldots,m\} \cup \{i \in A(x^*) \mid \lambda_i > 0\}; \\ f_i(x) \leq 0, i \in \{i \in A(x^*) \mid \lambda_i = 0\} \end{array} \right\}.$$

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

**Prova:** A demonstração é feita exatamente como na prova do Teorema 3.17.$\square$

Observamos que independente de $\lambda$, $\mathcal{L}_{\Omega_\lambda}(x^*) = C^S(x^*)$ e que $\Omega^+ \subset \Omega_\lambda$. O Teorema 3.20 garante que se vale Abadie para $\Omega_\lambda$, então vale a condição de segunda ordem para este multiplicador $\lambda$ e para todos os outros $\lambda'$ tais que $\Omega_\lambda \subset \Omega_{\lambda'}$, ou seja, se o conjunto de índices de multiplicadores positivos de $\lambda$ contém o conjunto de índices de multiplicadores positivos de $\lambda'$. Se $\lambda$ é tomado com $\lambda_i > 0$ para todo $i \in A^+(x^*)$ e a condição do Teorema 3.20 é satisfeita, então $\Omega_\lambda \subset \Omega_{\lambda'}$ para qualquer multiplicador de Lagrange $\lambda'$ associado a $x^*$ e como $\Omega_\lambda = \Omega^+$, a consequência é a mesma do Teorema 3.17.

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 3.20 seja satisfeita para algum multiplicador, mas não para todos. De fato, considere o problema

$$\begin{array}{ll} \text{Minimizar} & x_2, \\ \text{Sujeito a} & f_1(x) := -x_1^2 - x_2 \leq 0, \\ & f_2(x) := -x_2 \leq 0, \\ & f_3(x) := x_1 \leq 0 \end{array}$$

na solução $x^* = (0, 0)$. Temos que os multiplicadores são da forma $\lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_1 + \lambda_2 = 1, \lambda_3 = 0$. 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, $\lambda_1 = \lambda_2 = \frac{1}{2}, \lambda_3 = 0$ temos $\Omega_\lambda = \Omega^+ = \{(0, 0)\}$, 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] mostramos também que se um minimizador local $x^* \in \Omega$ satisfaz MFCQ e para todo $\mathcal{I} \subset \{1, \ldots, m\} \cup A^+(x^*)$ com cardinalidade um a menos que $\{1, \ldots, m\} \cup A^+(x^*)$ vale que $\{\nabla f_i(x)\}_{i \in \mathcal{I}}$ 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

Em [30], é considerado o problema quadrático em $\mathbb{R}^4$ de minimizar $f_0(x) := \frac{1}{2} x^\mathtt{T} H x$, sujeito a $x \geq 0$, onde $H = I - \frac{3}{2} \frac{zz^\mathtt{T}}{\|z\|^2}$ 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 parâmetro de barreira $\mu_k \to 0^+$, definimos as funções $b_k(x) = f_0(x) - \mu_k \sum_{i=1}^4 \log(x_i)$

(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 $e_1^{\mathtt{T}} H e_1 < 0$. 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. De fato, a mera verificação da condição forte de segunda ordem é um problema NP-difícil [43]. Por estes motivos, em se tratando de algoritmos práticos, a condição de otimalidade de segunda ordem de interesse é a condição fraca. Há diversos algoritmos na literatura que com pequenas modificações tem convergência global a pontos estacionários de segunda ordem: Lagrangiano aumentado, programação quadrática sequencial, pontos interiores, regiões de confiança, entre outros.

Nesta sessão vamos definir um análogo de segunda ordem para a condição sequencial de otimalidade AKKT e vamos mostrar a relação desta condição com condições de otimalidade pontuais de segunda ordem. Em particular, estamos interessados em condições de qualificação (CQ2) tais que

$$\text{WSONC ou não-CQ2}$$

é uma condição de otimalidade, onde WSONC (weak second order necessary optimality condition) é a condição necessária de otimalidade fraca de segunda ordem.

**Definição 3.21** *Dizemos que $x \in \Omega$ é um ponto Aproximadamente Estacionário de Segunda Ordem (AKKT2) se existem sequências $x^k \to x$, $\{\lambda^k\} \subset \mathbb{R}^{m+r}, \lambda_i^k \geq 0, i = m+1, \dots, m+r, \{\theta^k\} \subset \mathbb{R}^{m+r}, \theta_i^k \geq 0, i = 1, \dots, m+r$, $\delta_k \geq 0, \delta_k \to 0^+$ tais que*

$$\nabla_x L(x^k, \lambda^k) = \nabla f_0(x^k) + \sum_{i=1}^{m+r} \lambda_i^k \nabla f_i(x^k) \to 0,$$

$$\nabla_{xx}^2 L(x^k, \lambda^k) + \sum_{i=1}^{m+r} \theta_i^k \nabla f_i(x^k) \nabla f_i(x^k)^{\mathtt{T}} \succeq -\delta_k Id,$$

*para $k$ suficientemente grande, onde $A \succeq B$ significa que a matriz $A - B$ é semidefinida positiva e $Id$ é a matriz identidade.*

Vamos mostrar que AKKT2 é uma condição de otimalidade (sequencial), isto é, se $x^* \in \Omega$ é um minimizador local, então $x^*$ satisfaz AKKT2. Para isso, vamos utilizar o lema abaixo de [12]:

**Lema 3.22** *Se $f : \mathbb{R}^n \to \mathbb{R}, g_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \dots, p$ admitem segunda derivada contínua na vizinhança de $x^*$, onde $x^*$ é um minimizador local de $F(x) = f(x) + \frac{1}{2}\sum_{i=1}^p g_i(x)_+^2$, sendo $g_i(x)_+ = \max\{0, g_i(x)\}$, então $H(x) = \nabla^2 f(x) + \sum_{i=1}^p g_i(x)_+ \nabla^2 g_i(x) + \sum_{i|g_i(x^*)\geq 0} \nabla g_i(x) \nabla g_i(x)^{\mathtt{T}}$ é semidefinida positiva em $x^*$.*

**Prova:** Defina $\bar{F}(x) = f(x) + \frac{1}{2}\sum_{i|g_i(x^*)\geq 0} g_i(x)^2$. Note que $\bar{F}(x^*) = F(x^*)$, $\nabla\bar{F}(x^*) = \nabla F(x^*)$ e que $\bar{F}$ é duas vezes continuamente diferenciável em $x^*$ sendo $\nabla^2\bar{F}(x^*) = H(x^*)$. Sendo $F(x) \leq \bar{F}(x)$, suponha que $H(x^*)$ não é semi-definida positiva. Então $x^*$ não é um minimizador local de $\bar{F}$. Logo existe uma sequência $x^k \to x^*$ com $F(x^k) \leq \bar{F}(x^k) < \bar{F}(x^*) = F(x^*)$, o que contradiz a definição de $x^*$. □

**Teorema 3.23** *Se $x^* \in \Omega$ é um minimizador local, entao $x^*$ satisfaz AKKT2.*

**Prova:** Consideramos uma construção análoga à prova do Teorema 2.6, a saber, vamos aplicar o método de penalidade externa ao problema:

$$\begin{array}{ll} \text{Minimizar} & f_0(x) + \frac{1}{4}\|x - x^*\|^4, \\ \text{Sujeito a} & f_i(x) = 0, i = 1, \ldots, m, \\ & f_i(x) \leq 0, i = m+1, \ldots, m+p, \\ & \|x - x^*\|^2 \leq \delta, \end{array} .$$

obtendo uma sequência $x^k \to x^*$ definida como uma solução global de minimizar $F(x) := f_0(x) + \frac{1}{4}\|x - x^*\|^4 + \rho_k(\sum_{i=1}^{m} f_i(x)^2 + \sum_{i=m+1}^{m+p} f_i(x)_+^2)$, sujeito a $\|x - x^*\|^2 \leq \delta$. Sendo $\|x^k - x^*\|^2 < \delta$ para $k$ suficientemente grande, temos que $x^k$ é um minimizador local irrestrito de $F(x)$ e portanto pelo Lema 3.22 temos:

$$\nabla f_0(x^k) + \sum_{i=1}^{m} 2\rho_k f_i(x^k)\nabla f_i(x^k) + \sum_{i=m+1}^{m+p} 2\rho_k f_i(x^k)_+\nabla f_i(x^k) + \|x^k - x^*\|^3(x^k - x^*) = 0,$$

$$\nabla^2 f_0(x^k) + \sum_{i=1}^{m} 2\rho_k f_i(x^k)\nabla^2 f_i(x^k) + \sum_{i=m+1}^{m+p} 2\rho_k f_i(x^k)_+\nabla^2 f_i(x^k) + \sum_{i=1}^{m} 2\rho_k \nabla f_i(x^k)\nabla f_i(x^k)^{\mathtt{T}}$$
$$+ \sum_{i=m+1|f_i(x^k)\geq 0}^{m+p} 2\rho_k \nabla f_i(x^k)\nabla f_i(x^k)^{\mathtt{T}} + 2(x^k - x^*)(x^k - x^*)^{\mathtt{T}} + 3\|x^k - x^*\|^2 Id \succeq 0.$$

Definindo $\lambda_i^k = 2\rho_k f_i(x^k), i = 1, \ldots, m; \lambda_i^k = 2\rho_k f_i(x^k)_+, i = m+1, \ldots, m+p; \theta_i^k = 2\rho_k, i \in \{1, \ldots, m\} \cup \{i \in \{m+1, \ldots, m+p\} \mid f_i(x^k) \geq 0\}; \theta_i^k = 0, i \in \{i \in \{m+1, \ldots, m+p\} \mid f_i(x^k) < 0\}; \delta_k = 3\|x^k - x^*\|^2$ notemos que $\lambda_i^k = 0$ e $\theta_i^k = 0$ para $i = m+r+1, \ldots, m+p$, além disso, $\delta_k \to 0$ e $2(x^k - x^*)(x^k - x^*)^{\mathtt{T}} \succeq 0$, portanto, $x^*$ satisfaz AKKT2. □

Em [12] um algoritmo de Lagrangiano aumentado de segunda ordem é construído com convergência global para a condição de otimalidade WSONC ou não-CQ2 onde a condição de qualificação CQ2 usada é MFCQ + WCR. Mostraremos que AKKT2 é mais forte que WSONC ou não-(MFCQ+WCR) e que o algoritmo de Lagrangiano aumentado de segunda ordem, entre outros, gera sequências AKKT2. Um fato principal é o lema abaixo de [12]:

**Lema 3.24** *Se $x \in \Omega$ satisfaz WCR, então para todo $d \in C^W(x)$ e para toda sequência $x^k \to x$, existe sequência $d^k \to d$ tal que $\nabla f_i(x^k)^{\mathtt{T}} d^k = 0, i = 1, \ldots, m+r$ para $k$ suficientemente grande.*

**Prova:** Seja $I \subset \{1,\dots,m+r\}$ tal que $\{\nabla f_i(x)\}_{i\in I}$ seja uma base para $\text{span}\{\nabla f_i(x)\}_{i=1}^{m+r}$. Como $x$ satisfaz WCR, $\{\nabla f_i(y)\}_{i\in I}$ é uma base para $\text{span}\{\nabla f_i(y)\}_{i=1}^{m+r}$ se $y$ está suficientemente próximo de $x$. Definimos a matriz $\nabla f_I(x)$ sendo $\nabla f_i(x)$ sua $i$-ésima coluna, $i \in I$. Definindo $d^k = [Id - \nabla f_I(x^k)(\nabla f_I(x^k)^\mathtt{T}\nabla f_I(x^k))^{-1}\nabla f_I(x^k)^\mathtt{T}]d$, a projeção de $d$ no espaço ortogonal às colunas de $\nabla f_I(x^k)$, temos que $d^k$ está bem definido para $k$ suficientemente grande e $\nabla f_i(x^k)^\mathtt{T} d^k = 0, i = 1,\dots,m+r$. Tomando o limite, pela continuidade das funções temos $d^k \to [Id - \nabla f_I(x)(\nabla f_I(x)^\mathtt{T}\nabla f_I(x))^{-1}\nabla f_I(x)^\mathtt{T}]d$, a projeção de $d$ em $C^W(x)$, que coincide com $d$ já que $d \in C^W(x)$. $\square$

**Teorema 3.25** *Seja $x \in \Omega$ tal que vale AKKT2. Se MFCQ e WCR valem em $x$, então existe um multiplicador de Lagrange associado a $x$ que satisfaz a condição de otimalidade fraca de segunda ordem.*

**Prova:** Da definição de AKKT2 existem sequências $x^k \to x$, $\{\lambda^k\} \subset \mathbb{R}^{m+r}, \lambda_i^k \geq 0, i = m+1,\dots,m+r, \{\theta^k\} \subset \mathbb{R}^{m+r}, \theta_i^k \geq 0, i = 1,\dots,m+r$, $\delta_k \geq 0, \delta_k \to 0^+$ tais que

$$\varepsilon_k := \nabla_x L(x^k,\lambda^k) \to 0,$$

$$M_k := \nabla^2_{xx} L(x^k,\lambda^k) + \sum_{i=1}^{m+r} \theta_i^k \nabla f_i(x^k)\nabla f_i(x^k)^\mathtt{T} \succeq -\delta_k Id.$$

A sequência $\{\lambda^k\}$ é limitada, caso contrário, tomando o limite em uma subsequência para $\frac{\varepsilon_k}{\|\lambda^k\|}$ teríamos uma contradição com MFCQ. Vamos tomar uma subsequência tal que $\lambda^k \to \lambda, \lambda_i \geq 0, i \in A(x^*)$. O limite $\varepsilon_k \to 0$ mostra que $\lambda$ é um multiplicador de Lagrange associado a $x$. Seja $d \in C^W(x^*)$ e tome $d^k \to d$ dado pelo Lema 3.24. Segue que

$$(d^k)^\mathtt{T} M_k d^k = (d^k)^\mathtt{T}\nabla^2_{xx}L(x^k,\lambda^k)d^k + \sum_{i=1}^{m+r}\theta_i^k(\nabla f_i(x^k)^\mathtt{T} d^k)^2 \geq -\delta_k\|d^k\|^2,$$

e portanto $(d^k)^\mathtt{T}\nabla^2_{xx}L(x^k,\lambda^k)d^k \geq -\delta_k\|d^k\|^2$. Tomando o limite em $k$ temos $d^\mathtt{T}\nabla^2_{xx}L(x,\lambda)d \geq 0$. $\square$

Vamos mostrar que o algoritmo de Lagrangiano aumentado de segunda ordem [12, 3] gera sequências AKKT2. Em [8] mostramos que o algoritmo SQP regularizado [28], de regiões de confiança [25], entre outros, também geram sequências AKKT2.

Como o lagrangiano aumentado $L_\rho$ não é duas vezes diferenciável nos pontos em que $f_i(x) + \frac{\lambda_i}{\rho} = 0, i = m+1,\dots,m+p$, definimos o simbolo $\bar{\nabla}^2$ que coincide com $\nabla^2$ quando a função é duas vezes diferenciável e $\bar{\nabla}^2\left(\max\left\{0, f_i(x) + \frac{\lambda_i}{\rho}\right\}\right)^2 := \nabla^2\left(f_i(x) + \frac{\lambda_i}{\rho}\right)^2$ se $f_i(x) + \frac{\lambda_i}{\rho} = 0, i = m+1,\dots,m+p$. O algoritmo de Lagrangiano aumentado de segunda ordem é o mesmo definido para o Teorema 2.8 com a diferença que o Passo 1 é substituído por:

Passo 1' (minimização): Encontre um minimizador aproximado $x^k$ de $L_{\rho_k}(x, \lambda^k)$, isto é,

$$\|\nabla L_{\rho_k}(x^k, \lambda^k)\| \leq \varepsilon_k \text{ e } \bar{\nabla}^2 L_{\rho_k}(x^k, \lambda^k) \succeq -\varepsilon_k Id.$$

**Teorema 3.26** *Assuma que a sequência $\{x^k\}$ gerada pelo algoritmo de Lagrangiano aumentado de segunda ordem admite um ponto limite viável $x$. Então $x$ é um ponto AKKT2.*

**Prova:** O fato que $x$ satisfaz AKKT foi demonstrado no Teorema 2.8. De fato

$$\|\nabla f_0(x^k) + \sum_{i=1}^{m+p} \hat{\lambda}_i^k \nabla f_i(x^k)\| \leq \varepsilon_k,$$

onde $\hat{\lambda}_i^k = \lambda_i^k + \rho_k f_i(x^k), i = 1, \ldots, m; \hat{\lambda}_i^k = \max\{0, \lambda_i^k + \rho_k f_i(x^k)\}, i = m + 1, \ldots, m+r$, $\hat{\lambda}_i^k = 0, i = m+r+1, \ldots, m+p$. Pelo Passo 1' do algoritmo temos:

$$\nabla^2 L(x^k, \hat{\lambda}^k) + \rho_k \sum_{i=1}^{m} \nabla f_i(x^k) \nabla f_i(x^k)^{\mathtt{T}} + \rho_k \sum_{f_i(x^k) + \frac{\lambda_i^k}{\rho_k} \geq 0, i=m+1}^{m+p} \nabla f_i(x^k) \nabla f_i(x^k)^{\mathtt{T}} \succeq \varepsilon_k Id.$$

Com argumento análogo ao do Teorema 2.8 verificamos que o índice do segundo somatório varia em $A(x)$ e portanto vale AKKT2. $\square$

A seguir vamos provar um resultado análogo ao Teorema 2.12. Isto é, vamos definir uma condição de qualificação, chamada *continuidade do cone de segunda ordem*, CCP2, mais fraca que MFCQ+WCR de modo que CCP2 em $x^*$ é equivalente ao fato que para qualquer função objetivo $f_0$ tal que AKKT2 é satisfeita em $x^*$ implica que $x^*$ satisfaz a condição fraca de segunda ordem para algum multiplicador de Lagrange.

Fixado $x^* \in \Omega$, definimos para cada $x \in \mathbb{R}^n$, o cone $C^W(x, x^*) = \{d \in \mathbb{R}^n \mid \nabla f_i(x)^{\mathtt{T}} d = 0, i \in \{1, \ldots, m\} \cup A(x^*)\}$. Note que $C^W(x^*, x^*) = C^W(x^*)$. Definimos então o cone

$$K_2^W(x) := \bigcup_{\lambda \in \mathbb{R}^m \times \mathbb{R}_+^r} \left\{ (\sum_{i=1}^{m+r} \lambda_i \nabla f_i(x), H) \mid H \preceq \sum_{i=1}^{m+r} \lambda_i \nabla^2 f_i(x) \text{ em } C^W(x, x^*), H \text{ simétrica} \right\}.$$

Note que a condição necessária fraca de segunda ordem (WSONC) é dada por $(-\nabla f_0(x^*), -\nabla^2 f_0(x^*)) \in K_2^W(x^*)$. Definimos abaixo a condição de qualificação CCP2 e mostramos que esta é a condição de qualificação mais fraca possível associada a AKKT2.

**Definição 3.27** *Dizemos que $x^* \in \Omega$ satisfaz CCP2 se a multifunção $K_2^W(x)$ é semi-contínua exteriormente em $x^*$, isto é,* $\limsup_{x \to x^*} K_2^W(x) \subset K_2^W(x^*)$.

**Teorema 3.28** *$x^* \in \Omega$ satisfaz CCP2, se e somente se, para qualquer função objetivo $f_0$ tal que $x^*$ é AKKT2, vale que $x^*$ satisfaz WSONC para algum multiplicador de Lagrange.*

**Prova:** Veja [8]. □

Os Teoremas 3.25 e 3.28 garantem que CCP2 é mais fraca que MFCQ+WCR (de fato, considerando $\Omega = \{x \mid x \leq 0, -x \leq 0\} \subset \mathbb{R}$ verificamos que CCP2 é estritamente mais fraca que MFCQ+WCR). O teorema a seguir garante que CCP2 é mais fraca que RCRCQ. Este resultado, em particular, mostra que algoritmos que geram sequências AKKT2 tem convergência global para pontos estacionários de segunda ordem sob RCRCQ.

**Teorema 3.29** *Se $x^* \in \Omega$ satisfaz RCRCQ, então $x^*$ satisfaz CCP2.*

**Prova:** Veja [8]. □

Considerando o conjunto viável $\Omega = \{(x_1, x_2) \mid x_1 = 0, -x_1^2 + x_2 \leq 0, -x_1^2 + x_2^3 \leq 0\}$ verificamos que a recíproca do Teorema 3.29 não é válida.

A seguir mostramos a relação entre as condições de qualificação apresentadas.

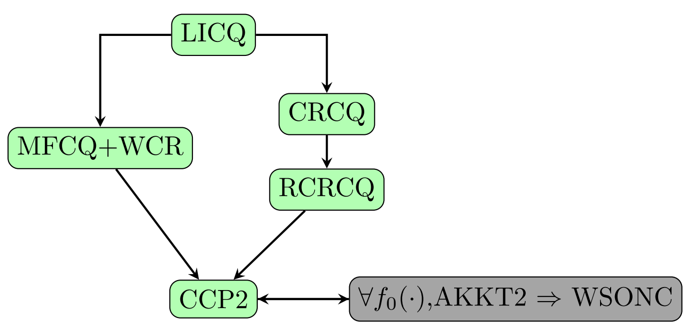



Figura 3.2: Relação entre as condições de qualificação apresentadas.

# Capítulo 4

# Conclusões

Em geral, um algoritmo de primeira ordem gera uma sequência $\{x^k\}$ de iterandos cujos pontos limites satisfazem uma condição de otimalidade pontual do tipo KKT ou não-CQ, para alguma condição de qualificação (CQ). Esta condição não fornece naturalmente um critério de parada para o algoritmo, pois ela só pode ser verificada no limite. Descrevemos a teoria de condições sequenciais de otimalidade, em particular, a condição AKKT, que justifica critérios práticos de parada utilizados por diversos algoritmos. O desenvolvimento desta teoria permite a prova de convergência global de algoritmos a pontos KKT sob condições de qualificação mais fracas do que as usuais. O mesmo pode ser feito para algoritmos de segunda ordem. Com a teoria de segunda ordem desenvolvida provamos que a convergência global de algoritmos de segunda ordem pode ser obtida sob condições de qualificação mais fracas, em particular, sob a condição de posto constante. Como trabalho futuro pretendemos generalizar a teoria desenvolvida para problemas de otimização de outra natureza, como otimização semi-definida. Outra questão relevante de pesquisa está relacionada ao problema de medir a distância entre um iterando e uma solução do problema quando um algoritmos para satisfazendo AKKT com uma tolerância $\varepsilon > 0$ pequena.

# Bibliografia

[1] J.M. Abadie. On the Kuhn-Tucker theorem. In J.M. Abadie, editor, *Nonlinear Programming*, pages 21–36. John Wiley, 1967.

[2] R. Andreani, R. Behling, G. Haeser, and P.J.S. Silva. On second order optimality conditions for nonlinear optimization. 2014. submitted.

[3] R. Andreani, E. G. Birgin, J. M. Martinez, and M. L. Schuverdt. Second-order negative-curvature methods for box-constrained and general constrained optimization. *Computational Optimization and Applications*, 45:209–236, 2010.

[4] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt. On augmented lagrangian methods with general lower–level constraints. *SIAM Journal on Optimization*, 18:1286–1309, 2007.

[5] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt. Augmented lagrangian methods under the constant positive linear dependence constraint qualification. *Mathematical Programming*, 112:5–32, 2008.

[6] R. Andreani, C. E. Echagüe, and M. L. Schuverdt. Constant-Rank Condition and Second-Order Constraint Qualification. *Journal of Optimization Theory and Applications*, 146(2):255–266, February 2010.

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

[8] R. Andreani, G. Haeser, A. Ramos, and P.J.S. Silva. On second-order sequential optimality conditions for nonlinear optimization and applications. 2015. submitted.

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

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

[11] R. Andreani, J. M. Martinez, A. Ramos, and P.J.S. Silva. A cone-continuity constraint qualification and algorithmic consequences. *Optimization Online*, February, 2015.

[12] R. Andreani, J. M. Martínez, and M. L. Schuverdt. On second-order optimality conditions for nonlinear programming. *Optimization*, 56(5-6):529–542, 2007.

[13] M. Anitescu. Degenerate nonlinear programming with a quadratic growth condition. *SIAM Journal on Optimization*, 10(4):1116–1135, 2000.

[14] A. V. Arutyunov. Smooth abnormal problems in extremum theory and analysis. *Russian Math. Surveys*, 67(3):403–457, 2012.

[15] A.V. Arutyunov. Perturbations of extremal problems with constraints and necessary optimality conditions. *Journal of Soviet Mathematics*, 54(6):1342–1400, 1991.

[16] E.R. Avakov, A.V. Arutyunov, and A.F. Izmailov. Necessary conditions for an extremum in a mathematical programming problem. *Proceedings of the Steklov Institute of Mathematics*, 256(1):2–25, 2007.

[17] D. Azé. A characterization of the Lagrange-Karush-Kuhn-Tucker property. *Optimization Online*, December, 2014.

[18] A. Baccari. On the Classical Necessary Second-Order Optimality Conditions. *Journal of Optimization Theory and Applications*, 123(1):213–221, October 2004.

[19] A. Baccari and A. Trad. On the Classical Necessary Second-Order Optimality Conditions in the Presence of Equality and Inequality Constraints. *SIAM Journal on Optimization*, 15(2):394–408, January 2005.

[20] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty. *Nonlinear Programming: Theory and Algorithms*. Wiley, third edition, 2006.

[21] A. Ben-Tal. Second-order and related extremality conditions in nonlinear programming. *Journal of Optimization Theory and Applications*, 31(2):143–165, 1980.

[22] A. Ben-Tal and J. Zowe. A unified theory of first and second order conditions for extremum problems in topological vector spaces. In Monique Guignard, editor, *Optimality and Stability in Mathematical Programming*, volume 19 of *Mathematical Programming Studies*, pages 39–76. Springer Berlin Heidelberg, 1982.

[23] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, 2000.

[24] L. Chen and D. Goldfarb. Interior-point $\ell_2$-penalty methods for nonlinear programming with strong global convergence properties. *Mathematical Programming*, 108:1–26, 2006.

[25] J. E. Dennis and L. N. Vicente. On the convergence theory of trust-region-based algorithms for equality-constrained optimization. *SIAM Journal of Optimization*, 7(4):927–950, 1997.

[26] A. Fischer and A. Friedlander. A new line search inexact restoration approach for nonlinear programming. *Computational Optimization and Applications*, 46(2):333–346, 2010.

[27] J. Gauvin. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Mathematical Programming*, 12:136–138, 1977.

[28] P. E. Gill, V. Kungurtsev, and D. P. Robinson. A regularized SQP method with convergence to second-order optimal points. *Optimization Online*, 2013.

[29] F.J. Gould and J.W. Tolle. A necessary and sufficient qualification for constrained optimization. *SIAM Journal on Applied Mathematics*, 20:164–172, 1969.

[30] N. I. M. Gould and P. L. Toint. A note on the convergence of barrier algorithms to second-order necessary points. *Mathematical Programming*, 85(2):433–438, June 1999.

[31] M. Guignard. Generalized Kuhn-Tucker conditions for mathematical programming problems in Banach space. *SIAM Journal on Control*, 7:232–241, 1969.

[32] R. Janin. Directional derivative of the marginal function in nonlinear programming. *Mathematical Programming Study*, 21:110–126, 1984.

[33] F. John. Extremum problems with inequalities as subsidiary conditions. In K.O. Friedrichs et al., editor, *Studies and Essays, Courant Anniversary Volume*, pages 187–204. Wiley/Interscience, 1948.

[34] W. Karush. Minima of functions of several variables with inequalities as side constraints. Master's thesis, Dept. of Mathematics, Univ. of Chicago, 1939.

[35] H.W. Kuhn and A.W. Tucker. Nonlinear programming. In J. Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492. University of California Press, 1951.

[36] J. Kyparisis. On uniqueness of Kuhn-Tucker multipliers in nonlinear programming. *Mathematical Programming*, 32:242–246, 1985.

[37] Shu Lu. Implications of the constant rank constraint qualification. *Mathematical Programming*, 126(2):365–392, 2011.

[38] M. C. Maciel, S. A. Santos, and G. N. Sottosanto. On second-order optimality conditions for vector optimization. *Journal of Optimization Theory and Applications*, 149(2):332–351, 2011.

[39] M. C. Maciel, S. A. Santos, and G. N. Sottosanto. On second-order optimality conditions for vector optimization: Addendum. *Journal of Optimization Theory and Applications*, pages 1–6, 2012.

[40] O.L. Mangasarian and S. Fromovitz. The Fritz John optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 17:37–47, 1967.

[41] J. M. Martínez and E. A. Pilotta. Inexact restoration methods for nonlinear programming: advances and perspectives. In L.Q. Qi, K.L. Teo, and X.Q. Yang, editors, *Optimization and Control with applications*, pages 271–292. Springer, 2005.

[42] L. Minchenko and S. Stakhovski. Parametric Nonlinear Programming Problems under the Relaxed Constant Rank Condition. *SIAM Journal on Optimization*, 21(1):314–332, 2011.

[43] K. G. Murty and S. N. Kabadi. Some np-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39(2):117–129, 1987.

[44] E. R. Panier and A. L. Tits. On combining feasibility, descent and superlinear convergence in inequality constrained optimization. *Mathematical Programming*, 59(1-3):261–276, March 1993.

[45] L. Qi and Z. Wei. On the constant positive linear dependence condition and its applications to SQP methods. *SIAM Journal on Optimization*, 10(4):963–981, 2000.

[46] A. Ramos. *Tópicos em condições de otimalidade para otimização não linear*. PhD thesis, IME-USP, Departamento de Matemática Aplicada, 2015.

[47] R.T. Rockafellar. Lagrange multipliers and optimality. *SIAM Review*, 35(2):183–238, 1993.

[48] R.T. Rockafellar and R. J-B Wets. *Variational Analysis*. Springer, 2009.

[49] M. Spivak. *A Comprehensive Introduction to Differential Geometry*, volume 1. Publish or Perish, third edition, 1999.

[50] G. Wachsmuth. On LICQ and the uniqueness of Lagrange multipliers. *Operations Research Letters*, 41(1):78 – 80, 2013.