Universidade Estadual de Campinas
Instituto de Matemática Estatística e Computação Científica
Departamento de Matemática Aplicada

Condições sequenciais de otimalidade

Gabriel Haeser
Doutorado em Matemática Aplicada

Orientador: **José Mario Martínez**

Trabalho financiado pela FAPESP

Campinas
Setembro de 2009

Condições sequenciais de otimalidade

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por **Gabriel Haeser** e aprovada pela comissão julgadora.

Campinas, 04 de Setembro de 2009.

José Mario Martínez
Orientador

Banca examinadora:
José Mario Martínez (IMECC/UNICAMP)
Marcia Aparecida Gomes Ruggiero (IMECC/UNICAMP)
Paulo José da Silva e Silva (IME/USP)
Luis Mauricio Graña Drummond (UFRJ)
Susana Scheimberg de Makler (UFRJ)

Tese apresentada ao Instituto de Matemática Estatística e Computação Científica, UNICAMP, como requisito parcial para a obtenção do título de **Doutor em Matemática Aplicada**.

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP
Bibliotecária: Miriam Cristina Alves CRB8a / 8596

Haeser, Gabriel

H119c Condições sequenciais de otimalidade/Gabriel Haeser-- Campinas, [S.P. :s.n.], 2009.

Orientador : José Mario Martínez

Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1.Programação não-linear. 2.Condições de otimalidade. 3.Condições de qualificação. I. Martínez, José Mario. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica Computação. III. Título.

Título em inglês: Sequential Optimality conditions

Palavras-chave em inglês (Keywords): 1. Nonlinear programming. 2. Optimality conditions.
3. Constraint qualifications.

Área de concentração: Otimização

Titulação: Doutor em Matemática Aplicada

Banca examinadora: Prof. Dr. José Mario Martínez (IMECC-UNICAMP)
Profa. Dra. Marcia Aparecida Gomes Ruggiero (IMECC-UNICAMP)
Prof. Dr. Paulo José da Silva e Silva (IME-USP)
Prof. Dr. Luis Mauricio Graña Drummond (UFRJ)
Profa. Dra. Susana Scheimberg de Makler (UFRJ)

Data da defesa: 04/09/2009

Programa de pós-graduação: Doutorado em Matemática Aplicada

Tese de Doutorado defendida em 04 de setembro de 2009 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). JOSÉ MARIO MARTÍNEZ PÉREZ

Prof(a). Dr(a). MARCIA APARECIDA GOMES RUGGIERO

Prof(a). Dr(a). PAULO JOSÉ DA SILVA E SILVA

Prof(a). Dr(a). LUIS MAURICIO GRAÑA DRUMMOND

Prof(a). Dr(a). SUSANA SCHEIMBERG DE MAKLER

à Etiene.

Agradecimentos

À Etiene pelo incentivo e motivação.

Ao Eloi, Neuza, Lucas e Karina.

Ao Seu Roberto, por me ceder uma sala na Alvesmaq.

À Fapesp.

Ao Martínez, pelas grandes idéias e por sempre responder cuidadosamente meus insistentes e-mails (mesmo na França durante o natal).

Ao Nino e Andreas Fischer, pela colaboração.

Por fim, agradeço a mim mesmo por ter terminado este trabalho.

Resumo

Estudamos as condições de otimalidade provenientes dos algoritmos de penalidade externa, penalidade interna, penalidade interna-externa e restauração inexata, e mostramos relações com a CPLD, uma nova condição de qualificação estritamente mais fraca que a condição de Mangasarian-Fromovitz e a condição de posto constante de Janin. Estendemos o resultado do clássico Lema de Carathéodory, onde mostramos um limitante para o tamanho dos novos multiplicadores. Apresentamos novas condições de otimalidade relacionadas à condição AGP (Approximate Gradient Projection). Quando há um conjunto extra de restrições lineares, definimos uma condição do tipo AGP e provamos relações com a CPLD e as equações KKT. Resultados similares são obtidos quando há um conjunto extra de restrições convexas. Mostramos também algumas generalizações e relações com um algoritmo de restauração inexata.

Palavras chave: programação não-linear; condições de otimalidade; condições de qualificação.

Abstract

We study optimality conditions generated by the external penalty, internal penalty, internal-external penalty and inexact restoration algorithms, and we show relations with the CPLD, a new constraint qualification strictly weaker than the Mangasarian-Fromovitz condition and the constant rank condition of Janin. We extend the result of the classical Carathéodory's Lemma, where we show a bound for the size of the new multipliers. We present new optimality conditions related to the Approximate Gradient Projection condition (AGP). When there is an extra set of linear constraints, we define an AGP type condition and prove relations with CPLD and KKT conditions. Similar results are obtained when there is an extra set of convex constraints. We provide some further generalizations and relations to an inexact restoration algorithm.

Key words: nonlinear programming; optimality conditions; constraint qualifications.

Sumário

Introdução

Neste trabalho nos propomos a analisar as chamadas condições sequenciais de otimalidade, isto é, condições de otimalidades que estão intimamente relacionadas com algoritmos.

A condição de otimalidade oriunda do algoritmo de penalidade externa foi analisada em [59, 42], onde os autores mostram que esta condição implica KKT quando o problema satisfaz a condição de qualificação CPLD [51, 7], uma condição estritamente mais fraca que Mangasarian-Fromovitz [40] e Posto Constante [34].

Nos mesmos moldes, estudamos a condição de otimalidade oriunda do algoritmo de penalidade interna, aplicado a problemas apenas com desigualdades, e mostramos que esta condição implica KKT quando o problema satisfaz Mangasarian-Fromovitz. Mostramos que este resultado não pode ser enfraquecido, pois o algoritmo de penalidade interna só pode ser aplicado quando os pontos viáveis são arbitrariamente aproximados por pontos interiores com respeito às restrições, e esta condição, juntamente com CPLD é equivalente a Mangasarian-Fromovitz. Estudamos também um algoritmo misto de penalidade interna-externa para o problema com restrições de igualdade e desigualdade, e mostramos que, sob a CPLD, a condição de otimalidade obtida implica KKT, desde que os pontos viáveis possam ser arbitrariamente aproximados por pontos interiores viáveis.

Outro resultado obtido nesta tese se refere ao Lema de Carathéodory, a ferramenta fundamental para trabalharmos com a CPLD, onde mostramos um limitante para o tamanho dos novos multiplicadores. Este resultado se mostrou muito útil na análise do algoritmo de penalidade interna e em outras demonstrações.

Estudamos o algoritmo de restauração inexata [43, 41, 44, 25, 24, 17] e a condição de otimalidade oriunda deste algoritmo [45], denominada AGP. Em [45], os autores mostram que AGP implica KKT sob a condição de Mangasarian-Fromovitz, e em [24] este resultado é fortalecido utilizando a CPLD. Quando o problema contém um conjunto adicional de restrições convexas ou lineares, definimos respectivamente as condições de otimalidade C-AGP e L-AGP, e mostramos que L-AGP implica AGP, herdando portanto a propriedade de AGP que sob CPLD implica KKT. Mostramos que o mesmo não acontece com C-AGP, apesar disto, esta é uma condição de otimalidade forte, pois sob Mangasarian-Fromovitz implica KKT.

Introduzimos uma generalização das condições AGP, L-AGP e C-AGP, o que nos fornece uma nova classe de direções úteis para a utilização na fase de otimalidade do algoritmo de restauração inexata proposto por Fischer e Friedlander [17].

Esta tese está organizada da seguinte maneira: no Capítulo 1, apresentamos as definições básicas e as condições de qualificação clássicas. Mostramos também definições equivalentes e propriedades, além do novo resultado sobre o Lema de Carathéodory. No Capítulo 2 mostramos os resultados referentes aos algoritmos de penalidade externa, penalidade interna e o algoritmo misto de penalidade interna-externa. No Capítulo 3 apresentamos as condições de otimalidade do tipo AGP e generalizações. No Capítulo 4 mostramos alguns algoritmos estudados, e resultados obtidos com condições de qualificação fracas.

Os resultados referentes ao Lema de Carathéodory estendido e a impossibilidade de enfraquecer o resultado de convergência do método de penalidade interna estão submetidos no artigo *New bounds on the multipliers given by Carathéodory's theorem and a result on the internal penalty method*[1]. As condições de otimalidade apresentadas no Capítulo 2 foram unificadas, dando origem ao artigo *On sequential optimality conditions for smooth constrained optimization*[2], aceito para publicação na revista *Optimization*, onde também incluímos as condições C-AGP e L-AGP. O artigo referente a C-AGP-generalizada e o algoritmo de restauração inexata está em preparação.

[1]disponível em `http://www.optimization-online.org/DB_HTML/2009/06/2332.html`
[2]disponível em `http://www.optimization-online.org/DB_HTML/2009/06/2326.html`

Notações:

- $||\cdot|| = ||\cdot||_2$.
- $\mathbb{R}^p_+ = \{x \in \mathbb{R}^p | x_i \geq 0, \quad \forall i \in \{1, \ldots, p\}\}$.
- $V(x)$ denota uma vizinhança (aberta) de x.
- Se $h\colon \mathbb{R}^n \to \mathbb{R}^m$ é diferenciável, $\nabla h(x)$ denota a matriz $n \times m$ cuja i-ésima coluna é o gradiente da i-ésima coordenada de h, isto é, $\nabla h_i(x)$. Temos que $\nabla h(x) = h'(x)^{\mathtt{T}}$, isto é, o Jacobiano transposto de h.
- $|\mathtt{I}|$ denota a quantidade de elementos do conjunto finito $\mathtt{I}$.
- $e_i \in \mathbb{R}^n$ denota o vetor cuja i-ésima coordenada é 1 e as demais são 0.
- $K \underset{\infty}{\subset} \mathbb{N}$ significa que K é um subconjunto infinito de $\mathbb{N}$.

Capítulo 1

Condições de qualificação

Neste capítulo apresentamos definições básicas e as condições de qualificação: regularidade (LICQ), Mangasarian-Fromovitz (MF), posto constante (CRCQ) e dependência linear positiva constante (CPLD) como uma série de generalizações naturais. Em seguida mostramos algumas propriedades clássicas dessas condições, como o fato de que, sob MF, a não existência de direções de descida viáveis é equivalente à existência de multiplicadores de Lagrange. Mostramos também que ao reformular um problema incluindo variáveis de folga, condições como MF, CRCQ e CPLD não são perdidas.

Obtemos também um novo resultado sobre o Lema de Carathéodory, onde mostramos um limitante para os multiplicadores, além disso, apresentamos condições de qualificação mais fracas como a quase-normalidade (QN), Abadie e Guignard, sendo a última equivalente à existência de multiplicadores de Lagrange.

1.1 Definições

Em otimização contínua, estamos interessados em resolver o problema

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & x \in X, \end{array}$$

onde $f\colon \mathbb{R}^n \to \mathbb{R}$ é uma função com derivadas primeiras contínuas e $X \subset \mathbb{R}^n$, fechado, é o conjunto dos pontos viáveis, geralmente definido por restrições de igualdade e desigualdade, isto é, $X = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0\}$, onde $h : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^p$, também com derivadas primeiras contínuas. Um ponto qualquer $x \in X$ é dito um ponto viável. A relação de ordem parcial usada em $\mathbb{R}^p$ é definida por coordenadas, isto é, $u \leq v \Leftrightarrow u_i \leq v_i, \quad \forall i \in \{1, \dots, p\}$. Veja mais detalhes no Apêndice 4.2.

Resolver o problema significa encontrar um minimizador global de f restrito a X, isto é, encontrar um ponto viável com menor valor da função objetivo f, ou seja, $x^* \in X$ tal que $f(x^*) \leq f(x)$ para todo $x \in X$. Um minimizador global pode não existir, é o caso quando $X = \emptyset$, onde dizemos que o problema é inviável, ou quando existe uma sequência $\{x^k\} \subset X$ tal que $f(x^k) \to -\infty$, onde dizemos que o problema é ilimitado. Uma condição suficiente para a existência de um minimizador global é que X seja compacto e não vazio.

Na prática, encontrar minimizadores globais é muito difícil, então o que se faz é encontrar um ponto que satisfaça alguma propriedade que sabemos ser satisfeita pelos minimizadores globais, isso é o que chamamos de **condição de otimalidade**.

Definição 1.1 (Condição de Otimalidade)**.** *Uma propriedade P sobre pontos de $\mathbb{R}^n$ é dita uma condição de otimalidade se*

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

Uma condição de otimalidade útil na prática deve ser forte, no sentido de que existam poucos pontos que a satisfaçam mas que não sejam minimizadores globais. Um exemplo de condição de otimalidade muito fraca é a condição $x \in \mathbb{R}^n$.

Uma condição de otimalidade bastante forte é a condição de ser um minimizador local, isto é, um ponto $x^* \in X$ tal que $f(x^*) \leq f(x)$ para todo $x \in X \cap V$, para alguma vizinhança V de x^*. O problema é que esta condição é muito difícil de ser verificada. Naturalmente, condições de otimalidade úteis na prática devem ser de fácil verificação.

Outra qualidade desejada em uma condição de otimalidade é que seja possível construir algoritmos de modo a garantir que os pontos limites satisfaçam a condição.

Considere o caso em que $X = \mathbb{R}^n$, isto é, o problema é irrestrito. Neste caso, uma condição de otimalidade razoável é que $\nabla f(x) = 0$. Um ponto que satisfaz esta condição é chamado ponto estacionário.

Para o caso com restrições, dizemos que x é estacionário quando satisfaz a condição de Karush-Kuhn-Tucker (KKT)[35, 37].

Definição 1.2 (KKT)**.** *Dizemos que x satisfaz a condição KKT quando $x \in X = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0\}$ e existem $\lambda \in \mathbb{R}^m$, $\mu_j \geq 0 \quad \forall j \in \mathtt{I}(x)$ tais que*

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

*onde $\mathtt{I}(x) = \{j \in \{1, \ldots, p\} | g_j(x) = 0\}$ é o conjunto das restrições ativas em x. Os vetores λ e μ são denominados **multiplicadores de Lagrange**.*

Diferentemente do caso irrestrito, em geral, a condição KKT não é uma condição de otimalidade, basta considerar o problema

$$\begin{array}{ll} \text{Minimizar} & x \\ \text{Sujeito a} & x^2 = 0. \end{array}$$

O minimizador global $x^* = 0$ não é KKT pois não existe $\lambda \in \mathbb{R}$ tal que $1 + \lambda 0 = 0$.

Para que a condição KKT seja uma condição de otimalidade é preciso assumir alguma hipótese adicional às restrições do problema, qualquer hipótese com essa propriedade é chamada uma condição de qualificação.

Definição 1.3 (Condição de Qualificação)**.** *Dizemos que uma propriedade CQ sobre pontos de $\mathbb{R}^n$ é uma condição de qualificação se*

x é um minimizador global e x satisfaz CQ $\Rightarrow$ x satisfaz a condição KKT.

Note que toda condição de qualificação CQ define uma condição de otimalidade dada por KKT ou não-CQ, já que $p \wedge q \Rightarrow r$ é logicamente equivalente a $p \Rightarrow r \vee \neg q$.

Observe que quanto mais fraca for a condição de qualificação, mais forte será a condição de otimalidade associada, deste modo, estamos interessados em condições de qualificação fracas.

A condição de qualificação mais conhecida é a condição de independência linear dos gradientes das restrições ativas (LICQ, do inglês *Linear Independence Constraint Qualification*), também chamada de regularidade:

Definição 1.4 (LICQ). *Dizemos que um ponto $x \in X$ satisfaz LICQ se $\{\nabla h_i(x)\}_{i=1}^m \bigcup \{\nabla g_j(x)\}_{j \in \mathtt{I}(x)}$ é linearmente independente.*

Observação: A rigor, nesta definição (e nas seguintes) estamos utilizando o conceito de multi-conjunto, isto é, um conjunto que admite elementos repetidos. Os multi-conjuntos podem ser definidos pela teoria usual dos conjuntos, basta indicar a quantidade de elementos repetidos em forma de par ordenado, por exemplo, o multi-conjunto de três elementos $\{a, a, b\}$ é definido como um conjunto com dois pares ordenados $\{(a, 2), (b, 1)\}$.

Com o objetivo de motivar a introdução das demais condições de qualificação, temos a proposição abaixo.

Proposição 1.5. *$x \in X$ satisfaz LICQ se, e somente se, $\forall \mathtt{I} \subset \{1, \ldots, m\}, \forall \mathtt{J} \subset \mathtt{I}(x)$ temos $\{\nabla h_i(y)\}_{i \in \mathtt{I}} \bigcup \{\nabla g_j(y)\}_{j \in \mathtt{J}}$ é linearmente independente $\forall y \in V(x)$, para alguma vizinhança $V(x)$ de x.*

Demonstração: Basta observar que todo subconjunto de um conjunto linearmente independente é também linearmente independente, além disso, se temos t funções contínuas $v_i : \mathbb{R}^n \to \mathbb{R}^q$ de modo que $\{v_1(x), \ldots, v_t(x)\}$ é linearmente independente então existe uma vizinhança $V(x)$ de x tal que $\{v_1(y), \ldots, v_t(y)\}$ é linearmente independente para todo $y \in V(x)$. Com efeito, suponha que exista uma sequência $y^k \to x$ tal que $\{v_1(y^k), \ldots, v_t(y^k)\}$ é linearmente dependente para todo k, logo existem vetores $\alpha^k = (\alpha_1^k, \ldots, \alpha_t^k)$ não nulos de modo que

$$\sum_{i=1}^{t} \alpha_i^k v_i(y^k) = 0. \tag{1.1}$$

Defina $M_k = \max\{|\alpha_i^k|, i = 1, \ldots, t\} \neq 0$ para cada k. Como $\left\{\dfrac{\alpha^k}{M_k}\right\}$ é limitada, podemos tomar uma subsequência de modo que $\dfrac{\alpha^k}{M_k} \to \alpha$, assim, dividindo (1.1) por M_k e tomando o limite para esta subsequência, obtemos

$$\sum_{i=1}^{t} \alpha_i v_i(x) = 0,$$

mas como $||\frac{\alpha^k}{M_k}||_\infty = 1$ e a norma é contínua, temos que $||\alpha||_\infty = 1$, logo $\{v_1(x), \ldots, v_t(x)\}$ é linearmente dependente, o que é uma contradição. $\square$

Os minimizadores que satisfazem LICQ possuem uma propriedade notável: os multiplicadores de Lagrange associados são únicos. Isso é consequência imediata do fato que a matriz do sistema linear das equações KKT tem posto completo.

Ao observarmos a demonstração via penalidade externa de que a LICQ é de fato uma condição de qualificação (mostraremos detalhes desta técnica na Seção 1.4), fica claro que é possível enfraquecer a hipótese. A idéia é impedir a existência de uma sequência de gradientes linearmente independentes, que seja linearmente dependente no limite. Assim, ao invés de pedir independência linear dos subconjuntos de gradientes na vizinhança de x, vamos impor que a dependência linear dos gradientes em x implique na dependência linear dos gradientes numa vizinhança de x.

Definição 1.6 (Posto Constante (CRCQ))**.** *Dizemos que $x \in X$ satisfaz a condição de qualificação de posto constante se* $\forall \mathtt{I} \subset \{1, \ldots, m\}, \forall \mathtt{J} \subset \mathtt{I}(x)$ *temos*

$$\{\nabla h_i(x)\}_{i\in\mathtt{I}} \bigcup \{\nabla g_j(x)\}_{j\in\mathtt{J}} \text{ é LD } \Rightarrow \{\nabla h_i(y)\}_{i\in\mathtt{I}} \bigcup \{\nabla g_j(y)\}_{j\in\mathtt{J}} \text{ é LD } \quad \forall y \in V(x),$$

para alguma vizinhança $V(x)$ de x.

A condição de posto constante CRCQ (*Constant Rank Constraint Qualification*) foi introduzida por Janin em [34]. Recentemente mostrou-se que a CRCQ é também uma condição de qualificação de segunda ordem, isto é, os multiplicadores de Lagrange associados a um minimizador local que satisfaz CRCQ são tais que a hessiana do lagrangiano é semi definida positiva em um certo subespaço tangente às restrições ativas [5].

Observe que todo problema apenas com restrições lineares satisfaz CRCQ, logo, em todo problema com somente restrições lineares, as equações KKT são satisfeitas pelos minimizadores.

Antes de introduzirmos a próxima condição de qualificação, precisamos de uma definição.

Definição 1.7 (Dependência linear positiva). *Dados $u_1, \dots, u_{p+m}$, vetores não necessariamente distintos em $\mathbb{R}^n$, dizemos que $(\{u_1, \dots, u_p\}, \{u_{p+1}, \dots, u_{p+m}\})$ é positivo linearmente dependente (PLD) se existem números reais $\alpha_1, \dots, \alpha_{p+m}$ não todos nulos, com $\alpha_{p+1} \geq 0, \dots, \alpha_{p+m} \geq 0$ tais que $\sum_{i=1}^{p+m} \alpha_i u_i = 0$. Caso contrário dizemos que $(\{u_1, \dots, u_p\}, \{u_{p+1}, \dots, u_{p+m}\})$ é positivo linearmente independente (PLI).*

Observação: Quando não houver perigo de confusão, eventualmente diremos $u_1, .., u_p, u_{p+1}, .., u_{p+m}$ ou $\{u_1, .., u_p\} \cup \{u_{p+1}, .., u_{p+m}\}$ é PLD (ou PLI). Observe que os conjuntos na definição são na verdade multi-conjuntos para permitirmos elementos repetidos, além disso, a notação de par ordenado é necessária para a identificação dos vetores associados aos escalares não negativos.

Utilizando o fato que nas equações KKT os multiplicadores associados às restrições de desigualdade ativas são não negativos, podemos enfraquecer a hipótese LICQ em outro sentido. Vamos impor que o gradiente das restrições de igualdade e desigualdade ativas sejam PLI. Esta é a condição de qualificação de Mangasarian-Fromovitz, introduzida em [40].

Definição 1.8 (MF). *Dizemos que $x \in X$ satisfaz a condição de qualificação de Mangasarian-Fromovitz (MF) se*

$$\left(\{\nabla h_i(x)\}_{i=1}^{m}, \{\nabla g_j(x)\}_{j \in \mathrm{I}(x)} \right) \text{ é PLI.}$$

Note que se os gradientes de duas restrições de desigualdade são iguais e não nulos em um certo ponto, a condição MF vale, enquanto LICQ não vale.

Como LI implica PLI, temos que a condição MF é mais fraca que LICQ. No caso de um minimizador que satisfaz MF, podemos garantir que os multiplicadores de Lagrange estão em um compacto. De fato, Gauvin mostrou que as duas coisas são equivalentes [22].

Analogamente à Proposição 1.5, temos que MF é equivalente ao fato que todo subconjunto dos gradientes das restrições ativas é PLI em uma vizinhança.

Proposição 1.9. *Se $x \in X$ satisfaz MF se, e somente se, $\forall \mathtt{I} \subset \{1, \ldots, m\}, \forall \mathtt{J} \subset \mathtt{I}(x)$ temos $\left(\{\nabla h_i(y)\}_{i \in \mathtt{I}}, \{\nabla g_j(y)\}_{j \in \mathtt{J}}\right)$ é PLI $\forall y \in V(x)$, para alguma vizinhança $V(x)$ de x.*

Demonstração: É análoga à demonstração da Proposição 1.5. Basta observar que quando $\frac{\alpha_i^k}{M_k} \to \alpha_i$, como $M_k > 0$ temos que α_i e α_i^k possuem o mesmo sinal para k suficientemente grande (isto é, $\alpha_i \alpha_i^k \geq 0$). $\square$

A condição MF é muito usada na prática, porém, existem problemas simples onde ela não é satisfeita. Por exemplo, quando temos uma restrição de igualdade $h(x) = 0$ e a transformamos em duas restrições de desigualdade $h(x) \leq 0$ e $-h(x) \leq 0$, a condição MF não vale, já que a soma dos gradientes é zero.

Com base nas três condições de qualificação apresentadas, é natural introduzirmos uma condição de qualificação análoga à CRCQ, mas impondo que se os subconjuntos dos gradientes são PLD, então devem ser PLD em uma vizinhança.

Definição 1.10 (CPLD)**.** *Dizemos que $x \in X$ satisfaz a CPLD se $\forall \mathtt{I} \subset \{1, \ldots, m\}, \forall \mathtt{J} \subset \mathtt{I}(x)$ temos que*

$$\left(\{\nabla h_i(x)\}_{i \in \mathtt{I}}, \{\nabla g_j(x)\}_{j \in \mathtt{J}}\right) \text{ é PLD} \Rightarrow \left(\{\nabla h_i(y)\}_{i \in \mathtt{I}}, \{\nabla g_j(y)\}_{j \in \mathtt{J}}\right) \text{ é PLD} \quad \forall y \in V(x),$$

para alguma vizinhança $V(x)$ de x.

A condição de dependência linear positiva constante (CPLD) foi introduzida em [51] por Qi e Wei e a demonstração de que ela é de fato uma condição de qualificação foi feita por Andreani, Martínez e Schuverdt em [7, 59]. Na verdade os autores mostraram que a CPLD implica a quase-normalidade, cuja definição se encontra na Seção 1.4.

A definição da CPLD, dada por por Qi e Wei é ligeiramente diferente da enunciada aqui. A definição original pedia a condição aparentemente mais fraca de que os subconjuntos de gradientes das restrições ativas que são PLD sejam LD em uma vizinhança. Em [59] foi mostrado que as duas definições são equivalentes.

A CPLD é uma condição de qualificação mais fraca que MF e CRCQ, de modo que a condição de otimalidade gerada por ela é mais forte. Além disso, a CPLD se mostrou uma condição de qualificação útil na prática, pois recentemente foi possível demonstrar a convergência de um algoritmo do tipo Lagrangiano Aumentado sob esta condição [59, 3, 4, 42].

Apesar da natureza genérica das condições de otimalidade [62], o estudo de condições de otimalidade mais fortes (ou condições de qualificação mais fracas) se justifica pois existem classes inteiras de problemas que não satisfazem as condições de qualificação usuais como LICQ e MF, é o caso de MPECs (*mathematical programs with equilibrium constraints*) e MPVCs (*mathematical programs with vanishing constraints*) [2, 29, 30, 31, 32].

1.2 Propriedades

Nesta seção mostraremos algumas propriedades satisfeitas pelas condições de qualificação apresentadas, assim como possíveis definições equivalentes.

Proposição 1.11. *$x \in X$ satisfaz MF se, e somente se existe $d \in \mathbb{R}^n$ tal que*

i) $\{\nabla h_i(x)\}_{i=1}^m$ *é LI*

ii) $\nabla h_i(x)^{\mathtt{T}} d = 0 \quad \forall i = 1, \ldots, m$

iii) $\nabla g_j(x)^{\mathtt{T}} d < 0 \quad \forall j \in \mathtt{I}(x)$

Demonstração: ($\Leftarrow$) Seja $d \in \mathbb{R}^n$ como no enunciado e suponha que $x \in X$ não satisfaz MF, logo existe $\lambda \in \mathbb{R}^m$ e $\mu_j \geq 0, \quad \forall j \in \mathtt{I}(x)$, não todos nulos tais que

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

Tomando o produto interno com d obtemos

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

De ii), iii) e do fato que $\mu_j \geq 0$ temos que $\mu_j \nabla g_j(x)^{\mathsf{T}} d = 0 \quad \forall j \in \mathtt{I}(x)$, logo $\mu_j = 0 \quad \forall j \in \mathtt{I}(x)$, e então temos que $\sum_{i=1}^{m} \lambda_i \nabla h_i(x) = 0$ com $\lambda \neq 0$, o que contradiz o fato de $\{\nabla h_i(x)\}_{i=1}^{m}$ ser LI.

($\Rightarrow$) Observe que a primeira condição é claramente satisfeita. Considere o problema linear

$$\begin{array}{ll} \text{Minimizar}_{(d,z)} & z \\ \text{Sujeito a} & \nabla h_i(x)^{\mathsf{T}} d = 0 \quad \forall i = 1, \ldots, m \\ & \nabla g_j(x)^{\mathsf{T}} d \leq z \quad \forall j \in \mathtt{I}(x). \end{array}$$

Note que a região viável é não vazia pois $(0,0)$ é viável. Suponha que exista uma solução ótima (d^*, z^*). Pela linearidade das restrições, temos que vale a condição KKT, isto é, existem $\lambda \in \mathbb{R}^m$ e $\mu_j \geq 0, \quad \forall j \in \mathtt{I}(x)$, tais que

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

com $\mu_j(\nabla g_j(x)^{\mathsf{T}} d^* - z^*) = 0, \forall j \in \mathtt{I}(x)$. Logo temos que $\sum_{j \in \mathtt{I}(x)} \mu_j = 1$ (o que implica em $\mu \neq 0$) e $\displaystyle\sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{j \in \mathtt{I}(x)} \mu_j \nabla g_j(x) = 0$, o que contradiz a hipótese de que x satisfaz MF. Assim, o problema linear é ilimitado e podemos garantir a existência da direção d do enunciado. $\square$

Na verdade, esta é a forma com que a condição MF foi originalmente enunciada. A formulação equivalente que enunciamos é devido a Rockafellar [58].

A seguir, damos uma prova simples da equivalência proposta por Robinson em [53]. Observe que int$C$ é o interior topológico de C, isto é, $x \in \text{int}C \Leftrightarrow$ existe uma vizinhança $V(x)$ de x inteiramente contida em C. É importante salientar que esta equivalência está feita em um contexto mais geral em [53].

Proposição 1.12. *Seja $x \in X$, então x satisfaz MF se, e somente se* $0 \in \text{int}\{(y,z) \in \mathbb{R}^p \times \mathbb{R}^m | y \geq g(x) + \nabla g(x)^\mathtt{T} d, z = h(x) + \nabla h(x)^\mathtt{T} d$ *para algum* $d \in \mathbb{R}^n\}$.

Demonstração: ($\Rightarrow$) Seja $x \in X$. Fixado $z \in \mathbb{R}^m$, o sistema linear $\nabla h(x)^\mathtt{T} d = z$ admite solução $\bar{d} \in \mathbb{R}^n$, já que suas linhas são LI pela Proposição 1.11. Seja $d' \in \mathbb{R}^n$ dado por esta Proposição e defina $d = \alpha\bar{d} + \beta d'$. Como $h(x) = 0$ e $\nabla h(x)^\mathtt{T} d' = 0$, temos $\alpha z = h(x) + \nabla h(x)^\mathtt{T} d$ para todo $z \in \mathbb{R}^m, \alpha > 0$ e $\beta > 0$.

Por outro lado, $g_i(x) + \nabla g_i(x)^\mathtt{T} d = g_i(x) + \alpha \nabla g_i(x)^\mathtt{T} \bar{d} + \beta \nabla g_i(x)^\mathtt{T} d'$. Caso $i \in \mathtt{I}$, então $g_i(x) = 0$ e $\nabla g_i(x)^\mathtt{T} d' < 0$, logo podemos tomar $\alpha > 0$ suficientemente pequeno tal que $g_i(x) + \nabla g_i(x)^\mathtt{T} d < 0$. Caso $i \notin \mathtt{I}$, então $g_i(x) < 0$ e podemos tomar $\alpha > 0$ e $\beta > 0$ suficientemente pequenos tais que $g_i(x) + \nabla g_i(x)^\mathtt{T} d < 0$. Logo se $||y||$ é suficientemente pequena então $g(x) + \nabla g(x)^\mathtt{T} d \leq y$.

($\Leftarrow$) Seja $x \in X$ e $\varepsilon > 0$ tal que $\forall (y,z) \in \mathbb{R}^p \times \mathbb{R}^m$ com $||(y,z)||_\infty < \varepsilon$, existe $d \in \mathbb{R}^n$ tal que $y \geq g(x) + \nabla g(x)^\mathtt{T} d, z = h(x) + \nabla h(x)^\mathtt{T} d$. Para cada $z' \in \mathbb{R}^m, z' \neq 0$, tome $z = \varepsilon \frac{z'}{||z'||_\infty}$, assim $||z||_\infty = \varepsilon$ e existe d tal que $z = h(x) + \nabla h(x)^\mathtt{T} d$. Tomando $d' = \frac{||z'||_\infty}{\varepsilon} d$, como $h(x) = 0$ temos que $\nabla h(x)^\mathtt{T} d' = z'$, logo o posto de $\nabla h(x) \in \mathbb{R}^{n \times m}$ é m, portanto suas colunas são LI, ou seja, $\{\nabla h_i(x)\}_{i=1}^m$ é LI.

Se $i \in \mathtt{I}(x)$, defina $y_i = -\varepsilon$, caso contrário defina $y_i = 0$. Defina também $z = 0$. Assim, $||(y,z)||_\infty \leq \varepsilon$ e existe $d \in \mathbb{R}^n$ com $y \geq g(x) + \nabla g(x)^\mathtt{T} d, z = h(x) + \nabla h(x)^\mathtt{T} d$. Como $h(x) = 0$ e $z = 0$, temos $\nabla h(x)^\mathtt{T} d = 0$. Para $i \in \mathtt{I}(x)$, temos $0 > -\varepsilon \geq g_i(x) + \nabla g_i(x)^\mathtt{T} d = \nabla g_i(x)^\mathtt{T} d$. Logo x satisfaz MF. □

A proposição a seguir mostra que se vale a condição MF em x, então x é KKT se, e somente se não existem direções de descida estritamente viáveis a partir de x. Na verdade, o resultado é mais forte:

Proposição 1.13. *Se $x \in X$ é KKT então não existe $d \in \mathbb{R}^n$ direção de descida viável, isto é, tal que* $\nabla f(x)^\mathtt{T} d < 0, \nabla h(x)^\mathtt{T} d = 0, \nabla g_j(x)^\mathtt{T} d \leq 0 \quad \forall j \in \mathtt{I}(x)$.

Demonstração: Temos que existem $\lambda \in \mathbb{R}^m, \mu_j \geq 0 \quad \forall j \in \mathtt{I}(x)$ tais que $\nabla f(x) + \nabla h(x)\lambda + \sum_{j \in \mathtt{I}(x)} \mu_j \nabla g_j(x) = 0$. Se existe uma direção de descida viável d, tomando o produto interno com as equações KKT temos

$$\nabla f(x)^{\mathtt{T}} d + \lambda^{\mathtt{T}} \nabla h(x)^{\mathtt{T}} d + \sum_{j \in \mathtt{I}(x)} \mu_j \nabla g_j(x)^{\mathtt{T}} d = 0,$$

o que é uma contradição pois o lado esquerdo é menor que zero. □

Proposição 1.14. *Se $x \in X$ satisfaz MF e não existe $d \in \mathbb{R}^n$ direção de descida estritamente viável, isto é, tal que $\nabla f(x)^{\mathtt{T}} d < 0, \nabla h(x)^{\mathtt{T}} d = 0, \nabla g_j(x)^{\mathtt{T}} d < 0 \quad \forall j \in \mathtt{I}(x)$, então x é KKT.*

Demonstração: Considere o problema

$$\begin{array}{ll} \text{Minimizar}_{(d,z)} & z \\ \text{Sujeito a} & \nabla h(x)^{\mathtt{T}} d = 0 \\ & \nabla g_j(x)^{\mathtt{T}} d \leq z, \forall j \in \mathtt{I}(x) \\ & \nabla f(x)^{\mathtt{T}} d \leq z. \end{array}$$

Temos que $(0,0)$ é viável. Suponha que o problema é ilimitado, em particular, para $z < 0$, o fato de que existe d tal que (d,z) é viável fornece uma direção de descida estritamente viável para o problema original, o que é uma contradição. Logo o problema admite uma solução (d^*, z^*). Pela linearidade das restrições, (d^*, z^*) é KKT, logo existem $\lambda \in \mathbb{R}^m, \rho \geq 0$, $\mu_j \geq 0 \quad \forall j \in \mathtt{I}(x)$ tais que

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

Se $\rho = 0$, então $\sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in \mathtt{I}(x)} \mu_j \nabla g_j(x) = 0$ com $\sum_{j \in \mathtt{I}(x)} \mu_j = 1$, contradizendo o fato que x satisfaz MF. Portanto $\rho > 0$. Dividindo as equações KKT por ρ temos $\nabla f(x) + \sum_{i=1}^m \frac{\lambda_i}{\rho} \nabla h_i(x) + \sum_{j \in \mathtt{I}(x)} \frac{\mu_j}{\rho} \nabla g_j(x) = 0, \frac{\mu_j}{\rho} \geq 0$, ou seja, x é KKT do problema original. □

Quando o conjunto viável é convexo, existe uma condição simples que garante a existência de multiplicadores de Lagrange. É a chamada condição de qualificação de Slater [61]. Mostraremos que a condição de Slater implica a condição MF para um subproblema apropriado, de forma que a existência de multiplicadores para o subproblema garanta a existência para o problema original. O desenvolvimento a seguir se baseia nos resultados em [9]. Antes de introduzirmos a condição de Slater, mostremos o lema abaixo:

Lema 1.15. *Se x^* é solução de*

$$\begin{array}{ll} \textit{Minimizar} & f(x) \\ \textit{Sujeito a} & Ax = b \\ & x \in \Omega \end{array}$$

com $A^{\mathtt{T}} = [a_1, \ldots, a_m]$ então x^* é solução de

$$\begin{array}{ll} \textit{Minimizar} & f(x) \\ \textit{Sujeito a} & a_i^{\mathtt{T}} x = b_i, \forall i \in \mathtt{I} \\ & x \in \Omega, \end{array}$$

para todo $\mathtt{I} \subset \{1, \ldots, m\}$, tal que $\{a_i\}_{i \in \mathtt{I}}$ é LI com $|\mathtt{I}| = posto(A)$.

Demonstração: Seja $\mathtt{I}$ como no enunciado. Se $\mathtt{I} = \{1, \ldots, m\}$, o resultado claramente vale. Senão, seja $j \notin \mathtt{I}$, logo existem $\lambda_i \in \mathbb{R}, \forall i \in \mathtt{I}$ tais que

$$a_j = \sum_{i \in \mathtt{I}} \lambda_i a_i, \tag{1.2}$$

logo, tomando o produto interno com x^* e usando o fato que x^* é viável, temos

$$b_j = \sum_{i \in \mathtt{I}} \lambda_i b_i. \tag{1.3}$$

Considere $x \in \mathbb{R}^n$ tal que x satisfaz todas as restrições cujos índices estão em $\mathtt{I}$, isto é, $a_i^{\mathtt{T}} x = b_i, \forall i \in \mathtt{I}$.

Multiplicando a equação (1.2) por x, temos

$$a_j^{\mathtt{T}}x = \sum_{i\in\mathtt{I}} \lambda_i a_i^{\mathtt{T}}x.$$

Usando a equação (1.3) e o fato que x satisfaz as restrições temos

$$\begin{aligned} a_j^{\mathtt{T}}x &= \sum_{i\in\mathtt{I}} \lambda_i a_i^{\mathtt{T}}x \\ &= \sum_{i\in\mathtt{I}} \lambda_i b_i \\ &= b_j. \end{aligned}$$

Logo, todas as restrições cujos índices não estão em $\mathtt{I}$ são automaticamente satisfeitas e podemos removê-las da definição do problema. $\square$

Definição 1.16. *Dizemos que $x \in X$ satisfaz a condição de qualifição de Slater quando as restrições de igualdade são afins, as restrições de desigualdade são convexas e existe $x_0 \in X$ tal que $g_j(x_0) < 0$ para todo $j \in \mathtt{I}(x)$.*

Proposição 1.17. *A condição de Slater é uma condição de qualificação.*

Demonstração: Se $x^* \in X$ é um minimizador de $f(x)$ sujeito a $x \in X$, com $X = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0\}$ com h_i afim e g_j convexa, pelo Lema 1.15 e ignorando as restrições de desigualdade inativas, temos que x^* é solução de

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & h_i(x) = 0, \forall i \in \mathtt{I} \\ & g_j(x) \leq 0, \forall j \in \mathtt{I}(x^*), \end{array} \tag{1.4}$$

onde $\mathtt{I} \subset \{1, \ldots, m\}$ é tal que $\{\nabla h_i(x^*)\}_{i\in\mathtt{I}}$ é linearmente independente com $|\mathtt{I}| =$ posto$(\{\nabla h_i(x^*)\}_{i=1}^m)$. Pela convexidade de g_j, para $j \in \mathtt{I}(x)$ temos

$$0 > g_j(x_0) \geq g_j(x) + \nabla g_j(x)^{\mathtt{T}}(x_0 - x) = \nabla g_j(x)^{\mathtt{T}}(x_0 - x).$$

Tomando $d = x_0 - x$ temos $\nabla g_j(x)^{\mathtt{T}}d < 0, \forall j \in \mathtt{I}(x)$. Como h_i é afim e $x_0, x \in X$, temos $h_i(x_0) = h_i(x) + \nabla h_i(x)^{\mathtt{T}}(x_0 - x)$, logo $\nabla h_i(x)^{\mathtt{T}}d = 0, \forall i \in \mathtt{I}$. Assim x^* satisfaz MF para o problema (1.4), portanto x^* é KKT. Tomando $\lambda_i = 0$ se $i \notin \mathtt{I}$ temos que x^* é KKT para o problema original. $\square$

Observação: Quando as funções g_j não estão definidas em todo $\mathbb{R}^n$, mas sim, digamos em $\mathcal{D} \subset \mathbb{R}^n$, a condição de Slater afirma que existe $x_0 \in \text{relint}\mathcal{D}$ com $g_j(x_0) < 0, \forall j \in \mathtt{I}(x)$, onde relint$\mathcal{D}$ é o interior de $\mathcal{D}$ com respeito à topologia induzida no fecho afim de $\mathcal{D}$ pela topologia usual de $\mathbb{R}^n$. Mais especificamente temos o fecho afim $\text{aff}\mathcal{D} = \{\theta_1 x_1 + \cdots + \theta_r x_r | x_i \in \mathcal{D}, \theta_1 + \cdots + \theta_r = 1\}$ com a topologia dada por $\{B \cap \text{aff}\mathcal{D} |\, B \text{ é aberto de } \mathbb{R}^n\}$. Nesta topologia, o interior de $\mathcal{D}$ é dado por $\text{relint}\mathcal{D} = \{y \in \mathcal{D} |\, V(y) \cap \text{aff}\mathcal{D} \subset \mathcal{D}$, para alguma vizinhança $V(y)$ de $y\}$. Veja [12] para um tratamento de otimização convexa neste sentido.

A seguir mostramos uma definição equivalente para a condição CRCQ, a condição de posto constante de Janin. Esta equivalência justifica o nome da condição, além disso, esta é a maneira como a condição foi enunciada por Janin [34].

Proposição 1.18. *x satisfaz CRCQ se, e somente se $\forall \mathtt{I} \subset \{1, \ldots, m\}, \mathtt{J} \subset \mathtt{I}(x)$, $\{\nabla h_i(y)\}_{i \in \mathtt{I}} \bigcup \{\nabla g_j(y)\}_{j \in \mathtt{J}}$ tem posto constante para todo $y \in V(x)$, para alguma vizinhança $V(x)$ de x.*

Demonstração: ($\Leftarrow$) Seja $\mathtt{I} \subset \{1, \ldots, m\}, \mathtt{J} \subset \mathtt{I}(x)$ e suponha $\{\nabla h_i(x)\}_{i \in \mathtt{I}} \bigcup \{\nabla g_j(x)\}_{j \in \mathtt{J}}$ é LD, portanto existe uma vizinhança $V(x)$ de x tal que o posto dos gradientes é constante nesta vizinhança, logo $\{\nabla h_i(y)\}_{i \in \mathtt{I}} \bigcup \{\nabla g_j(y)\}_{j \in \mathtt{J}}$ é LD, para todo $y \in V(x)$.

($\Rightarrow$) Seja $\mathtt{I} \subset \{1, \ldots, m\}, \mathtt{J} \subset \mathtt{I}(x)$ e suponha que o posto de $\{\nabla h_i(x)\}_{i \in \mathtt{I}} \bigcup \{\nabla g_j(x)\}_{j \in \mathtt{J}}$ é r. Sejam $\mathtt{I}_0 \subset \mathtt{I}$ e $\mathtt{J}_0 \subset \mathtt{J}$ com $|I_0| + |J_0| = r$ tal que $\{\nabla h_i(x)\}_{i \in \mathtt{I}_0} \bigcup \{\nabla g_j(x)\}_{j \in \mathtt{J}_0}$ seja LI. Pela demonstração da Proposição 1.5, existe uma vizinhança $V(x)$ de x tal que $\{\nabla h_i(y)\}_{i \in \mathtt{I}_0} \bigcup \{\nabla g_j(y)\}_{j \in \mathtt{J}_0}$ é LI. Então $\{\nabla h_i(y)\}_{i \in \mathtt{I}} \bigcup \{\nabla g_j(y)\}_{j \in \mathtt{J}}$ tem posto maior ou igual a r para todo $y \in V(x)$.

Suponha que exista uma sequência $y^k \to x$ tal que o posto de $\{\nabla h_i(y^k)\}_{i \in \mathtt{I}} \bigcup \{\nabla g_j(y^k)\}_{j \in \mathtt{J}}$ é maior que r. Logo para cada k, existe $\mathtt{I}_k \subset \mathtt{I}$ e $\mathtt{J}_k \subset \mathtt{J}$ com $|I_k| + |J_k| > r$ tal que $\{\nabla h_i(y^k)\}_{i \in \mathtt{I}_k} \bigcup \{\nabla g_j(y^k)\}_{j \in \mathtt{J}_k}$ é LI. Podemos tomar uma subsequência tal que os conjuntos são os mesmos, isto é, $I' = I_k, J' = J_k$ para todo k, com $|I'| + |J'| > r$. Assim, $\{\nabla h_i(x)\}_{i\in \mathtt{I}'} \bigcup \{\nabla g_j(x)\}_{j\in \mathtt{J}'}$ é LD (já que a quantidade de vetores é maior que o posto), com $\{\nabla h_i(y^k)\}_{i\in \mathtt{I}'} \bigcup \{\nabla g_j(y^k)\}_{j\in \mathtt{J}'}$ LI, o que contradiz a definição de CRCQ. $\square$

O resultado acima aparece, por exemplo, em [51].

A seguir, mostramos que MF e CPLD não são perdidas quando reformulamos o problema incluindo variáveis de folga.

Proposição 1.19. *O ponto x^* é viável e satisfaz MF para*

$$\begin{array}{ll} \textit{Minimizar} & f(x) \\ \textit{Sujeito a} & h(x) = 0 \\ & g(x) \leq 0 \end{array}$$

se, e somente se (x^*, s^*) é viável e satisfaz MF para

$$\begin{array}{ll} \textit{Minimizar} & f(x) \\ \textit{Sujeito a} & h(x) = 0 \\ & g(x) + s = 0 \\ & -s \leq 0. \end{array}$$

Demonstração: ($\Leftarrow$) Seja $d = (d', d'') \subset \mathbb{R}^n \times \mathbb{R}^p$ tal que

$$\left\{ \begin{pmatrix} \nabla h_1(x^*) \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} \nabla h_m(x^*) \\ 0 \end{pmatrix}, \begin{pmatrix} \nabla g_1(x^*) \\ e_1 \end{pmatrix}, \ldots, \begin{pmatrix} \nabla g_p(x^*) \\ e_p \end{pmatrix} \right\} \text{ é LI,} \quad (1.5)$$

onde $e_i \in \mathbb{R}^n$ é um vetor com 1 na i-ésima coordenada e 0 nas demais. Logo $\{\nabla h_1(x^*), \ldots, \nabla h_m(x^*)\}$ é LI. Além disso,

$$\begin{pmatrix} \nabla h_i(x^*) \\ 0 \end{pmatrix}^{\mathtt{T}} \begin{pmatrix} d' \\ d'' \end{pmatrix} = 0, \forall i = 1, \ldots, m,$$

portanto $\nabla h_i(x^*)^{\mathtt{T}} d' = 0, \forall i = 1, \ldots, m$. Temos também

$$\begin{pmatrix} \nabla g_j(x^*) \\ e_i \end{pmatrix}^{\mathtt{T}} \begin{pmatrix} d' \\ d'' \end{pmatrix} = 0, \forall j = 1, \ldots, p,$$

portanto $\nabla g_j(x^*)^{\mathtt{T}} d' + d''_j = 0, \forall j = 1, \ldots, p$. Analisando as restrições de desigualdade $-s \leq 0$, temos que se $s^*_j = 0$, ou seja, $g_j(x^*) = 0$, vale

$$\begin{pmatrix} 0 \\ -e_j \end{pmatrix}^{\mathtt{T}} \begin{pmatrix} d' \\ d'' \end{pmatrix} < 0,$$

e portanto $d''_j > 0$. Assim, $\nabla g_j(x^*)^{\mathtt{T}} d' < 0, \forall i \in \mathtt{I}(x^*)$.

($\Rightarrow$) Considere a direção $d' \in \mathbb{R}^n$ dada pela Proposição 1.11. Defina $d''_j = -\nabla g_j(x^*)^{\mathtt{T}} d'$ se $j \in \mathtt{I}(x^*)$ e $d''_j = 0$ caso contrário. É fácil verificar que $d = \begin{pmatrix} d' \\ d'' \end{pmatrix}$ satisfaz as propriedades da Proposição 1.11 para o problema com variáveis de folga. $\square$

Proposição 1.20. *O ponto x^* é viável e satisfaz CPLD para*

$$\begin{array}{ll} \textit{Minimizar} & f(x) \\ \textit{Sujeito a} & h(x) = 0 \\ & g(x) \leq 0 \end{array}$$

se, e somente se (x^*, s^*) é viável e satisfaz CPLD para

$$\begin{array}{ll} \textit{Minimizar} & f(x) \\ \textit{Sujeito a} & h(x) = 0 \\ & g(x) + s = 0 \\ & -s \leq 0. \end{array}$$

Demonstração: ($\Rightarrow$) Seja $\mathtt{I} \subset \{1, \ldots, m\}, \mathtt{J} \subset \{1, \ldots, p\}, \mathtt{K} \subset \{i \in \{1, \ldots, p\} | s^*_i = 0\}$ e suponha que existam $\lambda \in \mathbb{R}^n$, $\mu \in \mathbb{R}^p$, $\bar{\mu}_j \geq 0, \forall j \in \mathtt{K}$ com $(\lambda, \mu, \bar{\mu}) \neq 0$ tais que

$$\sum_{i\in\mathtt{I}} \lambda_i \begin{pmatrix} \nabla h_i(x^*) \\ 0 \end{pmatrix} + \sum_{j\in\mathtt{J}} \mu_j \begin{pmatrix} \nabla g_j(x^*) \\ e_j \end{pmatrix} + \sum_{j\in\mathtt{K}} \bar{\mu}_j \begin{pmatrix} 0 \\ -e_j \end{pmatrix} = 0.$$

Logo

$$\sum_{i\in\mathtt{I}} \lambda_i \nabla h_i(x^*) + \sum_{j\in\mathtt{J}} \mu_j \nabla g_j(x^*) = 0 \tag{1.6}$$

e

$$\sum_{j\in\mathtt{J}} \mu_j e_j - \sum_{j\in\mathtt{K}} \bar{\mu}_j e_j = 0, \tag{1.7}$$

assim, se $j \in \mathtt{J} \cap \mathtt{K}$, então $\mu_j = \bar{\mu}_j \geq 0$, além disso, se $j \in \mathtt{J} - \mathtt{K}$ temos $\mu_j = 0$ e se $j \in \mathtt{K} - \mathtt{J}$ temos que $\bar{\mu}_j = 0$, portanto $(\lambda, \mu, \bar{\mu}) \neq 0 \Leftrightarrow \lambda_i, i \in \mathtt{I}, \mu_j, j \in \mathtt{J} \cap \mathtt{K}$ são não todos nulos.

Como $K \subset \{j \in \{1, \ldots, p\} | s_i^* = 0\} = \mathtt{I}(x^*)$, então se definirmos $\mathtt{J}' = \mathtt{J} \cap \mathtt{K} \subset \mathtt{I}(x^*)$, de (1.6) podemos escrever

$$\sum_{i\in\mathtt{I}} \lambda_i \nabla h_i(x^*) + \sum_{j\in\mathtt{J}'} \mu_j \nabla g_j(x^*) = 0, \mu_j \geq 0, (\lambda, \mu) \neq 0,$$

ou seja, $\left(\{\nabla h_i(x^*)\}_{i\in\mathtt{I}}, \{\nabla g_j(x^*)\}_{j\in\mathtt{J}'}\right)$ é PLD. Como vale CPLD para o problema original, temos que existe uma vizinhança $V(x^*)$ de x^* tal que $\left(\{\nabla h_i(y)\}_{i\in\mathtt{I}}, \{\nabla g_j(y)\}_{j\in\mathtt{J}'}\right)$ é PLD para todo $y \in V(x^*)$. Ou seja, para cada $y \in V(x^*)$ existem $\lambda_i', \forall i \in \mathtt{I}$, $\mu_j' \geq 0, \forall j \in \mathtt{J}'$ não todos nulos tais que

$$\sum_{i\in\mathtt{I}} \lambda_i' \nabla h_i(y) + \sum_{j\in\mathtt{J}'} \mu_j' \nabla g_j(y) = 0.$$

Basta definir $\mu_j' = 0$ se $j \in \mathtt{J} - \mathtt{K}$, $\bar{\mu}_j' = \mu_j'$ se $j \in \mathtt{J} \cap \mathtt{K}$ e $\bar{\mu}_j' = 0$ se $j \in \mathtt{K} - \mathtt{J}$ e temos

$$\sum_{i\in\mathtt{I}} \lambda_i' \begin{pmatrix} \nabla h_i(y) \\ 0 \end{pmatrix} + \sum_{j\in\mathtt{J}} \mu_j' \begin{pmatrix} \nabla g_j(y) \\ e_j \end{pmatrix} + \sum_{j\in\mathtt{K}} \bar{\mu}_j' \begin{pmatrix} 0 \\ -e_j \end{pmatrix} = 0, \bar{\mu}' \geq 0, (\lambda, \mu, \bar{\mu}) \neq 0,$$

portanto vale a CPLD para o problema com variáveis de folga.

($\Leftarrow$) Seja $\mathtt{I} \subset \{1, \ldots, m\}$, $\mathtt{J} \subset \mathtt{I}(x^*)$ e suponha que existem $\lambda_i, \forall i \in \mathtt{I}, \mu_j \geq 0, \forall j \in \mathtt{J}$

não todos nulos tais que

$$\sum_{i\in \mathtt{I}} \lambda_i \nabla h_i(x^*) + \sum_{j\in \mathtt{J}} \mu_j \nabla g_j(x^*) = 0.$$

Defina $\mathtt{K} = \mathtt{J} \subset \{j \in \{1,\dots,p\} | s_i^* = 0\}$ e $\bar{\mu}_j = \mu_j$, assim

$$\sum_{i\in \mathtt{I}} \lambda_i \begin{pmatrix} \nabla h_i(x^*) \\ 0 \end{pmatrix} + \sum_{j\in \mathtt{J}} \mu_j \begin{pmatrix} \nabla g_j(x^*) \\ e_j \end{pmatrix} + \sum_{j\in \mathtt{K}} \bar{\mu}_j \begin{pmatrix} 0 \\ -e_j \end{pmatrix} = 0.$$

Como o problema com variáveis de folga satisfaz CPLD, temos que existe vizinhança V de (x^*, s^*) tal que para cada $(y, z) \in V$ existem $\lambda_i', \forall i \in \mathtt{I}, \mu_j', \forall j \in \mathtt{J}, \bar{\mu}_j' \geq 0, \forall j \in \mathtt{K}$ não todos nulos tais que

$$\sum_{i\in \mathtt{I}} \lambda_i' \begin{pmatrix} \nabla h_i(y) \\ 0 \end{pmatrix} + \sum_{j\in \mathtt{J}} \mu_j' \begin{pmatrix} \nabla g_j(y) \\ e_j \end{pmatrix} + \sum_{j\in \mathtt{K}} \bar{\mu}_j' \begin{pmatrix} 0 \\ -e_j \end{pmatrix} = 0,$$

ou seja

$$\sum_{i\in \mathtt{I}} \lambda_i' \nabla h_i(y) + \sum_{j\in J} \mu_j' \nabla g_j(y) = 0 \text{ e } \mu' = \bar{\mu}' \geq 0, (\lambda, \mu) \neq 0.$$

Portanto vale a CPLD para o problema original. $\square$

É imediato da demonstração da Proposição 1.20 que um resultado análogo vale para a CRCQ. A situação é diferente quando incluímos folgas irrestritas, como mostra a proposição abaixo.

Proposição 1.21. *O ponto x^* é viável e satisfaz LICQ para*

$$\begin{array}{ll} \textit{Minimizar} & f(x) \\ \textit{Sujeito a} & h(x) = 0 \\ & g(x) \leq 0 \end{array}$$

se, e somente se (x^*, s^*) é viável e satisfaz LICQ para

$$\begin{array}{ll} \textit{Minimizar} & f(x) \\ \textit{Sujeito a} & h(x) = 0 \\ & g(x) + s^2 = 0. \end{array}$$

Demonstração: Basta observar que o conjunto

$$\left\{ \begin{pmatrix} \nabla h_1(x^*) \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} \nabla h_m(x^*) \\ 0 \end{pmatrix}, \begin{pmatrix} \nabla g_1(x^*) \\ 2e_1 s_1^* \end{pmatrix}, \ldots, \begin{pmatrix} \nabla g_p(x^*) \\ 2e_p s_p^* \end{pmatrix} \right\} \text{ é LI,} \qquad (1.8)$$

se e somente se o conjunto

$$\{\nabla h_1, \ldots, \nabla h_m\} \bigcup \{\nabla g_j\}_{j \in \mathtt{I}(x^*)} \text{ é LI,}$$

já que $s_j^* = 0 \Leftrightarrow g_j(x^*) = 0$. □

Assim, se x^* satisfaz MF para

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & h(x) = 0 \\ & g(x) \leq 0 \end{array}$$

e não satisfaz LICQ, então temos que o par (x^*, s^*), viável para

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & h(x) = 0 \\ & g(x) + s^2 = 0, \end{array}$$

não satisfaz LICQ, que é o mesmo que MF, já que o problema possui apenas restrições de igualdade. Assim MF é perdida ao adicionarmos variáveis de folga irrestritas.

A CPLD também é perdida neste caso, basta considerar um problema com duas restrições $x \leq 0$ e $2x \leq 0$. Temos que $x^* = 0$ satisfaz CPLD, porém, para o problema com restrições $x + s_1^2 = 0$ e $2x + s_2^2 = 0$, o ponto $(x^*, s_1^*, s_2^*) = (0, 0, 0)$ não satisfaz CPLD, já que os gradientes $(1, 2s_2, 0)^{\mathtt{T}}$ e $(2, 0, 2s_2)^{\mathtt{T}}$ são LD em $(x, s_1, s_2) = (0, 0, 0)$ e LI para $(s_1, s_2) \neq 0$. Na verdade, este exemplo mostra que mesmo se x^* satisfaz MF para o problema original, não podemos garantir CPLD para o problema com variáveis de folga irrestritas.

Outra propriedade interessante da CPLD é que ela se mantém quando trocamos as restrições de igualdade $h(x) = 0$ por restrições de desigualdade $h(x) \leq 0$ e $-h(x) \leq 0$, o que não ocorre com MF. Veja a demonstração em [24].

1.3 Lema de Carathéodory estendido

Uma ferramenta muito útil para trabalharmos com condições de qualificação como a CPLD e CRCQ é o Lema de Carathéodory [9]. A seguir, enunciamos o celebrado Lema, incluindo um limitante para o tamanho dos novos multiplicadores.

Definição 1.22. *Seja* $\mathtt{J} = \{v_1, \ldots, v_m\} \subset \mathbb{R}^n$ *um multi-conjunto. Dizemos que* $\mathtt{I} \subset \mathtt{J}$ *é LI-maximal se* $\mathtt{I}$ *é LI e, caso* $\mathtt{I} \neq \mathtt{J}$*, então existe* $v \in \mathtt{J} - \mathtt{I}$ *tal que* $\mathtt{I} \cup \{v\}$ *é LD.*

Lema 1.23 (Carathéodory estendido)**.** *Sejam* $v_1, \ldots, v_m$ *vetores em* $\mathbb{R}^n$*. Se* $x = \sum_{i=1}^{m} \lambda_i v_i$*,* $\lambda_i \neq 0$*, então existem* $\mathrm{I} \subset \{1, \ldots, m\}$ *e escalares* $\lambda_i' \quad \forall i \in \mathtt{I}$ *tais que*

i) $x = \sum_{i \in \mathrm{I}} \lambda_i' v_i$

ii) $\{v_i\}_{i \in \mathrm{I}}$ *é LI-maximal*

iii) $\lambda_i \lambda_i' > 0 \quad \forall i \in \mathtt{I}$

iv) $|\lambda_i'| \leq 2^{m-1} |\lambda_i| \quad \forall i \in \mathrm{I}.$

Demonstração: Se $\{v_i\}_{i=1}^m$ é linearmente independente, tome $\mathrm{I} = \{1, \ldots, m\}$ e $\lambda_i' = \lambda_i$. Se $\{v_i\}_{i=1}^m$ é linearmente dependente, então existem $\alpha_i \in \mathbb{R}$ não todos nulos tais que $\sum_{i=1}^{m} \alpha_i v_i = 0$. Então, para todo $\gamma \in \mathbb{R}$ temos $x = \sum_{i=1}^{m} (\lambda_i - \gamma \alpha_i) v_i$. Seja $i^* = \operatorname{argmin}_i \left|\frac{\lambda_i}{\alpha_i}\right|$ e $\bar{\gamma} = \dfrac{\lambda_{i^*}}{\alpha_{i^*}}$, assim $\bar{\gamma}$ é o γ de menor módulo que anula pelo menos um dos coeficientes $\lambda_i - \gamma \alpha_i$. Se $\lambda_i(\lambda_i - \bar{\gamma}\alpha_i) < 0$ então $|\lambda_i|^2 = \lambda_i^2 < \lambda_i \bar{\gamma} \alpha_i = |\lambda_i||\bar{\gamma}||\alpha_i|$, com $|\alpha_i| \neq 0$, logo $|\bar{\gamma}| > \left|\dfrac{\lambda_i}{\alpha_i}\right|$, contradizendo a definição de $\bar{\gamma}$, assim $\lambda_i(\lambda_i - \bar{\gamma}\alpha_i) \geq 0$. Além disso, $|\lambda_i - \bar{\gamma}\alpha_i| \leq |\lambda_i| + |\bar{\gamma}||\alpha_i| \leq 2|\lambda_i|$, já que $|\bar{\gamma}| \leq \dfrac{|\lambda_i|}{|\alpha_i|}$. Incluindo em I apenas os índices tais que $\lambda_i' = \lambda_i - \bar{\gamma}\alpha_i \neq 0$, conseguimos escrever a combinação linear com pelo menos um vetor a menos. Assim, podemos repetir o argumento até que os vetores restantes sejam linearmente independentes e teremos $\lambda_i\lambda_i' > 0$ e $|\lambda_i'| \leq 2^{m-1}|\lambda_i| \quad \forall i \in \mathrm{I}$. □

Nas aplicações do Lema de Carathéodory, não nos preocuparemos com a restrição $\lambda_i \neq 0$, pois os índices com $\lambda_i = 0$ podem ser removidos da soma $x = \sum_{i=1}^{m} \lambda_i v_i$ sem alterar o valor de x (mesmo que todos os λ_i's sejam nulos).

Este novo item no Lema de Carathéodory, que nos dá um limitante para o tamanho dos novos multiplicadores, se mostrou muito útil. Ele foi usado na demonstração das Proposições 2.12, 2.22, 4.6 e 4.7. Para exemplificar a sua utilidade, vamos considerar uma sequência $x^k = \sum_{i=1}^{m} \lambda_i^k v_i^k$ e aplicamos o Lema 1.23 de Carathéodory, obtendo $x^k = \sum_{i \in \mathrm{I}_k} \lambda_i'^k v_i^k$, então se considerarmos uma subsequência com $\mathrm{I}_k = \mathrm{I}$, teremos que se $\{\lambda^k\}$ é limitada então $\{\lambda'^k\}$ também é limitada. Observe que a recíproca é falsa, basta considerar $v_1^k \neq 0$, $x^k = \lambda_1^k v_1^k + \lambda_2^k v_2^k$ e $\alpha_1^k v_1^k + \alpha_2^k v_2^k = 0$ com $\alpha_1^k = \alpha_2^k = 1$ e $\lambda_1^k = 1 + 10^k$, $\lambda_2^k = 10^k$. Assim $\left|\dfrac{\lambda_1^k}{\alpha_1^k}\right| > \left|\dfrac{\lambda_2^k}{\alpha_2^k}\right|$ para todo k, logo $\lambda_1'^k = \lambda_1^k - \left(\dfrac{\lambda_2^k}{\alpha_2^k}\right)\alpha_1^k = 1$ e $x^k = \lambda_1'^k v_1^k$ para todo k.

Do mesmo modo, se λ_i^k tende a zero então $\lambda_i'^k$ também tende a zero, sendo a recíproca falsa.

1.4 Condições de qualificação mais fracas

Nesta tese trabalhamos essencialmente com as condições MF e CPLD, porém, nesta seção apresentamos outras condições de qualificação, a saber, a quase-normalidade, a condição de Abadie e a condição de Guignard.

A seguir definimos a quase-normalidade (QN), proposta por Hestenes em [28]. Uma motivação para esta definição será dada na Seção 2.1, além de uma demonstração de que QN é uma condição de qualificação.

Definição 1.24 (quase-normalidade)**.** *Dizemos que $x^* \in X$ satisfaz a quase-normalidade (QN) quando $\left(\{\nabla h_i(x^*)\}_{i=1}^m, \{\nabla g_j(x^*)\}_{j\in\mathtt{I}(x^*)}\right)$ é PLI, ou, se $\lambda \in \mathbb{R}^m$ e $\mu_j \geq 0, \forall j \in \mathtt{I}(x^*)$, são tais que $(\lambda, \mu) \neq 0$ e $\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j\in\mathtt{I}(x^*)} \mu_j \nabla g_j(x^*) = 0$ então não existe sequência $x^k \to x^*$ tal que*

- $\lambda_i \neq 0 \Rightarrow \lambda_i h_i(x^k) > 0$ *para todo* $i \in \{1, \ldots, m\}$
- $\mu_j > 0 \Rightarrow \mu_j g_j(x^k) > 0$ *para todo* $j \in \mathtt{I}(x^*)$.

A seguir, mostraremos uma condição suficiente para que valha QN. Ressaltamos que este resultado também pode ser encontrado em [9].

Veremos que quando o conjunto viável X é definido por restrições de igualdade lineares e restrições de desigualdade côncavas, a QN é satisfeita automaticamente, de modo que quando o conjunto viável é côncavo, podemos garantir a existência de multiplicadores de Lagrange sem mais exigências.

Proposição 1.25. *Se h_i é afim para todo i e g_j é côncava para todo j então todo ponto viável $x \in X = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0\}$ satisfaz a QN.*

Demonstração: Seja $x \in X$ e suponha que

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

com $(\lambda, \mu) \neq 0, \mu \geq 0$. Assim, para todo $y \in \mathbb{R}^n$ temos

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

Como h_i é afim, g_j é côncava e x é viável, temos $h_i(y) = h_i(x) + \nabla h_i(x)^{\mathtt{T}}(y - x) = \nabla h_i(x)^{\mathtt{T}}(y-x)$ e $g_j(y) \leq g_j(x) + \nabla g_j(x)^{\mathtt{T}}(y-x) = \nabla g_j(x)^{\mathtt{T}}(y-x)$, para todo $i \in \{1, \ldots, m\}$

e todo $j \in \mathtt{I}(x)$. Portanto

$$\sum_{i=1}^{m} \lambda_i h_i(y) + \sum_{j \in \mathtt{I}(x)} \mu_j g_j(y) \leq 0, \quad \forall y \in \mathbb{R}^n.$$

Como $(\lambda, \mu) \neq 0$, concluímos que não existe sequência $x^k \to x$ tal que $\lambda_i \neq 0 \Rightarrow \lambda_i h_i(x^k) > 0$ e $\mu_j > 0 \Rightarrow \mu_j g_j(x^k) > 0$, isto é, vale QN. $\square$

Na verdade, é imediato da demonstração que conjunto viável côncavo implica a pseudonormalidade, uma condição mais forte que QN. Veja [9, 59].

Em [59, 7], os autores mostram que CPLD implica QN. Em [28, 9], mostra-se que QN implica a condição de Abadie (em [9], chamada de *quase-regularidade*). A seguir introduzimos alguns conceitos a fim de definirmos a condição de Abadie [1].

Dizemos que $K \subset \mathbb{R}^n$ é um *cone* se $tx \in K$ sempre que $x \in K$ e $t \geq 0$.

Definimos o *cone de direções viáveis*

$$L(x) = \{d \in \mathbb{R}^n | \nabla h_i(x)^{\mathtt{T}} d = 0, \forall i \in \{1, \ldots, m\}, \nabla g_j(x)^{\mathtt{T}} d \leq 0, \forall j \in \mathtt{I}(x)\}.$$

$L(x)$ é as vezes chamado de "cone de variações de primeira ordem" ou "cone linearizado". Observe que nem todo $d \in L(x)$ é uma direção "viável", no sentido que se $x \in X$ então $x + \alpha d \in X$ para todo α suficientemente pequeno. De fato, as direções com esta propriedade formam um cone contido em $L(x)$.

Definimos também o *cone tangente*

$$T(x) = \{d \in \mathbb{R}^n | \exists \{x^k\} \subset X, \exists t_k > 0 \text{ com } x^k \to x \text{ e } \frac{x^k - x}{t_k} \to d\}.$$

Definição 1.26 (Abadie). *Dizemos que $x \in X$ satisfaz a condição de qualificação de Abadie se $T(x) = L(x)$.*

Dado um cone $K \subset \mathbb{R}^n$, definimos o *cone dual* $K^* = \{v \in \mathbb{R}^n | v^{\mathtt{T}} d \geq 0, \forall v \in K\}$. A seguir apresentamos a condição de qualificação de Guignard, introduzida em [27].

Definição 1.27 (Guignard)**.** *Dizemos que* $x \in X$ *satisfaz a condição de qualificação de Guignard se* $T^*(x) = L^*(x)$.

Da definição, é imediato que a condição de Abadie implica a condição de Guignard. Em [26], os autores mostram que a condição de Guignard é necessária e suficiente para que valham as equações KKT, no seguinte sentido: se as restrições X são tais que para toda função objetivo f tal que x é minimizador de f em X, vale KKT, então x satisfaz Guignard.

A seguir mostramos uma definição equivalente para a condição de Abadie.

Proposição 1.28. *O ponto* $x \in X$ *satisfaz a condição de Abadie se, e somente se dado* $d \in L(x)$*, existem* $t_k > 0, t_k \to 0$ *e* $\{r_k\} \subset \mathbb{R}^n$ *tal que* $\frac{r_k}{t_k} \to 0$ *e* $x^k = x + t_k d + r_k \in X$ *para todo* k.

Demonstração: ($\Rightarrow$) Seja $d \in L(x)$. Pela condição de Abadie, $d \in T(x)$, logo existem $t_k > 0, t_k \to 0$ e $x^k \to x, x^k \in X$ tal que

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

Defina $r_k = x^k - x - t_k d$, assim

$$\frac{r_k}{t_k} = \frac{x^k - x}{t_k} - d \to 0,$$

e o resultado segue.

($\Leftarrow$) Seja $d \in L(x)$, por hipótese existem $t_k > 0, t_k \to 0$ e $\{r_k\} \subset \mathbb{R}^n$ tal que $\frac{r_k}{t_k} \to 0$ e $x^k = x + t_k d + r_k \in X$. Assim,

$$\frac{x^k - x}{t_k} = d + \frac{r_k}{t_k} \to d,$$

portanto $d \in T(x)$.

Seja $d \in T(x)$, logo existe $x^k \in X$, $x^k \to x$ e $t_k > 0, t_k \to 0$ tais que $\frac{x^k - x}{t_k} = d + \varepsilon_k$, com $\varepsilon_k \to 0$, assim, para $j \in \mathtt{I}(x)$ temos

$$0 \geq g_j(x^k) = g_j(x) + \nabla g_j(\tilde{x})^{\mathtt{T}}(x^k - x) = \nabla g_j(\tilde{x}_k)^{\mathtt{T}}(x^k - x) \Rightarrow \nabla g_j(\tilde{x}_k)^{\mathtt{T}}(d + \varepsilon_k) \leq 0,$$

para algum $\tilde{x}_k$ no segmento entre x e x^k. Tomando o limite, temos $\nabla g_j(x)^{\mathtt{T}} d \leq 0$ para $j \in \mathtt{I}(x)$. Analogamente temos $\nabla h_i(x)^{\mathtt{T}} d = 0$ para $i = 1, \ldots, m$, ou seja $d \in L(x)$. Assim, $T(x) = L(x)$ e vale a condição de Abadie. $\square$

Em [15] o autor utiliza uma condição similar à de Abadie para desenvolver uma teoria de otimização mais geral, com desigualdades generalizadas (veja o Apêndice 4.2). Para cada $d \in L(x)$, é exigido que exista uma função $r : (0, \bar{t}) \to \mathbb{R}^n$, para algum $\bar{t}$ suficientemente pequeno, tal que $\frac{r(t)}{t} \to 0$ e $x(t) = x + td + r(t) \in X$ para todo $t \in (0, \bar{t})$.

Na Figura 1.1 mostramos as relações entre as condições de qualificação apresentadas. As setas representam implicações estritas, isto é, as implicações contrárias são falsas.

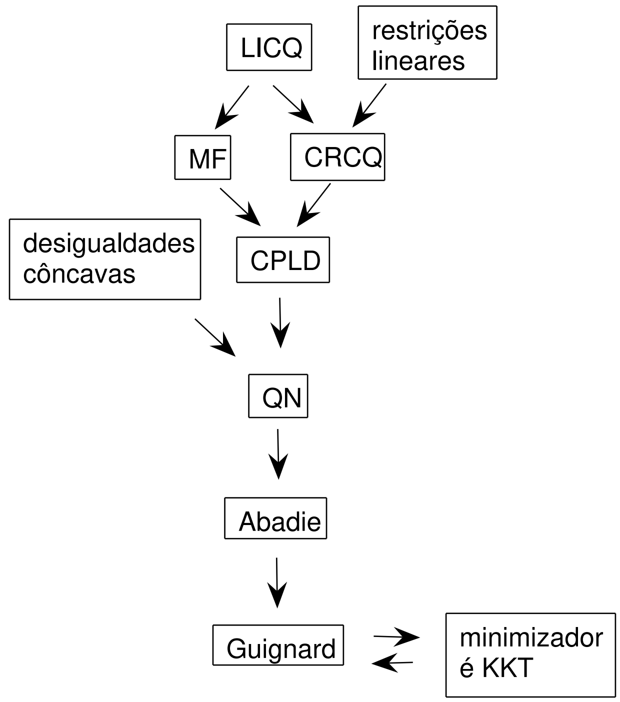

Figura 1.1: Relação entre as condições de qualificação.

Capítulo 2

Algoritmos de penalidade

Neste capítulo deduzimos as condições sequenciais de otimalidade relacionadas com os algoritmos de penalidade externa (CO-PE), penalidade interna (CO-PI) e um algoritmo misto de penalidade interna e externa (CO-PIE). No caso da penalidade externa obtemos CO-PE $\Rightarrow$ KKT ou não-CPLD. Para a penalidade interna, a condição de otimalidade obtida é mais fraca e temos CO-PI $\Rightarrow$ KKT ou não-MF, além disso mostramos que a condição H, necessária para a aplicação da penalidade interna, juntamente com a CPLD é equivalente à MF.

2.1 Penalidade externa

Vamos definir um algoritmo simples de penalidade externa, seguindo os moldes de [59, 42].

Considere o problema (P) abaixo:

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & x \in X, \end{array}$$

onde agora, além das restrições usuais $h(x) = 0$ e $g(x) \leq 0$, com funções continuamente diferenciáveis, o conjunto viável X é formado também por uma restrição adicional arbitrária do tipo $x \in \Omega$, com $\Omega \subset \mathbb{R}^n$ fechado e não vazio. Isto é $X = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0, x \in \Omega\}$.

Definimos a função $\Phi\colon \mathbb{R}^n \to \mathbb{R}$ que mede a inviabilidade, onde $Phi(x) = \frac{1}{2}\left(||h(x)||^2 + ||g^+(x)||^2\right)$, sendo $g_i^+(x) = \max\{0, g_i(x)\}$ e $g^+(x) = (g_1^+(x), \ldots, g_p^+(x))$. Observe que x é viável se, e somente se $Phi(x) = 0, x \in \Omega$.

Lema 2.1. *A função $Phi(x)$ é continuamente diferenciável.*

Demonstração: Provaremos que $r(x) = [\max\{0, s(x)\}]^2$ é continuamente diferenciável se $s\colon \mathbb{R}^n \to \mathbb{R}$ é continuamente diferenciável. Os casos $s(x) > 0$ e $s(x) < 0$ são imediatos já que r coincide com uma função continuamente diferenciável em uma vizinhança de x. Seja $x \in \mathbb{R}^n$ tal que $s(x) = 0$, se $s(x + te_i) < 0$ temos $\dfrac{r(x + te_i) - r(x)}{t} = 0$, senão temos $\dfrac{r(x + te_i) - r(x)}{t} = \dfrac{s(x + te_i)^2 - s(x)^2}{t}$. Em ambos os casos, $\dfrac{r(x + te_i) - r(x)}{t} \to \dfrac{\partial(s(x)^2)}{\partial x_i} = 2s(x)\dfrac{\partial s(x)}{\partial x_i} = 0$. Logo r é derivável e $\dfrac{\partial r(x)}{\partial x_i} = 2\max(0, s(x))\dfrac{\partial s(x)}{\partial x_i}$ é contínua. □

Seja $\{\rho_k\} \subset \mathbb{R}$, $\rho_k > 0$ com $\rho_k \to +\infty$. Para cada $k \in \mathbb{N}$ definimos o subproblema (P_k)

$$\begin{array}{ll} \text{Minimizar} & f(x) + \rho_k Phi(x) \\ \text{Sujeito a} & x \in \Omega. \end{array}$$

O algoritmo de penalidade externa consiste em encontrar uma sequência $\{x^k\}$ tal que para cada k, x^k é solução global do problema (P_k). Uma condição suficiente para a existência da solução x^k para todo k é que Ω seja também limitado.

Proposição 2.2. *Se (P) admite um ponto viável, então todo ponto limite de uma sequência $\{x^k\}$ gerada pelo algoritmo de penalidade externa é um minimizador global de (P).*

Demonstração: Seja z um ponto limite de $\{x^k\}$. Renomeando os índices se necessário, vamos escrever $x^k \to z$. Observe que $z \in \Omega$ já que Ω é fechado e $\{x^k\} \subset \Omega$.

Como x^k é solução do subproblema (P_k), então, sendo x um ponto viável arbitrário de (P), temos

$$f(x^k) + \rho_k Phi(x^k) \leq f(x) + \rho_k Phi(x) = f(x).$$

Além disso, como $\rho_k > 0$ e $Phi(x^k) \geq 0$, temos $f(x^k) \leq f(x^*)$. Assim, tomando o limite temos que $f(z) \leq f(x)$, pela continuidade de f.

Para mostrar que z é um minimizador global de (P), resta mostrar que z é viável. Suponha $Phi(z) > 0$, assim, pela continuidade de Phi temos que existe $\varepsilon > 0$ tal que $Phi(x^k) > \varepsilon$ para k suficientemente grande, logo

$$f(x^k) + \rho_k \varepsilon \leq f(x^k) + \rho_k Phi(x^k) \leq f(x),$$

para um ponto x viável qualquer. Como $f(x^k) \to f(z)$ e $\rho_k \to +\infty$, temos que o lado esquerdo tende para infinito, enquanto o lado direito não depende de k, o que é uma contradição. Logo $Phi(z) = 0$ e temos que z é minimizador global de (P). $\square$

Embora este resultado não seja utilizado neste trabalho, é interessante notar que mesmo quando o problema original não possui pontos viáveis, os pontos limites da sequência gerada pelo algoritmo de penalidade externa possuem uma propriedade agradável, a saber, eles minimizam a inviabilidade $Phi(x)$, sujeito a $x \in \Omega$. Na verdade o resultado é mais forte: os pontos limites minimizam $f(x)$ no conjunto dos minimizadores globais da inviabilidade, veja os detalhes em [42]. A justificativa para o nome do método deve-se ao fato de que os iterandos são quase sempre externos à região viável do problema. Com efeito, se x^k é viável para o problema original, do fato de que x^k é solução para o subproblema temos

$$f(x^k) = f(x^k) + \rho_k Phi(x^k) \leq f(z) + \rho_k Phi(z) = f(z),$$

para qualquer z viável, logo x^k seria solução global para o problema original.

Vamos considerar agora o problema geral de

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & h(x) = 0 \\ & g(x) \leq 0, \end{array}$$

e suponhamos que x^* é um minimizador local. Logo, existe $\delta > 0$ tal que x^* é um minimizador global de

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & h(x) = 0 \\ & g(x) \leq 0 \\ & ||x - x^*|| \leq \delta. \end{array}$$

Além disso, se definirmos a função regularizada $f_R(x) = f(x) + \frac{1}{2}||x - x^*||^2$, temos que x^* é o único minimizador global do problema regularizado (PR)

$$\begin{array}{ll} \text{Minimizar} & f_R(x) \\ \text{Sujeito a} & h(x) = 0 \\ & g(x) \leq 0 \\ & ||x - x^*|| \leq \delta. \end{array}$$

Considere uma sequência $\{x^k\}$ gerada pelo algoritmo de penalidade externa aplicado ao problema (PR), considerando $\Omega = \{x \in \mathbb{R}^n | \, ||x - x^*|| \leq \delta\}$. Observe que $\{x^k\}$ está bem definida, pois Ω é compacto e não vazio, além disso, este fato também garante que a sequência $\{x^k\}$ possui pelo menos um ponto limite z. Pela Proposição 2.2, z é um minimizador global de (PR), mas como o minimizador é único e igual a x^*, temos que a sequência possui exatamente um ponto limite $z = x^*$, isto é, $x^k \to x^*$.

Assim, para k suficientemente grande, a solução x^k para o k-ésimo subproblema é tal que $||x^k - x^*|| < \delta$, ou seja, x^k é solução local do problema irrestrito

$$\text{Minimizar} \quad f(x) + \frac{1}{2}||x - x^*||^2 + \rho_k Phi(x) \ .$$

Um cálculo simples mostra que o gradiente de $Phi(x)$ é dado por

$$\nabla Phi(x) = \sum_{i=1}^{m} h_i(x) \nabla h_i(x) + \sum_{j=1}^{p} g_j^+(x) \nabla g_j(x),$$

assim, temos que o gradiente da função objetivo do problema irrestrito é nulo em x^k, o que demonstra a Proposição abaixo.

Proposição 2.3. *Se x^* é um minimizador local, então para toda sequência $\rho_k > 0, \rho_k \to$*

$+\infty$, *existe* $\{x^k\} \subset \mathbb{R}^n, x^k \to x^*$ *tal que*

$$\nabla f(x^k) + \sum_{i=1}^{m} \rho_k h_i(x^k) \nabla h_i(x^k) + \sum_{j=1}^{p} \rho_k g_j^+(x^k) \nabla g_j^+(x^k) = x^* - x^k. \tag{2.1}$$

Vamos agora deduzir a condição de qualificação denominada quase-normalidade. Para isso, vamos tentar provar que x^*, o minimizador global do problema original, satisfaz as equações KKT, a partir da Proposição 2.3.

Defina $\lambda_i^k = \rho_k h_i(x^k), \forall i \in \{1, \ldots, m\}$ e $\mu_j^k = \rho_k g_j^+(x^k), \forall j \in \{1, \ldots, p\}$ e seja $M_k = \max\{|\lambda_i^k|, \mu_j^k, \forall i, j\}$.

Primeiro caso: se $\{M_k\}$ é limitada, então existe $\lambda \in \mathbb{R}^m$ e $\mu \in \mathbb{R}_+^p$ tal que, passando a uma subsequência se necessário, $\lambda^k \to \lambda$ e $\mu^k \to \mu$. Logo, tomando o limite em (2.1), obtemos

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

Além disso, se $g_j(x^*) < 0$, então $g_j(x^k) < 0$ para todo k suficientemente grande, de modo que $\mu_j^k = \rho_k g_j^+(x^k) = 0$ para todo k suficientemente grande, logo $\mu_j = 0$, concluindo que x^* é KKT.

Segundo caso: se existe uma subsequência tal que $M_k \to +\infty$, então $\left|\left|\left(\frac{\lambda^k}{M_k}, \frac{\mu^k}{M_k}\right)\right|\right|_\infty = 1$, portanto $\left(\frac{\lambda^k}{M_k}, \frac{\mu^k}{M_k}\right) \to (\lambda, \mu)$ em uma subsequência, de modo que $||(\lambda, \mu)||_\infty = 1$ pela continuidade da norma. Dividindo (2.1) por M_k e tomando o limite, obtemos

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

sendo que, como anteriormente, $g_j(x^*) < 0 \Rightarrow \mu_j = 0$, de modo que a soma acima se reduz aos índices j tais que $j \in \mathtt{I}(x^*)$. Logo $\left(\{\nabla h_i(x^*)\}_{i=1}^m, \{\nabla g_j(x^*)\}_{j \in \mathtt{I}(x^*)}\right)$ é PLD, além disso, se $\lambda_i \neq 0$, então λ_i^k tem o mesmo sinal de λ_i para todo k suficientemente grande (já que $\frac{\lambda_i^k}{M_k} \to \lambda_i$ e $M_k > 0$), mas como $\lambda_i^k = \rho_k h_i(x^k)$, temos que $h_i(x^k)$ tem o mesmo sinal que λ_i. Do mesmo modo, se $\mu_j > 0$, então $\mu_j^k > 0$ para todo k suficientemente grande, logo, como $\mu_j^k = \rho_k g_j^+(x^k)$, temos $g_j(x^k) > 0$.

Queremos que o segundo caso nunca aconteça, isto é, vamos impor que não exista uma sequência $\{x^k\}$ satisfazendo $\lambda_i \neq 0 \Rightarrow \lambda_i h_i(x^k) > 0$ e $\mu_j > 0 \Rightarrow \mu_j g_j(x^k) > 0$, quando $\left(\{\nabla h_i(x^*)\}_{i=1}^m, \{\nabla g_j(x^*)\}_{j\in \mathtt{I}(x^*)}\right)$ é PLD com multiplicadores $(\lambda, \mu) \in \mathbb{R}^m \times \mathbb{R}^p_+$. Esta é a definição da quase-normalidade, apresentada na Definição (1.24).

Do que foi mostrado anteriormente, é claro que QN é uma condição de qualificação. Com esta mesma técnica, via penalidade externa, é imediato ver que LICQ e MF são de fato condições de qualificação. Para a verificação de que CRCQ e CPLD são condições de qualificação, basta aplicar o Lema de Carathéodory (1.23) após dividir a equação (2.1) por M_k no segundo caso, assim, obtemos subconjuntos $\mathtt{I} \subset \{1, \dots, m\}$ e $\mathtt{J} \subset \mathtt{I}(x^*)$ tais que $\left(\{\nabla h_i(x^k)\}_{i\in\mathtt{I}} \bigcup \{\nabla g_j(x^k)\}_{i\in\mathtt{J}}\right)$ é LI, com $(\{\nabla h_i(x^*)\}_{i\in\mathtt{I}} \bigcup \{\nabla g_j(x^*)\}_{i\in\mathtt{J}})$ PLD, o que contradiz a CPLD (ou CRCQ). Veja detalhes em [59].

A partir da Proposição 2.3 vamos definir uma condição sequencial de otimalidade de maneira bastante simples.

Definição 2.4 (CO-PE). *Dizemos que x viável satisfaz a condição de otimalidade da penalidade externa (CO-PE) se existem $x^k \to x$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p_+$ tais que*

- $\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k) \to 0$
- $g_j(x) < 0 \Rightarrow \mu_j^k = 0$ *para k suficientemente grande*

Vamos mostrar uma definição equivalente de CO-PE.

Proposição 2.5. *x viável satisfaz CO-PE se, e somente se existem $x^k \to x$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p_+$, $\varepsilon_k > 0, \varepsilon_k \to 0$ tais que*

- $||\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^p \mu_j^k \nabla g_j(x^k)|| \leq \varepsilon_k$
- $||h(x^k)|| \leq \varepsilon_k, ||g(x^k)^+|| \leq \varepsilon_k$
- $g_j(x^k) < -\varepsilon_k \Rightarrow \mu_j^k = 0$

Demonstração: ($\Leftarrow$) Como $||h(x^k)|| \to 0$ e $||g^+(x^k)|| \to 0$, com $x^k \to x$, pela continuidade das funções temos que x é viável. Além disso, se $g_j(x) < 0$ então, pela continuidade, $g_j(x^k) < -\varepsilon_k$ para k suficientemente grande, de modo que vale CO-PE considerando as mesmas sequências $\{x^k\}, \{\lambda^k\}, \{\mu^k\}$.

($\Rightarrow$) Como x é viável, h e g são contínuas e $x^k \to x$ então $||h(x^k)|| \to 0$ e $||g^+(x^k)|| \to 0$. Vamos tomar $\varepsilon_k \to 0$ tal que $\varepsilon_k \geq \max\{-g_j(x^k), j \in \mathtt{I}(x)\}$. Assim, se $g_j(x) = 0$ temos $\varepsilon_k \geq -g_j(x^k)$, ou seja $g_j(x^k) \geq -\varepsilon_k$, o que mostra que $g_j(x^k) < -\varepsilon_k$ apenas quando $g_j(x) < 0$, donde por hipótese temos $\mu_j^k = 0$ para k suficientemente grande. $\square$

A condição CO-PE foi estudada em [59] (lá com o nome de AKKT-2), onde a autora mostra que AGP $\Rightarrow$ CO-PE, sendo AGP a condição de otimalidade oriunda do algoritmo de restauração inexata, introduzida em [45], da qual trataremos no Capítulo 3.

É interessante observar que a condição de otimalidade da penalidade externa é bastante forte, já que se x é viável e satisfaz CO-PE e CPLD, então x é KKT. Em outras palavras, CO-PE implica a condição de otimalidade KKT ou não-CPLD. Para verificar isto, suponhamos que x viável satisfaça CO-PE, isto é, existem sequências $x^k \to x, \{\lambda^k\} \subset \mathbb{R}^m, \{\mu^k\} \subset \mathbb{R}^p_+$ tais que

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

Pelo Lema 1.23 de Carathéodory, existem $\mathtt{I}_k \subset \{1, \ldots, m\}, \mathtt{J}_k \subset \mathtt{I}(x), \lambda_i'^k \, \forall i \in \mathtt{I}, \mu_j'^k \geq 0 \, \forall j \in \mathtt{J}$ tais que

$$\left(\left\{\nabla h_i(x^k)\right\}_{i \in \mathtt{I}_k}, \left\{\nabla g_j(x^k)\right\}_{j \in \mathtt{J}_k}\right) \text{ é LI} \tag{2.2}$$

$$\delta_k = \nabla f(x^k) + \sum_{i \in \mathtt{I}_k} \lambda_i'^k \nabla h_i(x^k) + \sum_{j \in \mathtt{J}_k} \mu_j'^k \nabla g_j(x^k), \tag{2.3}$$

e vamos tomar uma subsequência tal que $\mathtt{I}_k = \mathtt{I}$ e $\mathtt{J}_k = \mathtt{J}$.

Defina $M_k = \max\{|\lambda_i'^k| \, \forall i \in \mathtt{I}, \mu_j'^k \, \forall j \in \mathtt{J}\}$. Se $\{M_k\}$ é limitada, basta considerar uma subsequência convergente de (λ'^k, μ'^k) e tomando o limite, como $\delta_k \to 0$, temos que x

é KKT. Caso contrário, considere uma subsequência tal que $M_k \to +\infty$ e $(\frac{\lambda'^k}{M_k}, \frac{\mu'^k}{M_k}) \to (\lambda, \mu) \neq 0$, assim, dividindo (2.3) por M_k, e tomando o limite, temos que

$$\left(\{\nabla h_i(x)\}_{i\in \mathtt{I}}, \{\nabla g_j(x)\}_{j\in \mathtt{J}}\right) \text{ é PLD,}$$

o que juntamente com (2.2) mostra que não vale a CPLD.

Observamos também que o resultado não pode ser fortalecido utilizando a quase-normalidade. De fato, considere o problema

$$\begin{array}{ll} \text{Minimizar} & x \\ \text{Sujeito a} & -x^2 \leq 0. \end{array}$$

Temos que $x^* = 0$ satisfaz QN, já que, definindo $g(x) = -x^2$, temos $\mu \nabla g(x^*) = 0$ para qualquer $\mu > 0$, mas não existe sequência $x^k \to x$ tal que $\mu g(x^k) > 0$ quando $\mu > 0$, pois $g(x) \leq 0$ para todo x. Porém vale CO-PE, basta definir $x^k = \frac{1}{k}$ e $\mu^k = \frac{k}{2}$, assim, para $f(x) = x$ temos $\nabla f(x^k) + \mu_k \nabla g(x^k) = 1 + \frac{k}{2}(-\frac{2}{k}) = 0$. Além disso, x^* não é KKT.

2.2 Penalidade interna

A seguir definiremos um algoritmo de penalidade interna, também chamado de método de barreira, e faremos uma análise dos seus pontos limites nos moldes do que foi feito com o algoritmo de penalidade externa na Seção 2.1.

Considere o problema apenas com desigualdades

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & c(x) \geq 0 \\ & x \in \Omega, \end{array} \tag{2.4}$$

onde $\Omega \subset \mathbb{R}^n$ é um conjunto fechado arbitrário e $f : \mathbb{R}^n \to \mathbb{R}$, $c : \mathbb{R}^n \to \mathbb{R}^p$ são continuamente diferenciáveis. Por simplicidade, definimos as desigualdades como $c(x) \geq 0$, mas estas devem ser pensadas como $g(x) := -c(x) \leq 0$, no que diz respeito às definições do capítulo anterior.

Definimos o conjunto viável $X = \{x \in \Omega \,|\, c(x) \geq 0\}$ e o conjuntos dos pontos interiores, relativo às restrições, $X^0 = \{x \in \Omega \,|\, c(x) > 0\}$. Note que X^0 não é o interior topológico de X (veja [21]).

Suponhamos $X^0 \neq \emptyset$ e $X \subset \text{fecho}(X^0)$, ou seja, qualquer ponto viável pode ser arbitrariamente aproximado por pontos interiores. Esta hipótese será denominada **condição H**.

Definição 2.6 (H)**.** *Dizemos que um ponto $x \in X$ satisfaz a condição H se existe uma sequência $x^k \to x$ tal que $x^k \in X^0$.*

É intessante observar que a condição H não é uma condição de qualificação em geral. Com efeito, basta considerar o problema

$$\begin{array}{ll} \text{Minimizar} & x_1 \\ \text{Sujeito a} & x_1^3 + x_2 \geq 0 \\ & x_1^3 - x_2 \geq 0. \end{array}$$

O minimizador $(0,0)$ satisfaz a condição H com a sequência $x^k = (\frac{1}{k}, 0)$, mas $(0,0)$ não é KKT, já que não existem $\mu_1 \geq 0$ e $\mu_2 \geq 0$ tais que

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \mu_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \mu_2 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = 0.$$

Porém, quando as restrições são convexas, H é uma condição de qualificação, já que claramente implica a condição de Slater.

Um exemplo onde a condição H não é satisfeita é quando transformamos uma restrição de igualdade $h(x) = 0$ em duas desigualdades $h(x) \geq 0$ e $-h(x) \geq 0$, neste caso, o interior relativo às restrições é vazio. Mesmo quando o interior relativo às restrições é não vazio, a condição H pode falhar. Considere a função

$$k(x) = \begin{cases} x^n \text{ se } x > 0 \\ 0 \text{ se } x \leq 0 \end{cases}$$

e note que k é derivável até ordem n. Vamos considerar um problema com duas variáveis x e y com restrições $y \leq k(x)$ e $y \geq -k(x)$. Os pontos da forma $x > 0, y = 0$ pertencem ao interior relativo, enquanto os pontos da forma $x < 0, y = 0$, apesar de viáveis, não podem ser aproximados por pontos do interior relativo, veja a Figura 2.1.

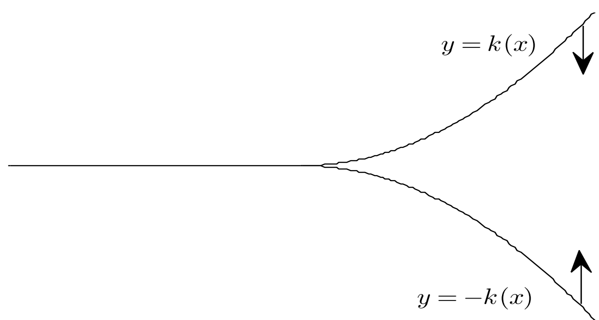

Figura 2.1: Exemplo com interior relativo não vazio tal que não vale a condição H.

Salientamos que é possível obter um exemplo como acima, de modo que as funções sejam infinitamente deriváveis.

Procedemos com a definição do algoritmo de penalidade interna.

Chamamos B de função barreira se $B : X^0 \to \mathbb{R}$ é contínua e se $y^k \in X^0$ com $c_i(y^k) \to 0$ para algum $i \in \{1, \ldots, p\}$ então $B(y^k) \to +\infty$. Escolhas comuns para B são $B(x) = -\sum_{i=1}^{p} \ln(c_i(x))$ (barreira logarítmica) e $B(x) = \sum_{i=1}^{p} \frac{1}{c_i(x)}$ (barreira inversa). Neste trabalho consideramos apenas a barreira logarítmica, mas resultados análogos valem para a barreira inversa.

O método de penalidade interna (ou método de barreira) é definido por uma sequência $\{r_k\} \subset \mathbb{R}$ com

$$r_k > 0 \text{ e } r_k \to 0,$$

e consiste em encontrar, para cada k, um minimizador global de

$$\begin{array}{ll} \text{Minimizar} & f(x) + r_k B(x) \\ \text{Sujeito a} & x \in X^0, \end{array}$$

onde geralmente, a solução x_{k-1} do $(k-1)$-ésimo problema é utilizada como aproximação inicial para resolver o k-ésimo problema.

As Proposições 2.7 e 2.8 a seguir mostram propriedades clássicas do algoritmo de penalidade interna. O primeiro resultado diz que todo ponto limite da sequência gerada pelo algoritmo é um minimizador global, e o segundo diz que o algoritmo está bem definido quando Ω é compacto, isto é, os subproblemas sempre admitem solução. Incluímos a demonstração por completude, embora estes resultados possam ser encontrados, por exemplo, em [8, 16].

Proposição 2.7. *Todo ponto limite da sequência $\{x^k\}$ gerada pelo método de penalidade interna aplicado ao problema (2.4) é um minimizador global.*

Demonstração: Seja x^* um ponto limite de $\{x^k\}$ e considere $K_0 \underset{\infty}{\subset} \mathbb{N}$ tal que $x^k \overset{k \in K_0}{\to} x^*$. Como Ω é fechado, $x^k \in X^0$ e c é contínua temos $x^* \in X$. Supondo que x^* não é minimizador global de (2.4) então existe $\bar{x} \in X$ tal que $f(\bar{x}) < f(x^*)$. Pela hipótese H e continuidade de f, existe $\tilde{x} \in X^0$ tal que $f(\tilde{x}) < f(x^*)$, assim, pela definição de x^k temos

$$f(x^k) + r_k B(x^k) \leq f(\tilde{x}) + r_k B(\tilde{x}). \tag{2.5}$$

Caso 1) Se $c(x^*) > 0$ temos $\lim_{k \in K_0} r_k B(x^k) = 0$, logo tomando o limite para $k \in K_0$ em (2.5) temos $f(x^*) \leq f(\tilde{x})$, contradição.

Caso 2) Se existe $i \in \{1, \ldots, p\}$ tal que $c_i(x^*) = 0$, então $B(x^k) \overset{k \in K_0}{\to} +\infty$, como $r_k > 0$ então para $k \in K_0$ suficientemente grande temos $r_k B(x^k) \geq 0$, logo por (2.5) temos $f(x^k) \leq f(\tilde{x}) + r_k B(\tilde{x})$, $\forall k \in K_0$ suficientemente grande. Tomando o limite para $k \in K_0$ temos $f(x^*) \leq f(\tilde{x})$, contradição.

Logo x^* é minimizador global de (2.4). $\square$

Proposição 2.8. *Se Ω é compacto, f é contínua e B é uma função barreira com $r_k > 0$ então o subproblema*

$$\begin{array}{ll} \textit{Minimizar} & \varphi_k(x) = f(x) + r_k B(x) \\ \textit{Sujeito a} & c(x) > 0 \\ & x \in \Omega \end{array}$$

admite solução para todo $k \in \mathbb{N}$.

Demonstração: Fixe $\varepsilon > 0$ e considere o problema de

$$\begin{array}{ll} \text{Minimizar} & \varphi_k(x) \\ \text{Sujeito a} & c(x) \geq \varepsilon \\ & x \in \Omega. \end{array}$$

Como $\{x \,|\, c(x) \geq \varepsilon\}$ é fechado, pois, pela continuidade de c, seu complementar é aberto e Ω é compacto, temos que o conjunto viável deste problema é compacto e pela continuidade de φ_k admite um minimizador global y.

Como $r_k > 0$, f é limitada inferiormente em $\{x \in \Omega | c(x) \geq 0\}$ e $B(x) \to +\infty$ se $c_i(x) \to 0$ então para cada $i \in \{1, \ldots, p\}$ existe $\varepsilon_i < \varepsilon, \varepsilon_i > 0$ tal que

$$0 < c_i(x) < \varepsilon_i, x \in \Omega, c(x) \geq 0 \Rightarrow \varphi_k(x) > \varphi_k(y).$$

Tome $\varepsilon' = \min\limits_{i=1,\ldots,p} \varepsilon_i > 0$, pela compacidade, o problema

$$\begin{array}{ll} \text{Minimizar} & \varphi_k(x) \\ \text{Sujeito a} & c(x) \geq \varepsilon' \\ & x \in \Omega \end{array}$$

admite solução global z tal que $\varphi_k(z) \leq \varphi_k(y)$, pois $\varepsilon' < \varepsilon$. Assim z também é solução global de

$$\begin{array}{ll} \text{Minimizar} & \varphi_k(x) \\ \text{Sujeito a} & c(x) > 0 \\ & x \in \Omega, \end{array}$$

pois se $x \in \Omega$ com $0 < c_i(x) < \varepsilon' \leq \varepsilon_i$ temos $\varphi_k(x) > \varphi_k(y) \geq \varphi_k(z)$. $\square$

Vamos agora deduzir a condição sequencial de otimalidade associada ao algoritmo de penalidade interna.

Lema 2.9. *Seja x^* um minimizador local de*

$$\begin{array}{ll} \textit{Minimizar} & f(x) \\ \textit{Sujeito a} & c(x) \geq 0 \end{array} \tag{2.6}$$

tal que vale a hipótese H, e considere a aplicação do método de penalidade interna ao problema regularizado

$$\begin{array}{ll} \textit{Minimizar} & f(x) + \frac{1}{2}||x - x^*||^2 \\ \textit{Sujeito a} & c(x) \geq 0 \\ & x \in \Omega = \{x \,|\, ||x - x^*|| \leq \delta\}, \end{array} \tag{2.7}$$

com função barreira $B(x) = -\sum_{i=1}^{p} \ln(c_i(x))$. Se $\delta > 0$ é suficientemente pequeno, então a sequência $\{x^k\}$ gerada é tal que $x^k \to x^*$,

$$\nabla f(x^k) - \sum_{i=1}^{p} \mu_i^k \nabla c_i(x^k) + x^k - x^* = 0$$

e $\mu_i^k = \dfrac{r_k}{c_i(x^k)} > 0$, onde r_k é o parâmetro de barreira.

Demonstração: Como x^* é minimizador local, existe $\delta > 0$ tal que x^* é solução global de

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & c(x) \geq 0 \\ & ||x - x^*|| \leq \delta, \end{array}$$

portanto x^* é a única solução global de (2.7). Pela Proposição 2.8, o subproblema

$$\begin{array}{ll} \text{Minimizar} & \varphi_k(x) = f(x) + \frac{1}{2}||x - x^*||^2 + r_k B(x) \\ \text{Sujeito a} & c(x) > 0 \\ & ||x - x^*|| \leq \delta \end{array}$$

admite solução global x^k, para todo k. Como $\{x^k\}$ é limitada, pois $||x^k - x^*|| \leq \delta$, temos que $\{x^k\}$ admite um ponto limite, que pela Proposição 2.7 é minimizador global de (2.7), como este é único temos que $x^k \to x^*$. Ou seja, para k suficientemente grande, x^k é minimizador local de

$$\begin{array}{ll} \text{Minimizar} & \varphi_k(x) \\ \text{Sujeito a} & c(x) > 0, \end{array}$$

logo $\nabla \varphi_k(x^k) = 0$, ou seja, $\nabla f(x^k) - \sum_{i=1}^p \mu_i^k \nabla c_i(x^k) + x^k - x^* = 0$ com $\mu_i^k = \dfrac{r_k}{c_i(x^k)} > 0$. No caso de barreira inversa teríamos $\mu_i^k = \frac{r_k}{c_i(x^k)^2} > 0$. □

O Lema 2.9 nos dá uma nova condição sequencial de otimalidade, associada ao algoritmo de penalidade interna, enunciada a seguir.

Definição 2.10 (CO-PI). *Dizemos que x viável satisfaz a condição de otimalidade da penalidade interna (CO-PI) se existem $x^k \to x$, $c(x^k) > 0$, $\{\mu^k\} \subset \mathbb{R}^p$, $\mu^k > 0$ tais que*

- $\nabla f(x^k) - \sum_{i=1}^p \mu_i^k \nabla c_i(x^k) \to 0$
- $c_i(x) > 0 \Rightarrow \mu_i^k \to 0$

Veremos a seguir que todo minimizador que satisfaz a condição H satisfaz CO-PI. Em outras palavras, CO-PI ou não-H é uma condição de otimalidade.

Proposição 2.11. *Se x é um minimizador local de* (2.6) *tal que vale H, então x satisfaz CO-PI*

Demonstração: Basta aplicar o Lema 2.9 e observar que se $c_i(x^*) > 0$, então existe $\varepsilon > 0$ tal que $c_i(x^k) > \varepsilon$, logo $\mu_i^k = \frac{r_k}{c_i(x^k)} \to 0$. □

Analogamente ao que foi feito no caso da penalidade externa, vamos mostrar que CO-PI é uma condição de otimalidade forte (desde que valha H). Veremos que sob a CPLD, todo ponto que satisfaz CO-PI é KKT. É o que mostram as duas próximas proposições.

Proposição 2.12. *Se x^* satisfaz CO-PI então existe $\mathtt{I} \subset \mathtt{I}(x^*)$ e $\mu_i, \forall i \in \mathtt{I} \cup \{0\}$ tais que* *i)* $\mu_0 \nabla f(x^*) - \sum_{i \in \mathtt{I}} \mu_i \nabla c_i(x^*) = 0$;

ii) $\mu_i \geq 0, \forall i \in I \cup \{0\}$;

iii) $\mu_i, \forall i \in I \cup \{0\}$ *são não todos nulos;*

iv) $c_i(x^*) > 0 \Rightarrow \mu_i = 0$;

v) Em toda vizinhança $V(x^*)$ *de* x^**, existe* $x \in V(x^*)$ *tal que* $(\emptyset, \{\nabla c_i(x)\}_{i \in \mathtt{I}})$ *é PLI.*

Demonstração: Sejam $\{x^k\}, \{\mu^k\}$ dados por CO-PI, assim

$$\nabla f(x^k) - \sum_{i=1}^{p} \mu_i^k \nabla c_i(x^k) + x^k - x^* = 0.$$

Pelo Lema 1.23 de Carathéodory existem $\mathtt{I}_k \subset \{1, \ldots, p\}, \mu_i^{'k} > 0 \, \forall i \in \mathtt{I}_k$ e $\mu_i^{'k} \leq 2^{p-1} \mu_i^k$ com $(\emptyset, \{\nabla c_i(x^k)\}_{i \in \mathtt{I}_k})$ PLI e

$$\nabla f(x^k) - \sum_{i \in \mathtt{I}_k} \mu_i^{'k} \nabla c_i(x^k) + x^k - x^* = 0.$$

Podemos considerar uma subsequência tal que $\mathtt{I}_k = \mathtt{I}$ (já que a quantidade de subconjuntos de $\{1, \ldots, p\}$ é finita), assim, definindo $\delta_k = \sqrt{1 + \sum_{i \in \mathtt{I}} (\mu_i^{'k})^2}, \mu_0^{''k} = \dfrac{1}{\delta_k}, \mu_i^{''k} = \dfrac{\mu_i^{'k}}{\delta_k}, \forall i \in \mathtt{I}$, temos

$$\mu_0^{''k} \nabla f(x^k) - \sum_{i \in \mathtt{I}} \mu_i^{''k} \nabla c_i(x^k) + \frac{x^k - x^*}{\delta_k} = 0.$$

Como $\{\mu_i^{''k}\}$ é limitada, considere uma subsequência tal que $\mu_i^{''k} \to \mu_i$, claramente $\mu_i \geq 0$ e $\mu \neq 0$ com

$$\mu_0 \nabla f(x^*) - \sum_{i \in \mathtt{I}} \mu_i \nabla c_i(x^*) = 0,$$

pois $x^k \to x^*$ e os gradientes de f e c são contínuos.

Por hipótese se $c_i(x^*) > 0$ então $\mu_i^k \to 0$, logo, como ${\mu_i'}^k \leq 2^{p-1}\mu_i^k \Rightarrow {\mu_i'}^k \to 0$, assim ${\mu_i''}^k = \frac{{\mu_i'}^k}{\delta_k} \to 0$ já que $\delta_k \geq 1$, ou seja, $\mu_i = 0$. $\square$

Observe que na demonstração da Proposição acima, utilizamos o novo resultado sobre a limitação dos multiplicadores obtidos no Lema de Carathéodory.

Proposição 2.13. *Se x^* viável satisfaz CO-PI e CPLD, então x^* satisfaz a condição KKT*

$$\nabla f(x^*) - \sum_{i=1}^{p} \mu_i \nabla c_i(x^*) = 0, \mu_i c_i(x^*) = 0, \mu_i \geq 0.$$

Demonstração: Basta observar que se $\mu_0 = 0$ na Proposição 2.12, temos que $(\emptyset, \{\nabla c_i(x^*)\}_{i\in I})$ é PLD pelas condições i), ii), iii) e $(\emptyset, \{\nabla c_i(x^k)\}_{i\in I})$ é PLI com $x^k \to x^*$ por v), contradizendo o fato que x^* satisfaz a CPLD. A condição $\mu_i c_i(x^*) = 0$ é equivalente à condição iv) no caso de $c_i(x^*) \geq 0$ e $\mu_i \geq 0$. $\square$

Este resultado mostra que este algoritmo simples de penalidade interna é convergente sob a CPLD, além de nos mostrar que sob H, CO-PI é uma condição de otimalidade forte pois implica a condição KKT ou não-CPLD. Observe que a condição H não pode ser dispensada em algoritmos de penalidade interna, pois o algoritmo gera pontos x^k que são internos (no sentido que satisfazem $c(x^k) > 0$), e se queremos obter convergência para a solução, devemos ter fecho$(X^0) \subset X$. Embora esta condição pareça ser "inofensiva", mostraremos a seguir o surpreendente fato de que CPLD+H é equivalente a MF, quando o problema possui apenas restrições de desigualdade.

Considere o problema

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & g(x) \leq 0, \end{array}$$

com $g : \mathbb{R}^n \to \mathbb{R}^p$, $g \in C^1$. Observe que neste caso x satisfaz a condição H quando existe uma sequência $x^k \to x$ com $g(x^k) < 0$.

É um fato conhecido que MF $\Rightarrow$ H; este resultado pode ser visto, por exemplo, em [21]. Incluímos aqui a demonstração por completude.

Lema 2.14. *Se x satisfaz MF então x satisfaz H.*

Demonstração: Seja $d \in \mathbb{R}^n$ dado pela Proposição 1.11, logo temos

$$g_i(x + \alpha d) = g_i(x) + \alpha \nabla g_i(x)^{\mathtt{T}} d + o(\alpha), \text{ com } \frac{o(\alpha)}{\alpha} \to 0 \text{ quando } \alpha \to 0.$$

Se $i \in \mathtt{I}$, temos $g_i(x) = 0$, e, dividindo por $\alpha > 0$ obtemos

$$\frac{g_i(x + \alpha d)}{\alpha} = \nabla g_i(x)^{\mathtt{T}} d + \frac{o(\alpha)}{\alpha},$$

como $\nabla g_i(x)^{\mathtt{T}} d < 0$ e $\frac{o(\alpha)}{\alpha} \to 0$, temos que $\frac{g_i(x+\alpha d)}{\alpha} < 0$ para $\alpha > 0$ suficientemente pequeno, ou seja, $g_i(x + \alpha d) < 0$. Se $i \notin \mathtt{I}(x)$, basta tomar α suficientemente pequeno e teremos $g_i(x+\alpha d) < 0$, pela continuidade de g. Logo, podemos tomar $\bar{\alpha} > 0$ satisfazendo $g_i(x + \alpha d) < 0$ para todo $i \in \{1, \ldots, p\}$ e todo $\alpha \in (0, \bar{\alpha}]$. Portanto x satisfaz H, basta considerar $x^k = x + \frac{\bar{\alpha}}{k} d$. □

Mostraremos agora que vale a recíproca do Lema 2.14, quando temos a CPLD.

Proposição 2.15. *x^* viável satisfaz CPLD e H se, e somente se x^* satisfaz MF.*

Demonstração: Seja x^* viável tal que vale H e CPLD. Observe que vale CPLD em x^* para o problema auxiliar de

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & -g_i(x) \leq 0, \forall i \in \mathtt{I}(x^*). \end{array}$$

Como CPLD $\Rightarrow$ QN (ver [7, 59]), temos que x^* satisfaz QN para o problema auxiliar. Se não vale MF, então existem $\mu_j \geq 0$ não todos nulos tais que

$$\sum_{j \in \mathtt{I}(x^*)} \mu_j \nabla g_j(x^*) = 0,$$

logo

$$\sum_{j\in \mathtt{I}(x^*)} \mu_j \nabla(-g_j(x^*)) = 0,$$

então, pela QN no problema auxiliar temos que não existe sequência $x^k \to x^*$ tal que

$$\mu_j > 0 \Rightarrow \mu_j(-g_j(x^k)) > 0.$$

Como existe um índice i_0 tal que $\mu_{i_0} > 0$, então não existe sequência tal que $g_{i_0}(x^k) < 0$, logo não vale H, o que é uma contradição. A implicação contrária é consequência imediata do Lema 2.14 e do fato que MF $\Rightarrow$ CPLD. $\square$

Note que QN+H $\not\Rightarrow$ MF, e o contra-exemplo é dado se consideramos a restrição $g(x) = -x^2$. No ponto $x^* = 0$ não vale MF, mas vale H com a sequência $x^k = \frac{1}{k}$ e vale QN pois não existe x com $g(x) > 0$.

É importante salientar que a Proposição 2.15 é válida quando o problema possui apenas restrições de desigualdade. Quando o problema inclui também restrições de igualdade, este resultado é, em geral, falso.

2.3 Penalidade interna-externa

Considere o problema

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & h(x) = 0 \\ & c(x) \geq 0 \\ & x \in \Omega, \end{array} \tag{2.8}$$

onde $\Omega \subset \mathbb{R}^n$ é um conjunto fechado arbitrário e $f\colon \mathbb{R}^n \to \mathbb{R}$, $h\colon \mathbb{R}^n \to \mathbb{R}^m$, $c\colon \mathbb{R}^n \to \mathbb{R}^p$ são continuamente diferenciáveis.

Definindo $\Omega' = \{x \in \Omega | h(x) = 0\}$, podemos aplicar o algoritmo de penalidade interna ao problema

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & c(x) \geq 0 \\ & x \in \Omega'. \end{array}$$

Ou seja, dada uma sequência $r_k > 0$, $r_k \to 0$ os subproblemas são dados por

$$\begin{array}{ll} \text{Minimizar} & f(x) - r_k \sum_{i=1}^{p} \ln(c_i(x)) \\ \text{Sujeito a} & h(x) = 0 \\ & c(x) > 0 \\ & x \in \Omega. \end{array}$$

A idéia do algoritmo de penalidade interna-externa é resolver cada subproblema utilizando penalidade externa, isto é, dada uma sequência $\rho_t > 0, \rho_t \to +\infty$, fixada a iteração k do algoritmo de penalidade interna, devemos resolver

$$\begin{array}{ll} \text{Minimizar} & f(x) - r_k \sum_{i=1}^{p} \ln(c_i(x)) + \frac{1}{2}\rho_t ||h(x)||^2 \\ \text{Sujeito a} & c(x) > 0 \\ & x \in \Omega, \end{array}$$

para todo t.

Para simplificar a notação, vamos supor que as sequências $\{r_k\}$ e $\{\rho_t\}$ estão indexadas pelo mesmo índice k. Note que isto não gera nenhuma restrição no algoritmo, já que a única imposição que temos é $r_k \to 0$ e $\rho_k \to +\infty$, independente de como estes parâmetros são atualizados, assim, o algoritmo consiste em a cada passo k, encontrar a solução x^k de

$$\begin{array}{ll} \text{Minimizar} & f(x) - r_k \sum_{i=1}^{p} \ln(c_i(x)) + \frac{1}{2}\rho_k ||h(x)||^2 \\ \text{Sujeito a} & c(x) > 0 \\ & x \in \Omega. \end{array}$$

Vamos definir uma condição análoga à condição H da penalidade interna.

Definição 2.16 (H1). *Dizemos que $x \in \Omega$, com $c(x) \geq 0$, satifaz H1 se existe $x^k \to x$ com $x^k \in \Omega$ e $c(x^k) > 0$.*

Note que H1 não depende das restrições de igualdade, portanto, se ignorarmos as restrições de igualdade do problema, temos que H1 coincide com H.

Para simplificar a notação, vamos definir $B(x) = -\sum_{i=1}^{p} \ln(c_i(x))$, a função barreira.

Vamos mostrar que mesmo quando o conjunto viável é vazio, o algoritmo possui a propriedade desejável de que os pontos limites minimizam uma medida de inviabilidade.

Proposição 2.17. *Se vale H1 então todo ponto limite da sequência gerada pelo algoritmo é solução global de*

$$\begin{array}{ll} \textit{Minimizar} & ||h(x)||^2 \\ \textit{Sujeito a} & c(x) \geq 0 \\ & x \in \Omega \end{array} \tag{2.9}$$

Demonstração: Seja x^* ponto limite de $\{x^k\}$ e considere uma subsequência tal que $x^k \to x^*$. Como Ω é fechado, temos que $x^* \in \Omega$ e pela continuidade de c temos $c(x^*) \geq 0$.

Suponha que x^* não é minimizador global de (2.9), logo existe $z' \in \Omega, c(z') \geq 0$ tal que $||h(z')||^2 < ||h(x^*)||^2$. Pela condição H1 e continuidade de $||h(\cdot)||^2$, existe $z \in \Omega, c(z) > 0$ tal que $||h(z)||^2 < ||h(x^*)||^2$. Novamente pela continuidade de $||h(\cdot)||^2$, existe $\varepsilon > 0$ tal que para k suficientemente grande vale $||h(z)||^2 + \varepsilon < ||h(x^k)||^2$, logo

$$f(x^k) + r_k B(x^k) + \frac{1}{2}\rho_k ||h(x^k)||^2 > f(z) + r_k B(z) +$$
$$+\frac{1}{2}\rho_k ||h(z)||^2 + \frac{1}{2}\rho_k \varepsilon + f(x^k) - f(z) + r_k B(x^k) - r_k B(z).$$

Como $f(x^k) - f(z) + r_k B(x^k) - r_k B(z)$ é limitada inferiormente, temos que

$$\frac{1}{2}\rho_k \varepsilon + f(x^k) - f(z) + r_k B(x^k) - r_k B(z) > 0,$$

para k suficientemente grande. Portanto

$$f(x^k) + r_k B(x^k) + \frac{1}{2}\rho_k ||h(x^k)||^2 > f(z) + r_k B(z) + \frac{1}{2}\rho_k ||h(z)||^2,$$

com $z \in \Omega, c(z) > 0$, o que contradiz a definição de x^k como solução do subproblema. $\square$

Em particular, se a região viável é não vazia, temos que todo ponto limite é viável.

Vamos mostrar que se a região viável é não vazia, então todo ponto limite é minimizador global do problema original. Para isso, vamos precisar da condição H2 dada abaixo.

Definição 2.18 (H2). *Dizemos que x viável para (2.8) satifaz H2 se existe $x^k \to x$ com $x^k \in \Omega$, $h(x^k) = 0$ e $c(x^k) > 0$.*

Proposição 2.19. *Se (2.8) admite um minimizador global que satisfaz H2 então todo ponto limite da sequência gerada pelo algoritmo é minimizador global de (2.8).*

Demonstração: Seja x^* ponto limite de $\{x^k\}$ e considere uma subsequência tal que $x^k \to x^*$. Suponha que exista $z' \in \Omega, c(z') \geq 0, h(z') = 0$ tal que $f(z') < f(x^*)$. Por H2 e continuidade de f, existe $z \in \Omega, c(z) > 0, h(z) = 0$ tal que $f(z) < f(x^*)$. Pela definição de x^k temos

$$f(x^k) + r_k B(x^k) + \frac{1}{2}\rho_k ||h(x^k)||^2 \leq f(z) + r_k B(z) + \frac{1}{2}\rho_k ||h(z)||^2,$$

como $r_k B(x^k) \to 0$ ou $r_k B(x^k) > 0$ para k suficientemente grande temos

$$f(x^k) + \frac{1}{2}\rho_k ||h(x^k)||^2 \leq f(z) + r_k B(z) + \frac{1}{2}\rho_k ||h(z)||^2,$$

mas como $r_k B(z) \to 0, h(z) = 0$ e $f(x^k) \to f(x^*)$, então $f(x^*) \leq f(z)$, o que é uma contradição. $\square$

Vamos mostrar que se não existe minimizador global satisfazendo H2, a conclusão da Proposição 2.19 pode ser falsa. Por simplicidade, vamos considerar a barreira inversa. Considere o problema

$$\begin{array}{ll} \text{Minimizar} & x \\ \text{Sujeito a} & x^3 \geq 0 \\ & x(x-1) = 0, \end{array}$$

e observe que o minimizador global $z = 0$ não satisfaz H2.

Vamos supor que o parâmetro de barreira r_k é tal que $r_k = \frac{1}{\rho_k}$, assim, os subproblemas do algoritmo de penalidade interna-externa, utilizando a barreira inversa são dados por

$$\begin{array}{ll} \text{Minimizar} & F(x,\rho_k) = x + \dfrac{1}{\rho_k x^3} + \rho_k x^2 (x-1)^2 \\ \text{Sujeito a} & x^3 > 0 \end{array} \quad , \rho_k \to +\infty.$$

Se $\{x^k\}$ é uma sequência gerada pelo algoritmo, vamos mostrar que $z = 0$ não é um ponto limite desta sequência. Como $F(1,\rho_k) = 1 + \frac{1}{\rho_k} \to 1$, é suficiente mostrar que para todo $\rho > 0$ e todo $x \in (0, \frac{1}{2})$ temos $F(x,\rho) > \sqrt{2}$. Com efeito, como $(1-x)^2 > \frac{1}{4}$ e $x > 0$ então

$$F(x,\rho) > \frac{1}{\rho x^3} + \frac{\rho x^2}{4},$$

além disso, como $\frac{1}{x} > 2$ temos

$$F(x,\rho) > \frac{2}{\rho x^2} + \frac{\rho x^2}{4}.$$

Vamos analisar a função $g(t) = \frac{2}{t} + \frac{t}{4}, t > 0$. Temos $g'(t) = -\frac{2}{t^2} + \frac{1}{4}$, assim, o único ponto estacionário de g é $t^* = \sqrt{8}$. De fato, t^* é o minimizador global de g pois $g''(t^*) = \frac{4}{(t^*)^3} > 0$ e $g(t) \to +\infty$ quando $t \to 0^+$ ou $t \to +\infty$. Portanto temos

$$F(x,\rho) > g(t^*) = \sqrt{2}, \, \forall \rho > 0, \, \forall x \in \left(0, \frac{1}{2}\right).$$

Assim, da definição de x^k e do fato que $F(1,\rho_k) = 1 + \frac{1}{\rho_k} \to 1$, para k suficientemente grande temos $x^k \geq \frac{1}{2}$, o que conclui a demonstração.

Observe que um teorema de boa definição do algoritmo também pode ser provado, como no caso de penalidade interna, quando Ω é compacto. Omitiremos esta demonstração, pois é análoga à da Proposição 2.8.

Procedendo como no caso de penalidade interna, supondo que x^* é solução local de

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & h(x) = 0 \\ & c(x) \geq 0 \end{array}$$

e aplicando o algoritmo de penalidade interna-externa para o problema regularizado

$$\begin{array}{ll} \text{Minimizar} & f(x) + \frac{1}{2}||x - x^*||^2 \\ \text{Sujeito a} & h(x) = 0 \\ & c(x) \geq 0 \\ & x \in \Omega = \{x \,|\, ||x - x^*|| \leq \delta\}, \end{array}$$

obtemos que a sequência gerada $\{x^k\}$, solução de

$$\begin{array}{ll} \text{Minimizar} & f(x) + \frac{1}{2}||x - x^*||^2 - r_k \sum_{i=1}^{p} \ln(c_i(x)) + \frac{1}{2}\rho_k||h(x)||^2 \\ \text{Sujeito a} & c(x) > 0 \\ & ||x - x^*|| \leq \delta, \end{array}$$

é tal que $||x^k - x^*|| < \delta$ para k suficientemente grande, logo vale

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

Sendo que se $c_j(x^*) > 0$, então existe $\varepsilon > 0$ tal que $c_j(x^k) > \varepsilon$ para todo k suficientemente grande, assim, $\frac{r_k}{c_j(x^k)} < \frac{r_k}{\varepsilon} \to 0$. Logo, definindo $\lambda_i^k = \rho_k h_i(x^k)$ e $\mu_j^k = \frac{r_k}{c_j(x^k)} \geq 0$ temos que

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

com $\mu_j^k \to 0$ se $c_j(x^*) > 0$.

A existência de sequências $\{x^k\}, \{\lambda^k\}, \{\mu^k\}$ com as propriedades acima nos fornece a condição de otimalidade satisfeita pelos pontos limites do algoritmo de penalidade interna-externa, desde que valha a condição H2.

Definição 2.20 (CO-PIE)**.** *Dizemos que x viável para (2.8 satisfaz a condição de otimalidade da penalidade interna-externa (CO-PIE) se existem $x^k \to x$, $c(x^k) > 0$, $\{\lambda^k\} \subset \mathbb{R}^m, \{\mu^k\} \subset \mathbb{R}^p$, $\mu^k > 0$ tais que*

- $\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) - \sum_{j=1}^{p} \mu_j^k \nabla c_j(x^k) \to 0$

- $c_j(x) > 0 \Rightarrow \mu_j^k \to 0$

Vamos mostrar uma definição equivalente para CO-PIE, que nos dirá em particular que CO-PIE $\Rightarrow$ CO-PE, a condição de otimalidade da penalidade externa.

Proposição 2.21. *O ponto x viável para (2.8) satisfaz CO-PIE se, e somente se existem $x^k \to x$, $c(x^k) > 0$, $\{\lambda^k\} \subset \mathbb{R}^m, \{\mu^k\} \subset \mathbb{R}^p$, $\mu^k \geq 0$ tais que*

- $\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) - \sum_{j=1}^p \mu_j^k \nabla c_j(x^k) \to 0$

- $c_j(x) > 0 \Rightarrow \mu_j^k = 0$ *para todo k suficientemente grande*

Demonstração: ($\Leftarrow$) É imediato.
($\Rightarrow$) Defina $\bar{\mu}_j^k = 0$ se $c_j(x) > 0$ e $\bar{\mu}_j^k = \mu_j^k$ caso contrário. Defina também

$$A_k = \nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) - \sum_{j=1}^p \mu_j^k \nabla c_j(x^k)$$

e

$$B_k = \sum_{j|c_j(x)>0} \mu_j^k \nabla c_j(x^k).$$

Temos que

$$\nabla f(x^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) - \sum_{j=1}^p \bar{\mu}_j^k \nabla c_j(x^k) = A_k + B_k,$$

com $A_k \to 0$ por hipótese e $B_k \to 0$ pela continuidade do gradiente de c_j e do fato que $\mu_j^k \to 0$ para todo j tal que $c_j(x) > 0$. Assim $A_k + B_k \to 0$ e as sequências $\{x^k\}, \{\lambda^k\}, \{\bar{\mu}_j^k\}$ satisfazem as condições do enunciado. $\square$

Vamos mostrar que CO-PIE é uma condição de otimalidade forte (desde que valha H2), pois se x^* satisfaz CO-PIE e CPLD, então x^* é KKT.

Proposição 2.22. *Se x^* viável satisfaz CO-PIE e CPLD então x^* é KKT.*

Demonstração: Sejam x^k, λ^k, μ^k dados por CO-PIE tais que

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) - \sum_{j=1}^{p} \mu_j^k \nabla c_j(x^k) = \varepsilon_k,$$

com $\varepsilon_k \to 0$ e $\mu_j^k \to 0$ se $c_j(x^*) > 0$.

Aplicando o Lema 1.23 de Carathéodory, temos que existem $\mathtt{I}_k \subset \{1, \ldots, m\}$, $\mathtt{J}_k \subset \{1, \ldots, p\}$, $\lambda_i'^k$ e $\mu_j'^k \geq 0$ tais que

$$\nabla f(x^k) + \sum_{i \in \mathtt{I}_k} \lambda_i'^k \nabla h_i(x^k) - \sum_{j \in \mathtt{J}_k} \mu_j'^k \nabla c_j(x^k) = \varepsilon_k, \tag{2.10}$$

com $\big(\{\nabla h_i(x^k)\}_{i \in \mathtt{I}_k}, \{\nabla g_j(x^k)\}_{j \in \mathtt{J}_k}\big)$ LI.

Observe que podemos tomar uma subsequência tal que $\mathtt{I}_k = \mathtt{I}$ e $\mathtt{J}_k = \mathtt{J}$.

Se $\{(\lambda'^k, \mu'^k)\}$ é limitada, podemos tomar uma subsequência convergente $\lambda'^k \to \lambda$ e $\mu'^k \to \mu \geq 0$. Pelo resultado estendido do Lema de Carathéodory, temos que $|\mu_j'^k| \leq 2^{p-1}|\mu_j|$, assim, se $c_j(x^*) > 0$, temos $\mu_j^k \to 0$, logo $\mu_j'^k \to 0$, ou seja $\mu_j = 0$ e x^* é KKT.

Se existe uma subsequência tal que $||(\lambda'^k, \mu'^k)||_\infty \to +\infty$, podemos tomar uma subsequência tal que

$$\frac{(\lambda'^k, \mu'^k)}{||(\lambda'^k, \mu'^k)||_\infty} \to (\lambda, \mu), \mu \geq 0,$$

já que a sequência é limitada. Além disso, pela continuidade da norma, temos $(\lambda, \mu) \neq 0$. Observe que se $c_j(x^*) > 0$, então como anteriormente temos $\mu_j = 0$.

Assim, dividindo (2.10) por $||(\lambda'^k, \mu'^k)||_\infty$ e tomando o limite, definindo $\mathtt{J}' = \mathtt{J} \cap \mathtt{I}(x^*)$ temos que $(\{\nabla h_i(x^*)\}_{i \in \mathtt{I}}, \{\nabla g_j(x^*)\}_{j \in \mathtt{J}'})$ é PLD com $\big(\{\nabla h_i(x^k)\}_{i \in \mathtt{I}}, \{\nabla g_j(x^k)\}_{j \in \mathtt{J}'}\big)$ LI, já que $\mathtt{J}' \subset \mathtt{J}$, o que contradiz a CPLD. $\square$

Observe que isto mostra que os pontos limites do algoritmo de penalidade interna-externa são estacionários quando vale CPLD e H2.

Capítulo 3

Condições de otimalidade do tipo AGP

Em [45], os autores definiram a condição de otimalidade AGP (*approximate gradient projection*), uma condição sequencial de otimalidade associada ao algoritmo de restauração inexata. Os autores mostraram que AGP $\Rightarrow$ KKT ou não-MF, sendo que este resultado foi fortalecido em [24] obtendo AGP $\Rightarrow$ KKT ou não-CPLD. Quando o problema possui um conjunto extra de restrições lineares K, definimos uma condição análoga, L-AGP, impondo que a sequência esteja em K. Neste caso obtemos L-AGP $\Rightarrow$ KKT ou não-CPLD. O mesmo pode ser feito quando K é um conjunto convexo qualquer (definido por restrições de igualdade e desigualdade), neste caso denominamos a condição de C-AGP e obtemos C-AGP $\Rightarrow$ KKT ou não-MF, além disso, obtemos um contra-exemplo que não podemos substituir MF por CPLD. Mostramos uma definição equivalente para as condições AGP, L-AGP e C-AGP, onde o problema de projeção é substituído por um problema de minimizar uma quadrática com Hessiana definida positiva em um certo espaço tangente, com restrições convexas, o que nos possibilita uma maior flexibilidade na escolha da direção de busca no passo de otimalidade de um algoritmo de restauração inexata.

3.1 Projeções

Seja $\Omega \subset \mathbb{R}^n$ fechado e não vazio, e considere o problema

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & x \in \Omega, \end{array}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$ é contínua e coerciva, isto é, $f(x) \to +\infty$ quando $||x|| \to +\infty$.

Vamos mostrar o resultado clássico de que este problema admite pelo menos uma solução. Como $\Omega \neq \emptyset$, seja $x_0 \in \Omega$ e defina $r_0 = ||x_0||$. Portanto existe x_1 solução de

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & x \in \Omega \\ & ||x|| \leq r_0, \end{array}$$

já que o conjunto viável é compacto e não vazio. Como f é coerciva, tome $M = f(x_1)$, logo existe $r_1 \geq r_0$ tal que $f(x) > M$ para todo $x \in \Omega$ com $||x|| > r_1$. Como $r_1 \geq r_0$, temos que o problema

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & x \in \Omega \\ & ||x|| \leq r_1, \end{array}$$

admite solução x_2, além disso, $f(x_2) \leq M = f(x_1)$, já que o conjunto viável contém o anterior. Assim, $f(x_2) \leq f(x), \forall x \in \Omega, ||x|| \leq r_1$ e $f(x) > M, \forall x \in \Omega, ||x|| > r_1$, portanto x_2 é solução para o problema original.

Considerando a função coerciva $f(x) = \frac{1}{2}||x-y||^2$, podemos então definir uma projeção ortogonal de y em Ω como sendo uma solução de

$$\begin{array}{ll} \text{Minimizar} & \frac{1}{2}||x-y||^2 \\ \text{Sujeito a} & y \in \Omega. \end{array} \tag{3.1}$$

Observe que a projeção ortogonal neste caso não é necessariamente única.

O caso mais interessante é quando Ω é convexo, fechado e não vazio. Neste caso, a projeção ortogonal é única. Uma maneira simples de ver isso é quando Ω é definido por igualdades lineares e desigualdades convexas, deste modo, o problema (3.1) é convexo. É um fato conhecido que um problema convexo (com função objetivo estritamente convexa) admite no máximo uma solução global (ver por exemplo [9]). Assim, quando $\Omega \neq \emptyset$ é convexo e fechado, para cada $x \in \mathbb{R}^n$ definimos a projeção ortogonal de x em Ω, denotado por $P_\Omega(x)$ como a única solução de (3.1).

Os dois próximos lemas serão muito úteis no decorrer deste trabalho. Estes resultados podem ser encontrados por exemplo em [9, 36]. É válido salientar que a projeção ortogonal pode ser definida em espaços mais gerais, como espaços de Hilbert (reais), de modo que os lemas abaixo valem com a mesma demonstração.

Lema 3.1. $z = P_\Omega(x)$ *se, e somente se* $z \in \Omega$ *e* $(x - z)^T(y - z) \leq 0 \quad \forall y \in \Omega$.

Demonstração: ($\Rightarrow$) Seja $z = P_\Omega(x)$ e $y \in \Omega$, como Ω é convexo e $z \in \Omega$, temos

$$(1 - t)z + ty = z + t(y - z) \in \Omega, \quad 0 \leq t \leq 1.$$

Pela definição da projeção, a função $\phi(t) = ||x - (z + t(y - z))||^2$, restrita a $0 \leq t \leq 1$ admite um minimizador em $t = 0$. Pelas equações KKT, existe $\mu \geq 0$ tal que $\phi'(0) + \mu(-1) = 0$, em outras palavras, $\phi'(0) \geq 0$.

Mas $\phi(t) = ||x - (z + t(y - z))||^2 = \langle x - z - t(y - z), x - z - t(y - z)\rangle =$
$= \langle x - z, x - z - t(y - z)\rangle - t\,\langle y - z, x - z - t(y - z)\rangle = ||x - z||^2 - 2t\,\langle x - z, y - z\rangle +$ $t^2||y - z||^2$. Como $\phi'(t) = -2\,\langle x - z, y - z\rangle + 2t||y - z||^2$, de $\phi'(0) \geq 0$ temos $\langle x - z, y - z\rangle \leq 0, \forall y \in \Omega$.

($\Leftarrow$) $z \in \Omega$ e $0 \geq \langle x - z, y - z\rangle = \langle x - z, y - x + x - z\rangle = \langle x - z, y - x\rangle + ||x - z||^2$, assim $||x - z||^2 \leq \langle z - x, y - x\rangle \leq ||z - x||.||y - x||$. Se $x = z$, então $x \in \Omega$ e da definição temos claramente $z = P_\Omega(x)$. Caso contrário, dividindo por $||z - x||$ temos $||z - x|| \leq ||y - x||, \forall y \in \Omega$, portanto $z = P_\Omega(x)$. $\square$

Lema 3.2. *A função* $P_\Omega(\cdot) : \mathbb{R}^n \to \Omega$ *é não expansiva, isto é, se* $x, x' \in \mathbb{R}^n$ *então* $||P_\Omega(x) - P_\Omega(x')|| \leq ||x - x'||$.

Demonstração: Pelo Lema anterior temos

$$\langle x - P_\Omega(x), y - P_\Omega(x)\rangle \leq 0, \forall y \in \Omega,$$

$$\langle x' - P_\Omega(x'), y' - P_\Omega(x')\rangle \leq 0, \forall y' \in \Omega.$$

Tomando $y = P_\Omega(x')$ e $y' = P_\Omega(x)$ e somando as duas desigualdades temos

$$\langle x - P_\Omega(x), P_\Omega(x') - P_\Omega(x)\rangle + \langle x' - P_\Omega(x'), P_\Omega(x) - P_\Omega(x')\rangle \leq 0 \Rightarrow$$

$$\langle x - P_\Omega(x), P_\Omega(x') - P_\Omega(x)\rangle - \langle x' - P_\Omega(x'), P_\Omega(x') - P_\Omega(x)\rangle \leq 0 \Rightarrow$$

$$\langle x - x', P_\Omega(x') - P_\Omega(x)\rangle + \langle P_\Omega(x') - P_\Omega(x), P_\Omega(x') - P_\Omega(x)\rangle \leq 0 \Rightarrow$$

$$||P_\Omega(x) - P_\Omega(x')||^2 \leq \langle x - x', P_\Omega(x) - P_\Omega(x')\rangle \leq ||x - x'||.||P_\Omega(x) - P_\Omega(x')||.$$

Portanto $||P_\Omega(x) - P_\Omega(x')|| \leq ||x - x'||$. $\square$

A Figura 3.1 a seguir ilustra o Lema 3.2.

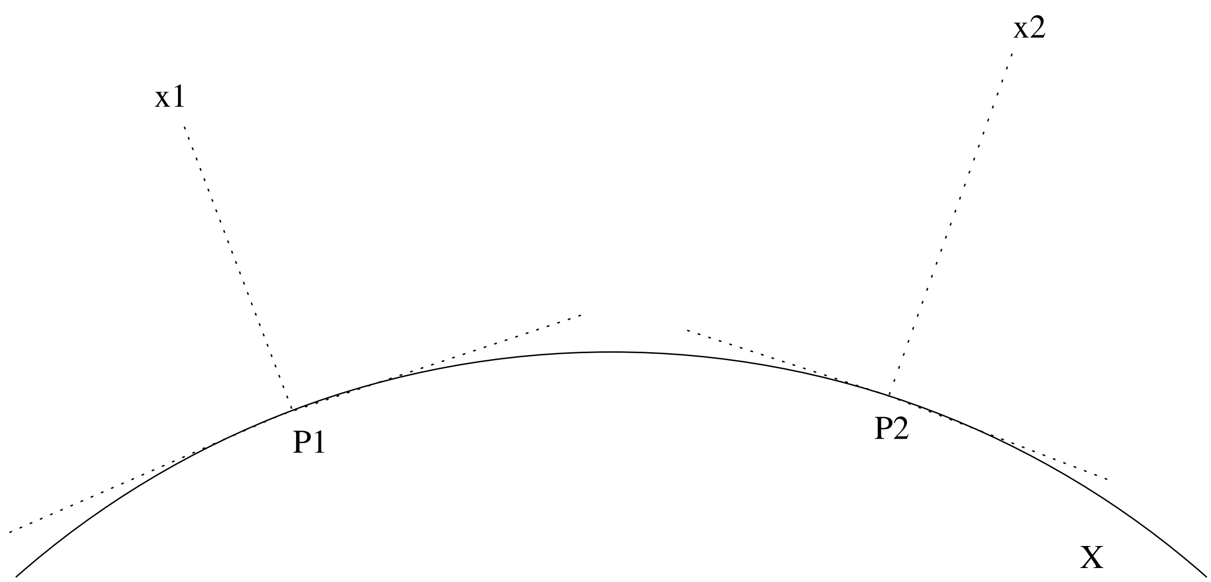

Figura 3.1: Temos que $||P_1 - P_2|| \leq ||x_1 - x_2||$, onde $P_1 = P_X(x_1)$ e $P_2 = P_X(x_2)$.

3.2 Condições do tipo AGP

Considere o problema

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & x \in \Omega, \end{array} \tag{3.2}$$

onde $\Omega = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0\}, f\colon \mathbb{R}^n \to \mathbb{R}, h\colon \mathbb{R}^n \to \mathbb{R}^m,\ g\colon \mathbb{R}^n \to \mathbb{R}^p$, todas com derivadas contínuas.

Vamos supor, por ora, que h é afim e g é convexa, assim Ω é convexo e fechado. Suponhamos que Ω é não vazio, assim, para cada $x \in \mathbb{R}^n$ podemos definir a projeção ortogonal de x em Ω, $P_\Omega(x)$.

O Lema a seguir se encontra em [45], mas o demonstramos aqui por completude.

Lema 3.3. *Seja* $x^* \in \Omega$, $\lambda \in \mathbb{R}^m, \mu_j \geq 0 \quad \forall j \in \mathtt{I}(x^*)$ *então*

$$P_\Omega(x^* + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j \in \mathtt{I}(x^*)} \mu_j \nabla g_j(x^*)) = x^*$$

Demonstração: Temos que $P_\Omega(x^* + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j\in \mathtt{I}(x^*)} \mu_j \nabla g_j(x^*))$ é solução do problema

$$\begin{array}{ll} \text{Minimizar} & ||x - (x^* + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j\in \mathtt{I}(x^*)} \mu_j \nabla g_j(x^*))||^2 \\ \text{Sujeito a} & x \in \Omega. \end{array}$$

Para valer KKT em $x \in \Omega$, devem existir $\lambda' \in \mathbb{R}^m, \mu' \geq 0$ tais que:

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

Para $\lambda' = \lambda$ e $\mu' = \mu$, temos que $x = x^*$ satisfaz KKT, assim, como o problema é convexo, temos que KKT é condição suficiente para ser minimizador global (vide [9]) e o resultado segue. □

Como motivação para a definição da condição de otimalidade AGP, mostremos as duas proposições a seguir, que permitem reescrever as equações KKT de outra maneira.

Proposição 3.4. *Se x^* é KKT para o problema (3.2) então $x^* = P_\Omega(x^* - \nabla f(x^*))$.*

Demonstração: Temos

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

$$\mu_j \geq 0 \quad \forall j \in \mathrm{I}(x^*)$$

$$h(x^*) = 0, g(x^*) \leq 0$$

$$\mathrm{I}(x^*) = \{j \in \{1, \ldots, p\} | g_j(x^*) = 0\}$$

Assim $(x^* - \nabla f(x^*)) - (x^* + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j \in \mathrm{I}(x^*)} \mu_j \nabla g_j(x^*)) = 0 \Rightarrow P_\Omega(x^* - \nabla f(x^*)) - P_\Omega(x^* + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j \in \mathrm{I}(x^*)} \mu_j \nabla g_j(x^*)) = 0 \Rightarrow P_\Omega(x^* - \nabla f(x^*)) - x^* = 0$, onde a última implicação decorre do Lema 3.3. $\square$

Proposição 3.5. *Se vale alguma condição de qualificação em Ω e $x^* = P_\Omega(x^* - \nabla f(x^*))$ então x^* é KKT para o problema (3.2).*

Demonstração: Como $x^* = P_\Omega(x^* - \nabla f(x^*))$, então x^* é solução de

$$\begin{array}{ll} \text{Minimizar} & ||x - (x^* - \nabla f(x^*))||^2 \\ \text{Sujeito a} & x \in \Omega, \end{array}$$

logo x^* satisfaz KKT para este problema, isto é,

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

ou seja, x^* é KKT para o problema (3.2). $\square$

A Figura 3.2 ilustra as Proposições acima.

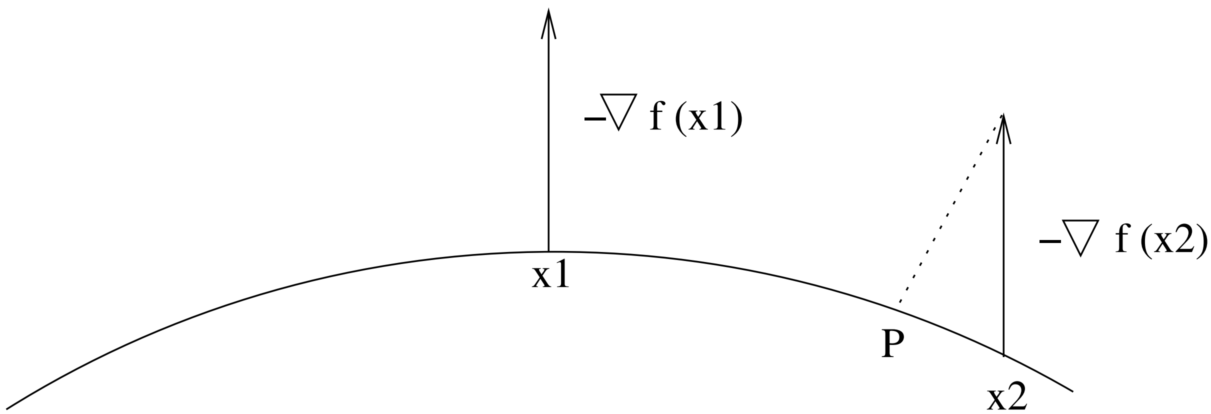

Figura 3.2: Temos que x_1 é minimizador, portanto $P_X(x_1 - \nabla f(x_1)) = x_1$. Já x_2 não é minimizador, logo $P_X(x_2 - \nabla f(x_2)) \neq x_2$.

É interessante observar que se não vale uma condição de qualificação em Ω, o resultado pode falhar. Considere o problema

$$\begin{array}{ll} \text{Minimizar} & x \\ \text{Sujeito a} & x^2 = 0, \end{array}$$

o minimizador $x^* = 0$ não é KKT, logo não satisfaz nenhuma condição de qualificação, mas observe que $P_\Omega(x^* - \nabla f(x^*)) = x^*$.

Vamos considerar o problema, em geral não convexo, a seguir

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & h(x) = 0 \\ & g(x) \leq 0, \end{array} \tag{3.3}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$ e $g : \mathbb{R}^n \to \mathbb{R}^p$ são funções com derivadas contínuas.

Baseados nos resultados acima, vamos definir uma condição de otimalidade para o problema (3.3).

Fixado $\gamma \in [-\infty, 0]$, considere o conjunto

$$\Omega_\gamma(x) = \left\{ z \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} g_i(x) + \nabla g_i(x)^{\mathtt{T}}(z-x) \leq 0 & \text{se } \gamma < g_i(x) < 0 \\ \nabla g_i(x)^{\mathtt{T}}(z-x) \leq 0 & \text{se } g_i(x) \geq 0 \\ \nabla h_i(x)^{\mathtt{T}}(z-x) = 0 & \end{array} \right\}.$$

Este conjunto pode ser interpretado como sendo a aproximação linear do conjunto dos pontos que não pioram a inviabilidade, isto é, o conjunto dos $z \in \mathbb{R}^n$ tal que $h(z) = h(x)$, $g_j(z) \leq 0$ se $\gamma < g_j(x) < 0$, $g_j(z) \leq g_j(x)$ se $g_j(x) \geq 0$. Quando $\gamma = 0$, temos que a primeira inequação não está presente na definição de Ω_γ.

Vamos definir $d(\Omega, z) = P_\Omega(z - \nabla f(z)) - z$, chamado de *gradiente projetado*. Existem muitas propriedades relacionadas ao gradiente projetado, principalmente quando f é convexa e Ω é convexo, veja [38, 48, 47]. Nestas referências podemos encontrar algoritmos para otimização convexa utilizando idéias de penalidade e projeção, juntamente com limitantes para o número de iterações necessárias para encontrar uma ε-solução.

A seguir, definimos a condição AGP (*Approximate Gradient Projection*), introduzida em [45], que surgiu do estudo de algoritmos de restauração inexata [44, 41, 43].

Definição 3.6 (AGP)**.** *Se x é um ponto viável para o problema (3.3), dizemos que x satisfaz AGP se existe $\{x^k\} \subset \mathbb{R}^n, x^k \to x$ tal que $d(\Omega_\gamma(x^k), x^k) \to 0$.*

Esta condição tem uma relação próxima com as equações KKT, como mostramos abaixo.

Proposição 3.7. *Se x é viável para (3.3), então $d(\Omega_\gamma(x), x) = 0$ se, e somente se x é KKT.*

Demonstração: Considere o problema

$$\begin{array}{ll} \text{Minimizar} & \frac{1}{2}||y - (x - \nabla f(x))||^2 \\ \text{Sujeito a} & \nabla h_i(x)^{\mathtt{T}}(y-x) = 0, \forall i = 1, \ldots, m \\ & \nabla g_j(x)^{\mathtt{T}}(y-x) \leq 0, \forall j \in \mathtt{I}(x) \\ & g_j(x) + \nabla g_j(x)^{\mathtt{T}}(y-x) \leq 0, \text{ se } \gamma < g_j(x) < 0, \end{array} \tag{3.4}$$

sendo que y viável é KKT para este problema se existem $\lambda \in \mathbb{R}^m, \mu_j \geq 0, \mu'_j \geq 0$ tais que

$$y - x + \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \sum_{j \in \mathtt{I}(x)} \mu_j \nabla g_j(x) + \sum_{\gamma < g_j(x) < 0} \mu'_j \nabla g_j(x) = 0, \tag{3.5}$$
$$\mu_j \nabla g_j(x)^{\mathtt{T}}(y - x) = 0, \mu'_j(g_j(x) + \nabla g_j(x)^{\mathtt{T}}(y - x)) = 0.$$

($\Rightarrow$) Como $d(\Omega_\gamma(x), x) = 0$, temos que $y = x$ é solução do problema (3.4), e pela linearidade das restrições, valem as equações KKT para esse problema. Substituindo $y = x$ nas equações (3.5) temos que $\mu'_j = 0$ e x é KKT para o problema (3.3).

($\Leftarrow$) Se x é KKT para o problema (3.3), com multiplicadores λ e μ, temos pelas equações (3.5) que $y = x$ satisfaz KKT para o problema (3.4) com os mesmos multiplicadores, basta definir $\mu' = 0$. Como o problema é convexo, temos que $y = x$ é solução, e portanto $d(\Omega_\gamma(x), x) = 0$. $\square$

Podemos interpretar $d(\Omega, x)$ de outra maneira. Da definição de projeção, temos que $P_\Omega(x - \nabla f(x))$ é a solução do problema

$$\begin{array}{ll} \text{Minimizar} & ||z - (x - \nabla f(x))||^2 \\ \text{Sujeito a} & z \in \Omega. \end{array}$$

Fazendo a mudança de variável $d = z - x$ temos

$$\begin{array}{ll} \text{Minimizar} & ||d + \nabla f(x)||^2 \\ \text{Sujeito a} & x + d \in \Omega. \end{array} \tag{3.6}$$

Como pode ser facilmente verificado, o conjunto $-x + \Omega = \{d \in \mathbb{R}^n | d = -x + z, z \in \Omega\}$ é um convexo fechado não vazio, quando o mesmo ocorre com Ω, assim, se d^* denota a solução de (3.6), temos $d^* = P_{-x+\Omega}(-\nabla f(x)) = P_\Omega(x - \nabla f(x)) - x = d(\Omega, x)$.

Às vezes pode ser útil reescrever o conjunto $\Omega_\gamma(x)$ da seguinte maneira

$$\Omega_\gamma(x) = \left\{ z \in \mathbb{R}^n \;\middle|\; \begin{array}{c} g_i(x)^- + \nabla g_i(x)^{\mathtt{T}}(z - x) \leq 0 \quad \text{se } \gamma < g_i(x) \\ \nabla h_i(x)^{\mathtt{T}}(z - x) = 0 \end{array} \right\},$$

onde $g_i(x)^- = \min\{0, g_i(x)\}$. Ou ainda, podemos denotar por $-x + \Omega_\gamma(x)$ o conjunto dos pontos da forma $-x + z$, com $z \in \Omega_\gamma(x)$, ou seja

$$-x + \Omega_\gamma(x) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} g_i(x)^- + \nabla g_i(x)^{\mathtt{T}} d \leq 0 & \text{se } \gamma < g_i(x) \\ \nabla h_i(x)^{\mathtt{T}} d = 0 & \end{array} \right\}.$$

Em [45], os autores mostram que AGP é de fato uma condição de otimalidade, além disso, mostram que AGP é mais forte que a condição KKT ou não-MF. Recentemente, em [24], os autores mostraram que o resultado é mais forte e vale AGP $\Rightarrow$ KKT ou não-CPLD. Incluímos aqui a demonstração por completude. Na demonstração apresentada em [24] os autores utilizam apenas desigualdades, trocando as eventuais igualdades $h(x) = 0$ por desigualdades $h(x) \leq 0$ e $-h(x) \leq 0$, mostrando que a CPLD se mantém para o novo problema só com desigualdades, enquanto a nossa demonstração considera o problema em sua forma original.

Proposição 3.8. *AGP $\Rightarrow$ KKT ou não-CPLD*

Demonstração: Seja x^* viável satisfazendo AGP, logo $h(x^*) = 0, g(x^*) \leq 0$ e existe $x^k \to x^*$ tal que $d(\Omega_\gamma(x^k), x^k) \to 0$. Defina $y^k = P_{\Omega_\gamma}(x^k)(x^k - \nabla f(x^k))$, logo y^k resolve o problema

$$\begin{array}{ll} \text{Minimizar} & \frac{1}{2}||y - (x^k - \nabla f(x^k)||^2 \\ \text{Sujeito a} & y \in \Omega_\gamma(x^k) \end{array},$$

como $\Omega_\gamma(x^k)$ é um politopo então existem $\lambda^k \in \mathbb{R}^m, \mu^k \geq 0$ tais que

$$y^k - (x^k - \nabla f(x^k)) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) = 0, y^k \in \Omega_\gamma(x^k) \tag{3.7}$$

com $\mu_j^k(g_j(x^k) + \nabla g_j(x^k)^{\mathtt{T}}(y^k - x^k)) = 0$ se $\gamma < g_j(x^k) < 0$,
$\mu_j^k(\nabla g_j(x^k)^{\mathtt{T}}(y^k - x^k)) = 0$ se $g_j(x^k) \geq 0$,
$\mu_j^k = 0$ caso contrário.

Se $g_j(x^*) < 0$, então existe $\varepsilon > 0$ tal que $g_j(x^k) < -\varepsilon$ para todo k suficientemente grande, assim, como $y^k - x^k \to 0$, temos $g_j(x^k) + \nabla g_j(x^k)^{\mathtt{T}}(y^k - x^k) < 0 \Rightarrow \mu_i^k = 0$ para

k suficientemente grande. Logo, de (3.7) temos:

$$(y^k - x^k) + \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j | g_j(x^*)=0} \mu_j^k \nabla g_j(x^k) = 0. \tag{3.8}$$

Pelo Lema de Carathéodory, existem $\mathrm{I}_k \subset \{1, \ldots, m\}, \mathrm{J}_k \subset \mathrm{I}(x^*),\ \lambda_i' \in \mathbb{R}^m, \mu_j' \geq 0$ (vamos tomar uma subsequência tal que $\mathrm{I}_k = \mathrm{I}$ e $\mathrm{J}_k = \mathrm{J}$) tais que

$$\left(\left\{ \nabla h_i(x^k) \right\}_{i \in \mathrm{I}}, \left\{ \nabla g_j(x^k) \right\}_{j \in \mathrm{J}} \right) \text{ é PLI,}$$

$$(y^k - x^k) + \nabla f(x^k) + \sum_{i \in \mathrm{I}} \lambda_i'^k \nabla h_i(x^k) + \sum_{j \in \mathrm{J}} \mu_j'^k \nabla g_j(x^k) = 0. \tag{3.9}$$

Se $\{(\lambda'^k, \mu'^k)\}$ é limitada, basta tomar uma subsequência convergente, assim, como $x^k - y^k \to 0$, temos que x^* é KKT.

Se $\{(\lambda'^k, \mu'^k)\}$ é ilimitada, defina $M_k = \max\{||\lambda'^k||_\infty, ||\mu'^k||_\infty\}$ em uma subsequência com $M_k \to +\infty$, assim $||(\frac{\lambda'^k}{M_k}, \frac{\mu'^k}{M_k})||_\infty = 1$ e considere uma subsequência convergindo para $(\hat{\lambda}, \hat{\mu})$ com $||(\hat{\lambda}, \hat{\mu})||_\infty = 1, \hat{\mu} \geq 0$. Logo, dividindo (3.9) por M_k e tomando o limite para essa subsequência temos $\sum_{i \in \mathrm{I}} \hat{\lambda}_i \nabla h_i(x^*) + \sum_{j \in \mathrm{J}} \hat{\mu}_j \nabla g_j(x^*) = 0$,
e $\left(\{\nabla h_i(x^*)\}_{i \in \mathrm{I}}, \{\nabla g_j(x^*)\}_{j \in \mathrm{J}} \right)$ é PLD, logo não vale CPLD em x^*. □

O parâmetro γ tem finalidade apenas prática, já que é mais fácil verificar a condição AGP, quando γ está próximo de zero. De fato, as condições obtidas com parâmetros diferentes são equivalentes, como mostra a proposição abaixo.

Proposição 3.9. *Seja $\gamma \in [-\infty, 0)$ tal que vale AGP, então vale AGP para todo $\gamma' \in [-\infty, 0)$.*

Demonstração: Veja [45]. □

A seguir, vamos mostrar que não podemos fortalecer o resultado da Proposição 3.8 trocando CPLD por QN.

Considere o problema

$$\begin{array}{ll} \text{Minimizar} & x_2 \\ \text{Sujeito a} & x_1 x_2 = 0 \\ & x_1 \geq 0 \\ & x_2 \geq 0. \end{array}$$

Vamos mostrar que $x^* = (0,1)$ é AGP. Para isso, definimos a sequência $x^k = (\frac{-1}{k}, 1)$ e vamos calcular $P_{\Omega_\gamma(x^k)}(x^k - (0,1))$=$P_{\Omega_\gamma(x^k)}((\frac{-1}{k}, 0))$ através da solução do problema

$$\begin{array}{ll} \text{Minimizar} & ||z - (\frac{-1}{k}, 0)||^2 \\ \text{Sujeito a} & z \in \Omega_\gamma(x^k) \end{array},$$

onde

$$\Omega_\gamma(x^k) = \left\{ z \in \mathbb{R}^2 \;\middle|\; \begin{array}{c} z_2 \geq 0 \\ z_1 \geq x_1^k \\ \nabla x_2^k (z_1 - x_1^k) + x_1^k (z_2 - x_2^k) = 0. \end{array} \right\}$$

Observe que $z = x^k$ satisfaz a condição KKT para esse problema, logo é a solução. Assim, $d(\Omega_\gamma(x^k), x^k) = 0$ e x^* é AGP. É fácil ver que x^* não é KKT e vale a quase-normalidade (QN), isso mostra que AGP $\not\Rightarrow$ KKT ou não-QN.

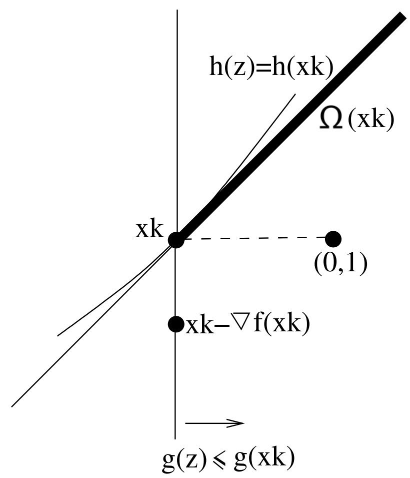

Figura 3.3: Temos $f(x) = x_2$, $g(x) = -x_1$, $h(x) = x_1x_2$. O conjunto $\Omega_\gamma(x^k)$ está destacado em negrito. Podemos ver que a projeção de $x^k - \nabla f(x^k)$ é o próprio ponto x^k, de modo que $P_{\Omega_\gamma(x^k)}(x^k - \nabla f(x^k)) - x^k = 0$.

Muitas vezes os algoritmos geram pontos que são sempre viáveis com respeito a um subconjunto K das restrições, que pode ser uma caixa, um politopo ou um conjunto convexo qualquer. Neste caso, vamos definir uma condição do tipo AGP, mas exigindo que a sequência esteja em K, de modo que ela seja possível de ser gerada por tais algoritmos. Além disso, podemos manter as restrições convexas como são, sem a necessidade de aproximações lineares.

Vamos então separar as restrições convexas do conjunto viável, isto é, vamos considerar o problema (3.3) onde as restrições são dadas por $x \in X = \Omega \cap K$, com $\Omega = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0\}, K = \{x \in \mathbb{R}^n | \bar{h}(x) = 0, \bar{g}(x) \leq 0\}$, onde

$$f\colon \mathbb{R}^n \to \mathbb{R}, h\colon \mathbb{R}^n \to \mathbb{R}^m, g\colon \mathbb{R}^n \to \mathbb{R}^p, \bar{h}\colon \mathbb{R}^n \to \mathbb{R}^{\bar{m}}, \bar{g}\colon \mathbb{R}^n \to \mathbb{R}^{\bar{p}},$$

todas com derivadas contínuas, sendo $\bar{h}$ afim e $\bar{g}$ convexa. Definimos $\Omega_\gamma(x)$ como anteriormente, levando em conta apenas as restrições definidas por Ω.

Definição 3.10 (C-AGP)**.** *Se x é um ponto viável para o problema (3.3), dizemos que x satisfaz C-AGP se existe $\{x^k\} \subset K, x^k \to x$ tal que $d(\Omega_\gamma(x^k) \cap K, x^k) \to 0$, onde K é convexo e satisfaz alguma condição de qualificação.*

A idéia de considerar a condição AGP restringindo a sequência x^k a um subconjunto das restrições apareceu pela primeira vez em [6], onde pensava-se que esta condição modificada era a mesma que AGP. Vamos mostrar que, na verdade, as condições são diferentes. Voltando ao exemplo anterior, vamos definir $\Omega = \{x \in \mathbb{R}^2 | x_1 x_2 = 0\}, K = \{x \in \mathbb{R}^2 | x_1 \geq 0, x_2 \geq 0\}$ e vamos mostrar que não existe $\{x^k\} \subset K, x^k \to x^* = (0,1)$ tal que $d(\Omega_\gamma(x^k) \cap K, x^k) \to 0$.

Seja $x^k \to x^*$ e $z^k = P_{\Omega_\gamma(x^k)\cap K}(x^k - (0,1))$, onde

$$\Omega_\gamma(x^k) = \{z \in \mathbb{R}^2 | x_2^k(z_1 - x_1^k) + x_1^k(z_2 - x_2^k) = 0\}.$$

Pela linearidade do conjunto viável, z^k satisfaz as equações KKT:

$$\begin{pmatrix} z_1^k - x_1^k \\ z_2^k - (x_2^k - 1) \end{pmatrix} + \lambda^k \begin{pmatrix} x_2^k \\ x_1^k \end{pmatrix} + \mu_1^k \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \mu_2^k \begin{pmatrix} 0 \\ -1 \end{pmatrix} = 0,$$

com $\mu_1 z_1^k = 0, \mu_2 z_2^k = 0, z_1^k \geq 0, z_2^k \geq 0, x_2^k(z_1^k - x_1^k) + x_1^k(z_2^k - x_2^k) = 0$. Se $z_2^k = 0$, como $x_2^k \to 1$ então $d(\Omega_\gamma(x^k) \cap K, x^k) \not\to 0$. Se $z_1^k = 0$ temos $z_2^k = x_2^k + \frac{x_2^k}{x_1^k}$, logo $d(\Omega_\gamma(x^k) \cap K, x^k) \not\to 0$. Se $z_1^k > 0, z_2^k > 0$, temos $z_1^k = x_1^k + \frac{x_1^k x_2^k}{(x_1^k)^2 + (x_2^k)^2}, z_2^k = x_2^k - \frac{(x_2^k)^2}{(x_1^k)^2 + (x_2^k)^2}$, logo $d(\Omega_\gamma(x^k) \cap K, x^k) \not\to 0$. Assim x^* não é C-AGP, o que mostra o resultado abaixo:

Proposição 3.11. *AGP $\not\Rightarrow$ C-AGP*

Observe que para o exemplo acima, o ponto $x^* = (0,1)$ não é solução, embora seja AGP. Isso mostra que, pelo menos para este problema, seria mais adequado aplicarmos um algoritmo cujos pontos limites satisfaçam C-AGP.

A seguir vamos mostrar que todo minimizador satisfaz C-AGP, isto é, C-AGP é de fato uma condição de otimalidade.

Proposição 3.12. *C-AGP é uma condição de otimalidade*

Demonstração: Seja x^* minimizador local de

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & x \in X = \Omega \cap K, \end{array}$$

com $K = \{x \in \mathbb{R}^n | \bar{h}(x) = 0, \bar{g}(x) \leq 0\}$ convexo e $\Omega = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0\}$. Seja $\varepsilon > 0$ tal que x^* é minimizador global de

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & x \in \Omega \cap K \\ & \|x - x^*\| \leq \varepsilon. \end{array}$$

Logo x^* é o único minimizador global de

$$\begin{array}{ll} \text{Minimizar} & f(x) + \frac{1}{2}\|x - x^*\|^2 \\ \text{Sujeito a} & x \in \Omega \cap K \\ & \|x - x^*\| \leq \varepsilon. \end{array}$$

Para cada $\rho > 0$, defina $\Phi_\rho(x) = f(x) + \frac{1}{2}\|x - x^*\|^2 + \frac{\rho}{2}(\|h(x)\|^2 + \|g^+(x)\|^2)$. Como $K \cap \{x | \|x - x^*\| \leq \varepsilon\}$ é compacto não vazio e $\Phi_\rho(\cdot)$ é contínua, temos que o problema

$$\begin{array}{ll} \text{Minimizar} & \Phi_\rho(x) \\ \text{Sujeito a} & x \in K \\ & \|x - x^*\| \leq \varepsilon \end{array} \tag{3.10}$$

admite solução x_ρ para cada $\rho > 0$. Fazendo $\rho \to +\infty$, obtemos uma sequência de soluções x^k, associada a $\rho = \rho^k$, no compacto $K \cap \{x | \|x - x^*\| \leq \varepsilon\}$, e usando o teorema da penalidade externa, todo ponto limite de $\{x^k\}$ é minimizador global de (3.10). Como (3.10) possui um único minimizador global e a sequência $\{x^k\}$ possui pelo menos um ponto limite, então $x^k \to x^*$. Assim, para k suficientemente grande, x^k é solução local de

$$\begin{array}{ll} \text{Minimizar} & \Phi_{\rho^k}(x) \\ \text{Sujeito a} & x \in K. \end{array}$$

Como K satisfaz uma condição de qualificação, então x^k satisfaz as equações KKT

$$\nabla \Phi_{\rho^k}(x^k) + \sum_{i=1}^{\bar{m}} \bar{\lambda}_i \nabla \bar{h}_i(x^k) + \sum_{j=1}^{\bar{p}} \bar{\mu}_j \nabla \bar{g}_j(x^k), \bar{\mu}_j \bar{g}_j(x^k) = 0,$$

para algum $\bar{\lambda} \in \mathbb{R}^{\bar{m}}, \bar{\mu} \geq 0$, ou seja

$$\nabla f(x^k) + \rho_k(\sum_{i=1}^{m} h_i(x^k)\nabla h_i(x^k) + \sum_{j|g_j(x^k)>0} g_j(x^k)\nabla g_j(x^k)) + (x^k - x^*) +$$
$$+ \sum_{i=1}^{\bar{m}} \bar{\lambda}_i \nabla \bar{h}_i(x^k) + \sum_{j|g_j(x^k)=0} \bar{\mu}_j \nabla \bar{g}_j(x^k) = 0.$$

Logo temos

$$(x^k - \nabla f(x^k)) - (x^k + \sum_{i=1}^{m} \rho_k h_i(x^k)\nabla h_i(x^k) + \sum_{j|g_j(x^k)>0} \rho_k g_j(x^k)\nabla g_j(x^k) +$$
$$+ \sum_{i=1}^{\bar{m}} \bar{\lambda}_i \nabla \bar{h}_i(x^k) + \sum_{j|g_j(x^k)=0} \bar{\mu}_j \nabla \bar{g}_j(x^k)) = x^k - x^*.$$

Como a projeção é não expansiva (Lema 3.2) temos

$$||P_{\Omega_\gamma(x^k)\cap K}(x^k - \nabla f(x^k)) - P_{\Omega_\gamma(x^k)\cap K}(x^k + \sum_{i=1}^{m} \rho_k h_i(x^k)\nabla h_i(x^k) +$$
$$+ \sum_{j|g_j(x^k)>0} \rho_k g_j(x^k)\nabla g_j(x^k) + \sum_{i=1}^{\bar{m}} \bar{\lambda}_i \nabla \bar{h}_i(x^k) + \sum_{j|g_j(x^k)=0} \bar{\mu}_j \nabla \bar{g}_j(x^k))|| \leq$$
$$\leq ||(x^k - \nabla f(x^k)) - (x^k + \sum_{i=1}^{m} \rho_k h_i(x^k)\nabla h_i(x^k) +$$
$$+ \sum_{j|g_j(x^k)>0} \rho_k g_j(x^k)\nabla g_j(x^k) + \sum_{i=1}^{\bar{m}} \bar{\lambda}_i \nabla \bar{h}_i(x^k) + \sum_{j|g_j(x^k)=0} \bar{\mu}_j \nabla \bar{g}_j(x^k))|| =$$
$$= ||x^k - x^*||,$$

e pelo Lema 3.3

$$x^k = P_{\Omega_\gamma(x^k)\cap K}(x^k + \sum_{i=1}^{m} \rho_k h_i(x^k)\nabla h_i(x^k) + \sum_{j|g_j(x^k)>0} \rho_k g_j(x^k)\nabla g_j(x^k) +$$
$$+ \sum_{i=1}^{\bar{m}} \bar{\lambda}_i \nabla \bar{h}_i(x^k) + \sum_{j|g_j(x^k)=0} \bar{\mu}_j \nabla \bar{g}_j(x^k)),$$

assim,

$$||P_{\Omega_\gamma(x^k)\cap K}(x^k - \nabla f(x^k)) - x^k|| \le ||x^k - x^*|| \to 0.$$

Como $\{x^k\} \subset K$ e $x^k \to x^*$, temos que x^* satisfaz C-AGP. $\square$

Agora vamos mostrar que a condição C-AGP é forte, pois implica KKT sob MF.

Proposição 3.13. *C-AGP $\Rightarrow$ KKT ou não-MF*

Demonstração: Seja x^* com $h(x^*) = 0, g(x^*) \le 0, x^* \in K$ tal que existe $\{x^k\} \subset K, x^k \to x^*$ com $d(\Omega_\gamma(x^k) \cap K, x^k) \to 0$ e suponha que x^* satisfaça MF com respeito a $x \in X \cap K$.

Defina $y^k = P_{\Omega_\gamma(x^k)\cap K}(x^k - \nabla f(x^k))$, logo y^k resolve

$$\begin{array}{ll} \text{Minimizar} & \frac{1}{2}||y - (x^k - \nabla f(x^k))||^2 \\ \text{Sujeito a} & y \in \Omega_\gamma(x^k) \cap K. \end{array}$$

Vamos mostrar que y^k satisfaz a condição KKT, quando k é suficientemente grande, para isso, basta mostrar que y^k satisfaz MF com respeito às restrições $\Omega_\gamma(x^k) \cap K$. Procedemos por contradição.

Suponha que

$$\left(\left\{\nabla h_i(x^k)\right\}_{i=1}^m \cup \left\{\nabla \bar{h}_i(y^k)\right\}_{i=1}^{\bar{m}}, \left\{\nabla g_j(x^k)\right\}_{j\in \mathtt{I}'(y^k)} \cup \left\{\nabla \bar{g}_j(y^k)\right\}_{j\in \bar{\mathtt{I}}(y^k)}\right) \text{ é PLD,}$$

onde $\bar{\mathtt{I}}(y) = \{j \in \{1, \ldots, \bar{p}\} | \bar{g}_j(y) = 0\}$ e $\mathtt{I}'(y)$ é o conjunto dos índices das restrições de desigualdade ativas com respeito às restrições

$$\Omega_\gamma(x^k) = \left\{ y \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} g_i(x^k) + \nabla g_i(x^k)^{\mathtt{T}}(y - x^k) \le 0 & \text{se } \gamma < g_i(x^k) < 0 \\ \nabla g_i(x^k)^{\mathtt{T}}(y - x^k) \le 0 & \text{se } g_i(x^k) \ge 0 \\ \nabla h_i(x^k)^{\mathtt{T}}(y - x^k) = 0 & \end{array} \right\}.$$

Logo existem $\lambda_i'^k, \bar{\lambda}_i'^k, \mu_j'^k \geq 0, \bar{\mu}_j'^k \geq 0$ não todos nulos tais que

$$\sum_{i=1}^{m} \lambda_i'^k \nabla h_i(x^k) + \sum_{i=1}^{\bar{m}} \bar{\lambda}_i'^k \nabla \bar{h}_i(y^k) + \sum_{j \in \mathtt{I}'(y^k)} \mu_j'^k \nabla g_j(x^k) + \sum_{j \in \bar{\mathtt{I}}(y^k)} \bar{\mu}_j'^k \nabla \bar{g}_j(y^k) = 0.$$

Vamos mostrar que $\mathtt{I}'(y^k) \subset \mathtt{I}(x^*)$ para todo k suficientemente grande. Seja $i \notin \mathtt{I}(x^*)$, logo $g_i(x^*) < 0$, assim, para k suficientemente grande temos que existe $\varepsilon > 0$ tal que $g_i(x^k) < -\varepsilon$. Se $g_i(x^k) \leq \gamma$ temos $i \notin \mathtt{I}'(y^k)$, já que este índice não faz parte das restrições do problema. Caso contrário temos que k satisfaz $g_i(x^k) + \nabla g_i(x^k)^{\mathtt{T}}(y^k - x^k) \leq 0$ e como $y^k - x^k \to 0$ temos que para k suficientemente grande vale $g_i(x^k) + \nabla g_i(x^k)^{\mathtt{T}}(y^k - x^k) < 0$, logo $i \notin \mathtt{I}'(y^k)$.

É fácil ver que $\bar{\mathtt{I}}(y^k) \subset \bar{\mathtt{I}}(x^*)$, pois se $j \notin \bar{\mathtt{I}}(x^*)$, isto é, $\bar{g}_j(x^*) < 0$, temos $\bar{g}_j(y^k) < 0$, pois $y^k \to x^*$ já que $x^k - y^k \to 0$ e $x^k \to x^*$.

Assim, incluindo multiplicadores nulos, temos

$$\sum_{i=1}^{m} \lambda_i'^k \nabla h_i(x^k) + \sum_{i=1}^{\bar{m}} \bar{\lambda}_i'^k \nabla \bar{h}_i(y^k) + \sum_{j \in \mathtt{I}(x^*)} \mu_j'^k \nabla g_j(x^k) + \sum_{j \in \bar{\mathtt{I}}(x^*)} \bar{\mu}_j'^k \nabla \bar{g}_j(y^k) = 0.$$

Dividindo por $M_k' = \max\{|\lambda_i'^k|, |\bar{\lambda}_i'^k|, \mu_j'^k, \bar{\mu}_j'^k, \forall i, j\} > 0$ e tomando o limite para uma subsequência convergindo para $\{\lambda', \bar{\lambda}', \mu' \geq 0, \bar{\mu}' \geq 0\} \neq 0$, temos

$$\sum_{i=1}^{m} \lambda_i' \nabla h_i(x^*) + \sum_{i=1}^{\bar{m}} \bar{\lambda}_i' \nabla \bar{h}_i(x^*) + \sum_{j \in \mathtt{I}(x^*)} \mu_j' \nabla g_j(x^*) + \sum_{j \in \bar{\mathtt{I}}(x^*)} \bar{\mu}_j' \nabla \bar{g}_j(x^*) = 0,$$

o que entra em contradição com o fato que x^* satisfaz MF para o problema original, assim, y^k satisfaz MF para o problema da projeção e portanto a condição KKT, isto é, existem $\lambda \in \mathbb{R}^m, \mu \geq 0, \bar{\lambda} \in \mathbb{R}^{\bar{m}}, \bar{\mu} \geq 0$, tais que

$$\nabla f(x^k) + (y^k - x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{i=1}^{\bar{m}} \bar{\lambda}_i^k \nabla \bar{h}_i(y^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) + \sum_{j=1}^{\bar{p}} \bar{\mu}_j^k \nabla \bar{g}_j(y^k) = 0$$

$$\mu_j^k(g_j(x^k) + \nabla g_j(x^k)^{\mathtt{T}}(y^k - x^k)) = 0 \text{ se } \gamma < g_j(x^k) < 0$$

$$\mu_j^k(\nabla g_j(x^k)^{\mathtt{T}}(y^k - x^k)) = 0 \text{ se } g_j(x^k) \geq 0$$

$$\mu_j^k = 0 \text{ caso contrário}$$

$$\bar{\mu}_j^k \bar{g}_j(y^k) = 0.$$

Como $\mathtt{I}'(y^k) \subset \mathtt{I}(x^*)$ e $\bar{\mathtt{I}}(y^k) \subset \bar{\mathtt{I}}(x^*)$, podemos incluir multiplicadores nulos e escrever

$$\nabla f(x^k)+(y^k-x^k)+\sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k)+\sum_{i=1}^{\bar{m}} \bar{\lambda}_i^k \nabla \bar{h}_i(y^k)+\sum_{j\in \mathtt{I}(x^*)} \mu_j^k \nabla g_j(x^k)+\sum_{j\in \bar{\mathtt{I}}(x^*)} \bar{\mu}_j^k \nabla \bar{g}_j(y^k) = 0. \tag{3.11}$$

Agora vamos analisar dois casos:

Se $||(\lambda^k, \mu^k, \bar{\lambda}^k, \bar{\mu}^k)||_\infty$ é limitada, então basta tomar uma subsequência convergente, assim, tomando o limite para esta subsequência em (3.11) temos que x^* é KKT para o problema original.

Se $M_k = ||(\lambda^k, \mu^k, \bar{\lambda}^k, \bar{\mu}^k)||_\infty$ é ilimitada, considere uma subsequência tal que $M_k \to \infty$ e $\frac{(\lambda^k,\mu^k,\bar{\lambda}^k,\bar{\mu}^k)}{M_k} \to (\lambda, \mu, \bar{\lambda}, \bar{\mu}) \neq 0, \mu \geq 0, \bar{\mu} \geq 0$, assim, dividindo por M_k em (3.11) e tomando o limite nesta subsequência temos

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{i=1}^{\bar{m}} \bar{\lambda}_i \nabla \bar{h}_i(x^*) + \sum_{j\in \mathtt{I}(x^*)} \mu_j \nabla g_j(x^*) + \sum_{j\in \bar{\mathtt{I}}(x^*)} \bar{\mu}_j \nabla \bar{g}_j(x^*) = 0,$$

contradizendo o fato que x^* satisfaz MF com respeito ao problema original. □

Vamos mostrar a seguir que não é possível substituir a condição MF na Proposição 3.13 por CPLD.

Proposição 3.14. *C-AGP $\not\Rightarrow$ KKT ou não-CPLD*

Demonstração: Considere o problema

$$\begin{array}{ll} \text{Minimizar} & x_1 \\ \text{Sujeito a} & -x_1^2 - x_2 \leq 0 \quad (x \in \Omega) \\ & x_1^2 + x_2 \leq 0 \quad (x \in K) \end{array}$$

e considere o ponto $x^* = (0, 0)$.

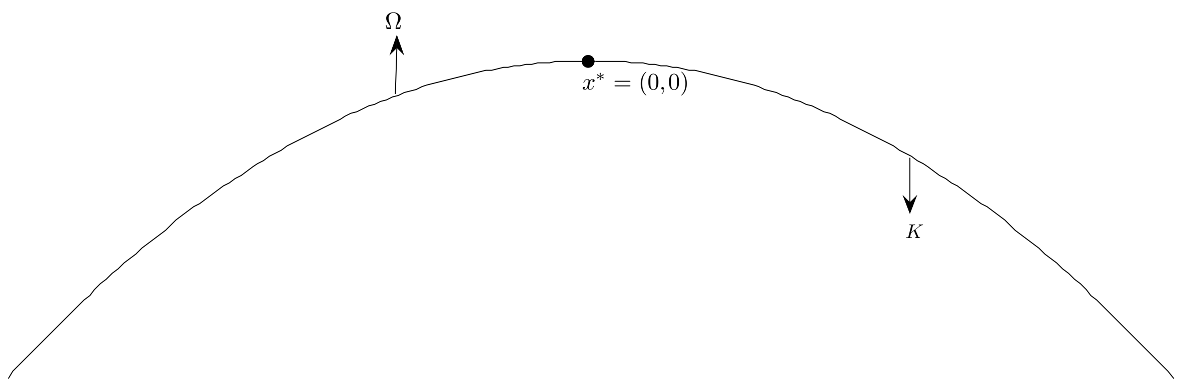

Figura 3.4: Temos que Ω é a região acima da parábola e K é a região abaixo da parábola, assim, o conjunto viável $\Omega \cap K$, é formado pelos pontos sobre a parábola.

Temos que x^* não é KKT, pois a equação

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \mu_1 \begin{pmatrix} 0 \\ -1 \end{pmatrix} + \mu_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0,$$

não é satisfeita para nenhum $\mu_1, \mu_2 \geq 0$.

Vamos mostrar que o problema satisfaz a CPLD.

Sejam $g_1(x) = -x_1^2 - x_2$ e $g_2(x) = x_1^2 + x_2$, logo $\nabla g_1(x) = (-2x_1, -1)$ e $\nabla g_2(x) = (2x_1, 1)$. Assim, $\{\nabla g_1(x^*)\} = \{(0, -1)\}$ e $\{\nabla g_2(x^*)\} = \{(0, 1)\}$ são conjuntos PLI.

Mas, $\{\nabla g_1(x^*), \nabla g_2(x^*)\} = \{(0, -1), (0, 1)\}$ é PLD pois $(0, -1) + (0, 1) = 0$ e em qualquer vizinhança de x^* temos $\{\nabla g_1(x), \nabla g_2(x)\} = \{(-2x, -1), (2x, 1)\}$ é PLD pois $(-2x, -1) + (2x, 1) = 0$.

Vamos definir $\Omega = \{x \in \mathbb{R}^2 | - x_1^2 - x_2 \leq 0\}$ e o convexo $K = \{x \in \mathbb{R}^2 | x_1^2 + x_2 \leq 0\}$ e vamos mostrar que o problema satisfaz C-AGP. Para isso, vamos definir a sequência constante $x^k = x^* \in K$. Temos

$$\Omega_\gamma(x^k) = \{x \in \mathbb{R}^2 | \nabla^{\mathtt{T}} g_2(x^k)(x - x^k) \leq 0\} = \{x \in \mathbb{R}^2 | (0, -1)^{\mathtt{T}} x \leq 0\} = \{x \in \mathbb{R}^2 | x_2 \geq 0\},$$

assim

$$\Omega_\gamma(x^k) \cap K = \{x \in \mathbb{R}^2 | x_2 \geq 0\} \cap \{x \in \mathbb{R}^2 | x_1^2 + x_2 \leq 0\} = \{(0, 0)\}.$$

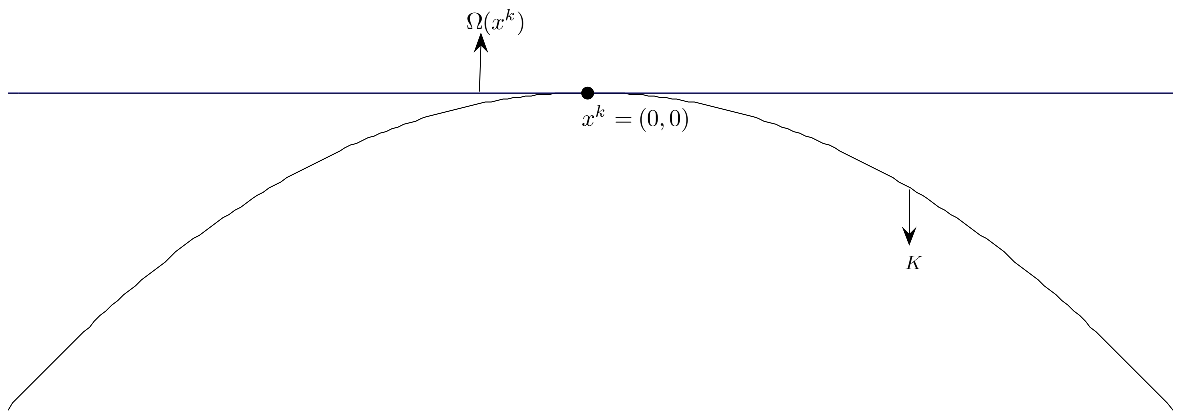

Figura 3.5: Temos $\Omega_\gamma(x^k) \cap K = \{(0,0)\}$.

Portanto, $P_{\Omega_\gamma(x^k)\cap K}(x^k - \nabla f(x^k)) - x^k = (0,0) - x^k = (0,0)$ e vale C-AGP. $\square$

Corolário 3.15. *C-AGP $\not\Rightarrow$ AGP*

Demonstração: É consequencia imediata das Proposições (3.8) e (3.14). $\square$

A Proposição 3.14 indica que C-AGP ainda não é o que procuramos. Então, ao invés de separarmos as restrições convexas do conjunto viável, vamos separar as restrições lineares, isto é, vamos considerar o problema (3.3) com restrições $x \in X = \Omega \cap K$, com $K = \{x \in \mathbb{R}^n | \bar{h}(x) = 0, \bar{g}(x) \leq 0\}$, sendo $\bar{h}, \bar{g}$ funções afins.

Definição 3.16 (L-AGP)**.** *Se x é um ponto viável para o problema (3.3), dizemos que x satisfaz L-AGP se existe $\{x^k\} \subset K, x^k \to x$ tal que $d(\Omega_\gamma(x^k) \cap K, x^k) \to 0$, onde K é um politopo.*

Proposição 3.17. *L-AGP é uma condição de otimalidade.*

Demonstração: Como todo politopo é convexo, o resultado vale pela Proposição (3.12). □

Vamos mostrar que L-AGP é uma condição de otimalidade mais forte que a AGP usual.

Proposição 3.18. *L-AGP $\Rightarrow$ AGP*

Demonstração: Seja x^* com $h(x^*) = 0, g(x^*) \leq 0, x^* \in K$ tal que existe $\{x^k\} \subset K, x^k \to x^*$ com $d(\Omega_\gamma(x^k) \cap K, x^k) \to 0$. Vamos considerar a definição de AGP, sendo $\Omega'(x^k)$ o conjunto linearizado das restrições, isto é

$$\Omega'(x^k) = \left\{ y \in \mathbb{R}^n \;\middle|\; \begin{array}{cl} g_i(x^k) + \nabla g_i(x^k)^{\mathtt{T}}(y - x^k) \leq 0 & \text{se } \gamma < g_i(x^k) < 0 \\ \nabla g_i(x^k)^{\mathtt{T}}(y - x^k) \leq 0 & \text{se } g_i(x^k) \geq 0 \\ \nabla h_i(x^k)^{\mathtt{T}}(y - x^k) = 0 & \\ \bar{g}_i(x^k) + \nabla \bar{g}_i(x^k)^{\mathtt{T}}(y - x^k) \leq 0 & \text{se } \gamma < \bar{g}_i(x^k) < 0 \\ \nabla \bar{g}_i(x^k)^{\mathtt{T}}(y - x^k) \leq 0 & \text{se } \bar{g}_i(x^k) \geq 0 \\ \nabla \bar{h}_i(x^k)^{\mathtt{T}}(y - x^k) = 0 & \end{array} \right\}.$$

Note que não temos nenhum tratamento especial para as restrições lineares.

Como $\bar{g}(x^k) \leq 0, \bar{h}(x^k) = 0$ e $\bar{h}_i(z) = \bar{h}_i(x^k) + \nabla^{\mathtt{T}}\bar{h}_i(x^k)(z - x^k), \bar{g}_j(z) = \bar{g}_j(x^k) + \nabla^{\mathtt{T}}\bar{g}_j(x^k)(z - x^k)$ pois $\bar{h}$ e $\bar{g}$ são funções afins, temos que $\Omega'(x^k) = \Omega_\gamma(x^k) \cap K$ para k suficientemente grande, logo, utilizando a sequência $\{x^k\}$, temos que vale AGP. □

É interessante observar que L-AGP não é o mesmo que AGP, como mostramos abaixo.

Proposição 3.19. *AGP $\not\Rightarrow$ L-AGP*

Demonstração: Considere o problema

$$\begin{array}{ll} \text{Minimizar} & x_2 \\ \text{Sujeito a} & x_1 x_2 = 0 \\ & x \in K = \{x \in \mathbb{R}^2 | x_1 \geq 0, x_2 \geq 0\}. \end{array}$$

Este é o exemplo que mostra a Proposição (3.11), como neste caso K é um politopo, então temos o resultado. $\square$

Observe que como L-AGP é mais forte que AGP, temos que L-AGP herda de AGP o fato de que sob CPLD, implica KKT.

Corolário 3.20. *L-AGP $\Rightarrow$ KKT ou não-CPLD*

Demonstração:
É consequência imediata das Proposições (3.8) e (3.18). $\square$

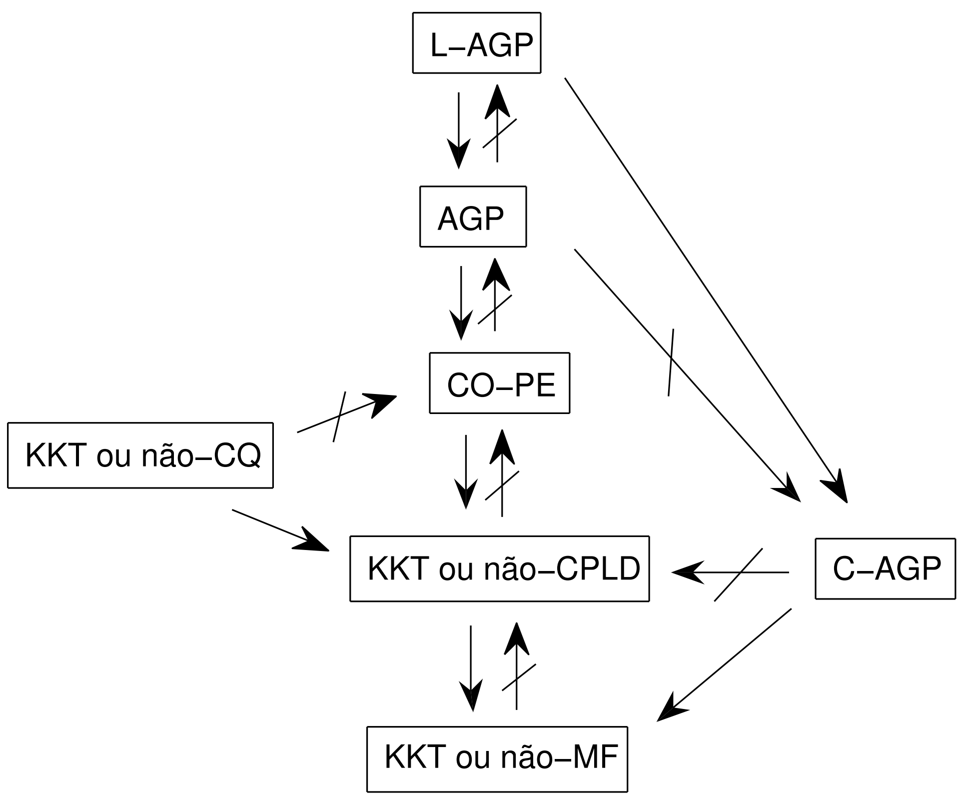

Figura 3.6: Relações entre as condições de otimalidade apresentadas.

Na figura 3.6 mostramos um resumo das relações entre as condições de otimalidade apresentadas. CQ é qualquer condição de qualificação fraca, tal que CPLD $\Rightarrow$ CQ. A implicação AGP $\Rightarrow$ CO-PE é simples, basta considerar as condições KKT do problema de projeção que define AGP. A demonstração em detalhes está feita em [59].

A seguir, mostraremos os contra-exemplos referentes às implicações que não valem, como mostra a figura 3.6.

Para estabelecermos que KKT ou não-CPLD não implica CO-PE, vamos considerar o problema

$$\begin{array}{ll} \text{Minimizar} & x_2 \\ \text{Sujeito a} & x_1^2 = 0, \end{array}$$

no ponto $x^* = (0,0)$. Temos que não vale CPLD, pois o gradiente da restrição, $(2x_1, 0)$, se anula em x^* (logo é PLD), mas é não nulo quando $x \neq (0,0)$ (portanto é PLI). Além disso, não vale CO-PE pois o gradiente da função objetivo é $(0,1)$ e portanto $||(0,1)+\lambda^k(2x_1^k, 0)||$ não vai a zero independentemente da escolha de λ^k e x^k. De fato é possível mostrar que a restrição $x_1^2 = 0$ não satisfaz nenhuma condição de qualificação (pois existe uma escolha de função objetivo tal que um minimizador não satisfaz KKT, a saber, basta escolher a função objetivo $f(x) = x_1$), logo vale o resultado mais geral que KKT ou não-CQ não implica CO-PE, qualquer que seja a condição de qualificação CQ.

Vamos mostrar que CO-PE não implica AGP. Para isso consideremos o problema

$$\begin{array}{ll} \text{Minimizar} & -x_2 \\ \text{Sujeito a} & x_1 x_2 = 0 \\ & -x_1 \leq 0, \end{array}$$

no ponto $x^* = (0,1)$. O gradiente da função objetivo $f(x) := -x_2$ é dado por $(0,-1)$ e o gradiente das restrições de igualdade $h(x) := x_1x_2 = 0$ e desigualdade $g(x) := -x_1 \leq 0$ são, respectivamente, (x_2, x_1) e $(-1, 0)$. Para mostrar que vale CO-PE, vamos considerar a sequência $x^k = (\frac{1}{k}, 1), \lambda^k = \mu^k = k$, assim

$$\begin{pmatrix} 0 \\ -1 \end{pmatrix} + \lambda^k \begin{pmatrix} x_2^k \\ x_1^k \end{pmatrix} + \mu^k \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix} + k \begin{pmatrix} 1 \\ \frac{1}{k} \end{pmatrix} + k \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

A condição $g(x^*) < 0 \Rightarrow \mu^k = 0$ é satisfeita pois $g(x^*) = -x_1^* = 0$, logo vale CO-PE.

Para mostrar que não vale AGP, vamos mostrar que não existe sequência $x^k \to x^*$ tal que $d^k = P_{-x^k+\Omega_\gamma(x^k)}(-\nabla f(x^k)) \to 0$, com $-\nabla f(x^k) = (0,1)$ e γ suficientemente negativo.

Supondo inicialmente $x_1^k > 0$ temos que $\gamma < g(x^k)$, assim

$$
\begin{aligned}
-x^k + \Omega_\gamma(x^k) = \left\{ d \in \mathbb{R}^2 \;\middle|\; \begin{array}{c} g(x^k)^- + \nabla g(x^k)^{\mathtt{T}} d \leq 0 \\ \nabla h(x^k)^{\mathtt{T}} d = 0 \end{array} \right\} &= \left\{ d \in \mathbb{R}^2 \;\middle|\; \begin{array}{c} -x_1^k + (-1,0)d \leq 0 \\ (x_2^k, x_1^k)d = 0 \end{array} \right\} \\
&= \left\{ d \in \mathbb{R}^2 \;\middle|\; \begin{array}{c} d_1 \geq -x_1^k \\ x_2^k d_1 + x_1^k d_2 = 0 \end{array} \right\}.
\end{aligned}
$$

Logo d^k é solução do problema

$$
\begin{array}{ll}
\text{Minimizar} & (d_1)^2 + (d_2 - 1)^2 \\
\text{Sujeito a} & d_1 \geq -x_1^k \\
& x_2^k d_1 + x_1^k d_2 = 0,
\end{array}
$$

mas como $x_1^k > 0$, temos que $d_2^k = -\dfrac{x_2^k d_1^k}{x_1^k}$ e d_1^k é solução de

$$
\begin{array}{ll}
\text{Minimizar} & F(d_1) := (d_1)^2 + (-\dfrac{x_2^k d_1}{x_1^k} - 1)^2 \\
\text{Sujeito a} & d_1 \geq -x_1^k.
\end{array}
$$

Se a solução $d_1^k = -x_1^k$ então $d_2^k = x_2^k \to 1$. Suponha então que $d_1^k > -x_1^k$, logo $F'(d_1^k) = 0$, ou seja

$$
2d_1^k + 2\left(\frac{x_2^k d_1^k}{x_1^k} + 1\right)\frac{x_2^k}{x_1^k} = 0 \Rightarrow d_1^k = \frac{-x_1^k x_2^k}{(x_1^k)^2 + (x_2^k)^2} \Rightarrow d_2^k = \frac{(x_2^k)^2}{(x_1^k)^2 + (x_2^k)^2},
$$

assim, como $x_1^k \to 0$ e $x_2^k \to 1$, então $d_2^k \to 1$. Vamos supor agora que $x_1^k \leq 0$, assim $g(x^k)^- = \min\{0, g(x^k)\} = \min\{0, -x_1^k\} = 0$, logo

$$
-x^k + \Omega_\gamma(x^k) = \left\{ d \in \mathbb{R}^2 \;\middle|\; \begin{array}{c} d_1 \geq 0 \\ x_2^k d_1 + x_1^k d_2 = 0 \end{array} \right\}.
$$

Sendo que d^k é solução de minimizar $(d_1)^2 + (d_2 - 1)^2$, sujeito a $d \in -x^k + \Omega_\gamma(x^k)$. Se $x_1^k = 0$ então claramente $d_2^k = 1$ para todo k. Suponha $x_1^k < 0$, então $d_2^k = -\frac{x_2^k d_1^k}{x_1^k}$ e d_1^k é solução de

$$\begin{array}{ll} \text{Minimizar} & F(d_1) := (d_1)^2 + (-\frac{x_2^k d_1}{x_1^k} - 1)^2 \\ \text{Sujeito a} & d_1 \geq 0. \end{array}$$

Se $d_1^k > 0$, então como anteriormente temos que $F'(d_1^k) = 0$ e concluímos do mesmo modo que $d_2^k \to 1$. Vamos mostrar que d_1^k não pode ser zero, com efeito, se fosse $d_1^k = 0$, pelas condições KKT deveríamos ter $F'(0) \geq 0$, mas $F'(0) = 1 + 2\frac{x_2^k}{x_1^k} \to -\infty$.

Em todos os casos concluímos que $d^k \not\to 0$, portanto não vale AGP.

O problema

$$\begin{array}{ll} \text{Minimizar} & x_2 \\ \text{Sujeito a} & x_1 x_2 = 0 \\ & x_1 \geq 0, x_2 \geq 0, \end{array}$$

considerando K as restrições de não-negatividade e o ponto $x^* = (0, 1)$, foi usado para mostrar que AGP não implica L-AGP (o mesmo para C-AGP). Em particular, como vale AGP, então vale CO-PE. Este resultado é importante pois mostra que o método de Lagrangiano aumentado [3, 4], cujos pontos limites satisfazem CO-PE, pode convergir para pontos que não satisfazem L-AGP.

Acreditamos que estas novas condições sequenciais de otimalidade podem ser aplicadas para o desenvolvimento de novos algoritmos. Em particular, algoritmos baseados em filtros [19, 20, 18, 23, 60].

3.3 Condições do tipo AGP generalizadas

Relembrando a definição da condição AGP, temos que x viável satisfaz AGP se existe uma sequência $x^k \to x$, tal que $d_k = P_{-x^k+\Omega_\gamma(x^k)}(-\nabla f(x^k)) \to 0$, ou seja, d_k é solução do problema

$$\begin{array}{ll} \text{Minimizar} & \frac{1}{2}||d + \nabla f(x^k)||^2 \\ \text{Sujeito a} & x^k + d \in \Omega_\gamma(x^k). \end{array}$$

Note que podemos reescrever a função objetivo como sendo

$$\frac{1}{2}||d + \nabla f(x^k)||^2 = \frac{1}{2}d^{\mathsf{T}}d + \nabla^{\mathsf{T}} f(x^k)d + c_k,$$

onde $c_k = \nabla^{\mathsf{T}} f(x^k)\nabla f(x^k)$ não depende de d e, portanto, pode ser descartado.

O nosso objetivo é generalizar a condição AGP, trocando a função objetivo do subproblema que define d_k por $Q_k(d) = \frac{1}{2}d^{\mathsf{T}}B_k d + \nabla^{\mathsf{T}} f(x^k)d$, assim teremos mais liberdade na escolha de d_k, o que pode gerar algoritmos mais eficientes. Mostraremos que essa nova condição é equivalente a AGP.

Podemos assumir sem perda de generalidade que as matrizes B_k são simétricas, caso contrário bastaria trocar B_k por $\frac{B_k + B_k^{\mathsf{T}}}{2}$, já que a função objetivo assume o mesmo valor. Vamos supor que $\{B_k\}$ é limitada, pois esta condição é equivalente à implicação: $d_k \to 0 \Rightarrow B_k d_k \to 0$, argumento que será essencial nas demonstrações. Além disso, se B é um ponto limite de $\{B_k\}$, será essencial nas demonstrações que B seja definida positiva. Como o conjunto das matrizes definidas positivas é um aberto no conjunto das matrizes reais com a topologia usual (isso é consequência do fato de que os autovalores são positivos e funções contínuas da matriz [63]), temos que B_k é definida positiva se B_k está suficientemente perto de B, assim, uma hipótese natural é que a sequência $\{B_k\}$ seja uniformemente definida positiva. Como consequência teremos que existem constantes $L_1 > 0, L_2 > 0$ tal que $L_1 \leq ||B_k|| \leq L_2$ e $L_1 \leq ||B_k^{-1}|| \leq L_2$.

Definição 3.21. *Dizemos que uma sequência de matrizes $\{B_k\}$ é uniformemente definida positiva se existe $\varepsilon > 0$ tal que $x^{\mathsf{T}}B_k x \geq \varepsilon$ para todo $x \in \mathbb{R}^n$ com $||x|| = 1$.*

Posteriormente vamos enfraquecer esta hipótese e imporemos que $\{B_k\}$ seja uniformemente definida positiva apenas no conjunto Ω_k a ser definido.

As proposições abaixo são de fácil verificação.

Proposição 3.22. *Se B_k é simétrica para todo k, então $\{B_k\}$ é uniformemente definida* *positiva se, e somente se, existe $\varepsilon > 0$ tal que o menor autovalor λ_k de B_k é tal que $\lambda_k \geq \varepsilon$ para todo k.*

Proposição 3.23. *$\{B_k\}$ é limitada se, e somente se, $x^k \to 0 \Rightarrow B_k x^k \to 0$.*

Proposição 3.24. *Se B_k é simétrica para todo k com $\{B_k\}$ uniformemente definida positiva e limitada, então existem constantes $L_1 > 0, L_2 > 0$ tais que $L_1 \leq ||B_k|| \leq L_2$ e $L_1 \leq ||B_k^{-1}|| \leq L_2$ para todo k.*

Vamos considerar o problema

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & x \in X, \end{array}$$

onde o conjunto viável é dado por $X = \Omega \cap K$, sendo $\Omega = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0\}$ definido por funções continuamente diferenciáveis e K um convexo fechado qualquer.

Definição 3.25 (C-AGP-generalizada)**.** *Dado um ponto viável x e uma sequência $\{B_k\}$ de matrizes simétricas, uniformemente definida positiva e limitada, diremos que x satisfaz C-AGP-generalizada se existe $\{x^k\} \subset K, x^k \to x$ com d_k solução de*

$$\begin{array}{ll} \text{Minimizar} & Q_k(d) \\ \text{Sujeito a} & x^k + d \in \Omega_\gamma(x^k) \\ & x^k + d \in K \end{array} \tag{3.12}$$

tal que $d_k \to 0$, onde $Q_k(d) = \frac{1}{2} d^\mathsf{T} B_k d + \nabla^\mathsf{T} f(x^k) d$.

Analagomente, definimos L-AGP-generalizada quando K é um politopo e AGP-generalizada quando $K = \mathbb{R}^n$.

Denotaremos o conjunto viável do subproblema por $\Omega_k = \{d \in \mathbb{R}^n | x^k + d \in \Omega_\gamma(x^k), x^k + d \in K\}$. Note que Ω_k é convexo.

Como B_k é simétrica e definida positiva, existe H_k simétrica e definida positiva tal que $B_k = H_k^2$, basta considerar a diagonalização ortogonal de $B_k = U \Sigma U^\mathsf{T}$ e definir $H_k = U \Sigma^{\frac{1}{2}} U^\mathsf{T}$, assim, podemos reescrever o subproblema (3.12)

$$\begin{array}{ll} \text{Minimzar} & \frac{1}{2}(H_k d)^{\mathrm{T}}(H_k d) + (H_k d)^{\mathrm{T}}(H_k^{-1}\nabla f(x^k)) \\ \text{Sujeito a} & d \in \Omega_k \end{array}$$

Fazendo a mudança de variável $\tilde{d} = H_k d$ e definindo $H_k\Omega_k = \{\tilde{d} | H_k^{-1}\tilde{d} \in \Omega_k\}$ podemos reescrever novamente

$$\begin{array}{ll} \text{Minimizar} & \frac{1}{2}\tilde{d}^{\mathrm{T}}\tilde{d} + \tilde{d}^{\mathrm{T}}(H_k^{-1}\nabla f(x^k)) \\ \text{Sujeito a} & \tilde{d} \in H_k\Omega_k. \end{array}$$

Observando que se Ω_k é um convexo, não vazio, fechado qualquer, então $H_k\Omega_k$ também será convexo, não vazio e fechado, temos que a solução $\tilde{d}^*$ do problema acima será dada por $\tilde{d}^* = P_{H_k\Omega_k}(-H_k^{-1}\nabla f(x^k))$, logo a solução d_k do subproblema (3.12) é dada por

$$d_k = H_k^{-1} P_{H_k\Omega_k}(-H_k^{-1}\nabla f(x^k)).$$

Podemos definir a projeção generalizada de y em $X \subset \mathbb{R}^n$ (convexo, não vazio e fechado), denotada por $P_X^B(y)$, como a solução de

$$\begin{array}{ll} \text{Minimizar} & \frac{1}{2}d^{\mathrm{T}}Bd - y^{\mathrm{T}}d \\ \text{Sujeito a} & d \in X, \end{array} \tag{3.13}$$

sendo B simétrica e definida positiva. Então temos

Proposição 3.26. *Se $x, x' \in \mathbb{R}^n$, então $||P_X^B(x) - P_X^B(x')|| \leq ||B^{-1}||||x - x'||$*

Demonstração: Seja H simétrica e definida positiva tal que $B = H^2$, assim $P_X^B(x) = H^{-1}P_{HX}(H^{-1}x)$ e $P_X^B(x') = H^{-1}P_{HX}(H^{-1}x')$, sendo $HX = \{d \in \mathbb{R}^n | H^{-1}d \in X\}$.

Logo

$$\begin{aligned} ||P_X^B(x) - P_X^B(x')|| = ||H^{-1}(P_{HX}(H^{-1}x) - P_{HX}(H^{-1}x'))|| \leq \\ \leq ||H^{-1}||||P_{HX}(H^{-1}x) - P_{HX}(H^{-1}x')|| \leq ||H^{-1}||||H^{-1}x - H^{-1}x'|| \leq \\ \leq ||H^{-1}||^2||x - x'|| = ||B||||x - x'||, \end{aligned}$$

sendo que aplicamos o Lema 3.2 e usamos que $||B^{-1}|| = ||H^{-1}||^2$, já que a norma é o maior autovalor, e os autovalores de H^{-1} são as raízes quadradas dos autovalores de B^{-1}. $\square$

Considerando B igual à identidade, temos que este resultado generaliza o Lema 3.2.

Com as técnicas descritas acima, podemos mostrar que C-AGP-generalizada (e os análogos L-AGP-generalizada e AGP-generalizada) são condições de otimalidade. Para isso, precisamos supor que o convexo K satisfaz alguma condição de qualificação qualquer e é definido por restrições de igualdade lineares $\bar{h}(x) = 0$ e de desigualdade convexas $\bar{g}(x) \leq 0$. Procedendo como foi feito na demonstração de que C-AGP é uma condição de otimalidade, obtemos, via penalidade externa, uma sequência $\{x^k\}$ convergindo para o minimizador global x^*, tal que

$$\nabla f(x^k) + y^k = (x^* - x^k),$$

onde

$$y^k = \sum_{i=1}^{m} \rho_k h_i(x^k) \nabla h_i(x^k) + \sum_{j | g_j(x^k) > 0} \rho_k g_j(x^k) \nabla g_j(x^k) +$$
$$+ \sum_{i=1}^{\bar{m}} \bar{\lambda}_i \nabla \bar{h}_i(x^k) + \sum_{j | g_j(x^k) = 0} \bar{\mu}_j \nabla \bar{g}_j(x^k) = (x^* - x^k).$$

Com uma verificação direta das equações KKT, análoga à que foi feita no Lema 3.3, temos $P_{\Omega_k}^{B_k}(y^k) = 0$. Assim

$$||d_k|| = ||P_{\Omega_k}^{B_k}(-\nabla f(x^k))|| = ||P_{\Omega_k}^{B_k}(-\nabla f(x^k)) - P_{\Omega_k}^{B_k}(y^k)|| \leq$$
$$\leq ||B_k^{-1}|| || - \nabla f(x^k) - y^k|| = ||B_k^{-1}|| ||x^k - x^*|| \to 0.$$

Agora vamos mostrar que a condição de otimalide C-AGP-generalizada é de fato equivalente a C-AGP usual.

Proposição 3.27. *Sejam $\{B_k\}$ e $\{A_k\}$ sequências de matrizes simétricas, uniformemente definidas positivas e limitadas. Se x satisfaz C-AGP-generalizada com a sequência $\{B_k\}$ então também satisfaz com a sequência $\{A_k\}$.*

Demonstração: Seja $x^k \to x$, $x^k \in K$ e $d_k \to 0$ solução de

$$\begin{array}{ll} \text{Minimizar} & \frac{1}{2}d^{\mathsf{T}}B_k d + \nabla^{\mathsf{T}} f(x^k)d \\ \text{Sujeito a} & d \in \Omega_k. \end{array}$$

Seja H_k simétrica e definida positiva tal que $B_k = H_k^2$, então $H_k d_k = P_{H_k\Omega_k}(-H_k^{-1}\nabla f(x^k))$, assim, pelo Lema 3.1, temos

$$(-H_k^{-1}\nabla f(x^k) - H_k d_k)^{\mathsf{T}}(x - H_k d_k) \leq 0 \quad \forall x \in H_k\Omega_k.$$

Como $x \in H_k\Omega_k \Leftrightarrow x = H_k d$, com $d \in \Omega_k$, temos

$$\begin{aligned} (-H_k^{-1}\nabla f(x^k) - H_k d_k)^{\mathsf{T}}(H_k d - H_k d_k) \leq 0 \Rightarrow \\ \Rightarrow -\nabla^{\mathsf{T}} f(x^k)H_k^{-1}H_k d - d_k^{\mathsf{T}}H_k^2 d + \nabla^{\mathsf{T}} f(x^k)H_k^{-1}H_k d_k + d_k^{\mathsf{T}}H_k^2 d_k \leq 0 \Rightarrow \\ \Rightarrow -\nabla^{\mathsf{T}} f(x^k)d - d_k^{\mathsf{T}}B_k d + \nabla^{\mathsf{T}} f(x^k)d_k + d_k^{\mathsf{T}}B_k d_k \leq 0 \quad \forall d \in \Omega_k. \end{aligned} \tag{3.14}$$

Analogamente, seja $\bar{d}_k$ solução de

$$\begin{array}{ll} \text{Minimizar} & \frac{1}{2}d^{\mathsf{T}}A_k d + \nabla^{\mathsf{T}} f(x^k)d \\ \text{Sujeito a} & d \in \Omega_k, \end{array}$$

então

$$-\nabla^{\mathsf{T}} f(x^k)d - (\bar{d}_k)^{\mathsf{T}}A_k d + \nabla^{\mathsf{T}} f(x^k)\bar{d}_k + (\bar{d}_k)^{\mathsf{T}}A_k\bar{d}_k \leq 0 \quad \forall d \in \Omega_k. \tag{3.15}$$

Vamos mostrar que $\{\bar{d}_k\}$ é limitada. Como $0 \in \Omega_k$ pois $x^k \in K \cap \Omega_\gamma(x^k)$, podemos tomar $d = 0$ em (3.15) para obtermos

$$\nabla^{\mathsf{T}} f(x^k)\bar{d}_k + (\bar{d}_k)^{\mathsf{T}}A_k\bar{d}_k \leq 0.$$

Supondo $||\bar{d}_k|| \to \infty$ e dividindo por $||\bar{d}_k||^2$ temos $\left(\frac{\bar{d}_k}{||\bar{d}_k||}\right)^{\mathsf{T}} A_k \left(\frac{\bar{d}_k}{||\bar{d}_k||}\right) \leq -\frac{1}{||\bar{d}_k||}\nabla^{\mathsf{T}} f(x^k)\frac{\bar{d}_k}{||\bar{d}_k||}$, como o lado direito tende a zero, temos uma contradição com o fato de $\{A_k\}$ ser uniformemente definida positiva.

Assim, considere uma subsequência tal que $\bar{d}_k \to \bar{d}$, tome $d = \bar{d}_k \in \Omega_k$ em (3.14) e aplique o limite. Usando que $d_k \to 0$ e $\{B_k\}$ é limitada temos

$$\nabla^{\mathsf{T}} f(x)\bar{d} \geq 0.$$

Escolhendo $d = d_k \in \Omega_k$ em (3.15) temos

$$\nabla^{\mathsf{T}} f(x^k)\bar{d}_k + (\bar{d}_k)^{\mathsf{T}} A_k \bar{d}_k \leq \nabla^{\mathsf{T}} f(x^k) d_k + (\bar{d}_k)^{\mathsf{T}} A_k d_k,$$

como $d_k \to 0$ e $\{A_k\}$ é limitada, temos que o lado direito tende para zero, assim, como $\nabla^{\mathsf{T}} f(x^k)\bar{d}_k \to \nabla^{\mathsf{T}} f(x)\bar{d} \geq 0$ e $\{A_k\}$ é uniformemente definida positiva, temos que $\bar{d}_k \to 0$.
□

Temos portanto o seguinte resultado:

Corolário 3.28. *C-AGP-generalizada* $\Leftrightarrow$ *C-AGP*

Demonstração: Basta considerar uma sequência de matrizes constantes e iguais à identidade, na proposição anterior. □

Corolário 3.29. *L-AGP-generalizada* $\Leftrightarrow$ *L-AGP e AGP-generalizada* $\Leftrightarrow$ *AGP*

Demonstração: Basta considerar K um politopo, respectivamente $K = \mathbb{R}^n$. □

Note que da demonstração da Proposição 3.27 obtemos também o seguinte resultado

Corolário 3.30. *Seja* $\{B_k\}$ *uma sequência de matrizes simétricas, uniformemente definida positiva e limitada, com* $\{x^k\} \subset K$ *e* d_k *solução de*

$$\begin{array}{ll} \min & \frac{1}{2} d^{\mathsf{T}} B_k d + \nabla^{\mathsf{T}} f(x^k) d \\ \text{suj.} & d \in \Omega_k, \end{array} \tag{3.16}$$

então $\{d_k\}$ *é limitada.*

É interessante observar que é possível enfraquecer a condição de que a sequência $\{A_k\}$ seja uniformemente definida positiva para obtermos a equivalência entre C-AGP-generalizada e C-AGP (e os análogos para L-AGP e AGP), a saber, podemos impor apenas que $\{A_k\}$ seja uniformemente definida positiva em Ω_k. Primeiro, observamos que as equações (3.14) e (3.15) na demonstração da Proposição 3.27 podem ser obtidas sem recorrer ao fato da hessiana ser definida positiva.

Proposição 3.31. *Seja d^* solução de*

$$\begin{array}{ll} \textit{Minimizar} & \frac{1}{2}d^{\mathtt{T}}Ad + b^{\mathtt{T}}d \\ \textit{Sujeito a} & d \in \Omega, \end{array}$$

com Ω convexo, fechado e satisfazendo alguma condição de qualificação. Então temos

$$-b^{\mathtt{T}}d - (d^*)^{\mathtt{T}}Ad + b^{\mathtt{T}}d + (d^*)^{\mathtt{T}}Ad^* \leq 0, \forall d \in \Omega.$$

Demonstração: Temos que a solução d^* satisfaz KKT, logo, pela Proposição 3.4 temos

$$P_\Omega(d^* - Ad^* - b) = d^*,$$

assim, pelo Lema 3.1 temos

$$(d^* - Ad^* - b - d^*)^{\mathtt{T}}(d - d^*) \leq 0, \forall d \in \Omega,$$

e efetuando o produto obtemos o resultado. □

Agora basta observar que o resto do argumento na demonstração da Proposição 3.27 segue da mesma maneira, impondo apenas que A_k seja uniformemente definida positiva em Ω_k.

3.4 Um algoritmo de restauração inexata

A seguir vamos mostrar a relação entre as condições de otimalidade do tipo AGP apresentadas e o algoritmo de restauração inexata proposto por Fischer e Friedlander em

[17].

Considere o problema

$$\begin{array}{ll} \min & f(x) \\ \text{suj.} & h(x) = 0 \\ & g(x) \leq 0 \\ & x \in K, \end{array} \tag{3.17}$$

onde f, g e h são continuamente diferenciáveis e $K = \{x \in \mathbb{R}^n | \bar{h}(x) = 0, \bar{g}(x) \leq 0\}$ é um convexo, compacto, satisfazendo alguma condição de qualificação.

A idéia da restauração inexata é primeiramente dar um passo que melhore a viabilidade, e em seguida obter uma melhora da função objetivo no espaço tangente às restrições.

Defina a função $H(x) = \max\{||h(x)||_\infty, ||g^+(x)||_\infty\}$ para $x \in K$ e a função de mérito $\Phi(x, p) = pf(x) + (1-p)H(x)$ com $p \in [0, 1]$ e $x \in K$.

Algoritmo (Fischer-Friedlander)

Seja $r \in [0, 1), \beta > 0$, $\hat{\gamma} > 0, \bar{\gamma} > 0$ e $\tau > 0$.

Passo 0: *Inicialização*

Escolha $x^0 \in K$ e $p_0 \in (0, 1)$. Defina $k = 0$.

Passo 1: *Restauração Inexata*

Calcule $y^k \in K$ tal que

$$\begin{array}{rcl} H(y^k) & \leq & rH(x^k) \\ f(y^k) & \leq & f(x^k) + \beta H(x^k) \end{array}$$

Passo 2: *Escolha do parâmetro de penalidade*

Calcule $p_{k+1} \in \{2^{-i} p_k | i \in \{0, 1, 2, \dots\}\}$ maior possível tal que

$$\Phi(y^k, p_{k+1}) - \Phi(x^k, p_{k+1}) \leq \frac{1}{2}(1-r)(H(y^k) - H(x^k))$$

Passo 3: *Escolha da direção de busca*

Calcular d_k tal que

$$f(y + td_k) \leq f(y^k) - \hat{\gamma}t||d_k||^2$$
$$H(y + td_k) \leq H(y^k) + \bar{\gamma}t^2||d_k||^2,$$

para todo $y \in K$ e todo $t \in [0, \tau]$.

Passo 4: *Minimização*

Calcule $t_k \in \{2^{-i}|i \in \{0, 1, 2, \dots\}\}$ maior possível tal que

$$\Phi(y^k + t_k d_k, p_{k+1}) - \Phi(x^k, p_{k+1}) \leq \frac{1}{2}(1 - r)(H(y^k) - H(x^k))$$

Passo 5: *Atualização*

Defina $x^{k+1} = y^k + t_k d_k$, $k := k + 1$ e volte para o Passo 1.

Convergência

O resultado de convergência presente em [17] é que $H(x^k) \to 0$ e $d_k \to 0$. Portanto é preciso escolher direções d_k tal que o fato de $d_k \to 0$ tenha um significado com relação a otimalidade dos pontos limites.

Vamos definir d_k como solução do subproblema

$$\begin{array}{ll} \min & Q_k(d) \\ \text{suj.} & y^k + d \in \Omega_\gamma(y^k) \\ & y^k + d \in K, \end{array} \tag{3.18}$$

onde $Q_k(d) = \frac{1}{2}d^{\mathrm{T}}B_k d + \nabla^{\mathrm{T}} f(y^k)d$ e $\{B_k\}$ é uma sequência de matrizes simétricas, uniformemente definida positiva e limitada.

Vamos mostrar que esta direção satisfaz as condições do algoritmo modelo de Fischer e Friedlander.

Lema 3.32. *Se $F: \mathbb{R} \to \mathbb{R}$ é continuamente diferenciável e $t \in \mathbb{R}$ então*

$$t \int_0^1 (F'(tx) - F'(0))dx = F(t) - F(0) - tF'(0).$$

Demonstração: Basta observar que se $t \neq 0$, $\int_0^1 F'(tx)dx = \frac{1}{t}(F(t) - F(0))$. $\square$

Lema 3.33. *Suponha que os gradientes de f, g e h são Lipschitz contínuas em K, então existem constantes positivas $\hat{\gamma}, \bar{\gamma}, \tau$ tais que*

$$f(y^k + td_k) \leq f(y^k) - \hat{\gamma}t||d_k||^2$$

e

$$H(y^k + td_k) \leq H(y^k) + \bar{\gamma}t^2||d_k||^2,$$

para todo $y^k \in K$ e todo $t \in [0, \tau]$.

Demonstração: Seja $t \in [0, 1]$ e $y^k \in K$ com d_k definido por (3.18). Como K é convexo temos $y^k + td_k \in K$. Pelo Lema 3.32 temos

$$f(y^k + td_k) = f(y^k) + t\nabla^{\mathsf{T}} f(y^k)d_k + t\int_0^1 (\nabla f(y^k + \sigma td_k) - \nabla f(y^k))^{\mathsf{T}} d_k d\sigma.$$

Como ∇f é Lipschitz contínua, existe $L > 0$ tal que

$$(\nabla f(y^k + \sigma td_k) - \nabla f(y^k))^{\mathsf{T}} d_k \leq ||\nabla f(y^k + \sigma td_k) - \nabla f(y^k)||||d_k|| \leq L||\sigma td_k||||d_k||.$$

Assim,

$$\int_0^1 (\nabla f(y^k + \sigma td_k) - \nabla f(y^k))^{\mathsf{T}} d_k d\sigma \leq \int_0^1 L\sigma t||d_k||^2 d\sigma = \frac{1}{2}Lt||d_k||^2.$$

Como $d = 0$ é viável para o subproblema, temos que $0 = Q_k(0) \geq Q_k(d_k) = \frac{1}{2}d_k^{\mathsf{T}}B_kd_k + \nabla^{\mathsf{T}} f(y^k)d_k$, logo $\nabla^{\mathsf{T}} f(y^k)d_k \leq -\frac{1}{2}d_k^{\mathsf{T}}B_kd_k$, assim, $f(y^k + td_k) \leq f(y^k) - \frac{1}{2}td_k^{\mathsf{T}}B_kd_k + \frac{1}{2}t^2L||d_k||^2$.

Como $\{B_k\}$ é uniformemente definida positiva, existe $\varepsilon > 0$ tal que $d_k^{\mathsf{T}} B_k d_k \geq \varepsilon d_k^{\mathsf{T}} d_k$, logo, $f(y^k + td_k) \leq f(y^k) - \frac{1}{2}t\varepsilon||d_k||^2 + \frac{1}{2}t^2 L||d_k||^2 \quad \forall t \in [0,1]$.

Defina $\tau = \min\{1, \frac{\varepsilon}{2L}\}$ e $\hat{\gamma} = \frac{\varepsilon}{4}$, assim, se $t \in [0,\tau]$ temos $t \leq \frac{\varepsilon - 2\hat{\gamma}}{L} \Rightarrow -\frac{1}{2}t\varepsilon + \frac{1}{2}t^2 L \leq -\hat{\gamma}t$, então $f(y^k + td_k) \leq f(y^k) - \hat{\gamma}t||d_k||^2 \quad \forall t \in [0,\tau]$.

Analogamente, como $\nabla^{\mathsf{T}} h_i(y^k) d_k = 0$, pois $d_k \in \Omega_\gamma(y^k)$, temos $h_i(y^k + td_k) \leq h_i(y^k) + \frac{1}{2}Lt^2||d_k||^2$ (podemos assumir que a constante $L > 0$ é a mesma para todas as funções).

Se $g_j(y^k) \geq 0$, então $\nabla^{\mathsf{T}} g_j(y^k) d_k \leq 0$, assim, $g_j(y^k + td_k) \leq g_j(y^k) + \frac{1}{2}Lt^2||d_k||^2$.

Se $\gamma < g_j(y^k) < 0$, então $\nabla^{\mathsf{T}} g_j(y^k) d_k \leq -g_j(y^k)$, logo $g_j(y^k + td_k) \leq \frac{1}{2}Lt^2||d_k||^2$.

Se $g_j(y^k) \leq \gamma$, como g_j é uniformemente contínua em K, pois K é compacto, então se $||td_k||$ é suficientemente pequeno temos $g(y^k + td_k) < 0$. Como $\{d_k\}$ é limitado (Corolário (3.30)), podemos reduzir τ se necessário tal que $g_j(y^k + td_k) < 0$ para todo $t \in [0,\tau]$ e todo $y^k \in K$ com $g(y^k) \leq \gamma$.

Em qualquer dos casos temos $|g_j^+(y^k + td_k)| \leq |g_j^+(y^k)| + \frac{1}{2}Lt^2||d_k||^2$ e $|h_i(y^k + td_k)| \leq |h_i(y^k)| + \frac{1}{2}Lt^2||d_k||^2$, logo $H(y^k + td_k) \leq H(y^k) + \bar{\gamma}t^2||d_k||^2 \quad \forall t \in [0,\tau]$, com $\bar{\gamma} = \frac{L}{2}$.

□

Agora podemos aplicar a teoria desenvolvida em [17]. Reescrevemos o resultado aqui por conveniência, utilizando a nomenclatura definida neste capítulo.

Proposição 3.34. *Se o Passo 1 do Algoritmo está bem definido para todo k, então todo ponto limite de uma sequência $\{x^k\}$ gerada pelo algoritmo é viável e satisfaz C-AGP*

Demonstração: O Teorema 1 em [17] mostra que $H(x^k) \to 0$ e o Teorema 2 em [17] juntamente com o Lema 3.33 mostra que $d_k \to 0$. □

No artigo [17], os autores consideram o uso da direção $d_k = P_{\Omega_\gamma(y^k)}(-\nabla f(y^k))$, resultando que os pontos limites são C-AGP. Com a generalização da condição C-AGP, podemos definir uma família de direções d_k mais promissoras, por exemplo, definindo a função lagrangiana $L(x, \lambda, \mu) = f(x) + \lambda^\mathsf{T} h(x) + \mu^\mathsf{T} g(x)$, vamos definir as matrizes $B_k = \nabla^2 L(y^k, \lambda^k, \mu^k)$, e vamos supor que perto da solução, esta sequência de matrizes é limitada e uniformemente definida positiva em $\Omega_\gamma(y^k)$. Observe que os multiplicadores de Lagrange associados ao subproblema que define d_k fornecem uma nova aproximação para λ^k e μ^k.

Capítulo 4

Alguns resultados

Neste capítulo mostramos alguns resultados sobre a teoria dos algoritmos propostos em [10] e [13].

Em [10] os autores propõem um método de Lagrangiano aumentado para a obtenção do minimizador global de um problema de otimização. Nosso resultado trata do algoritmo considerando a obtenção de pontos aproximadamente estacionários do Lagrangiano, ao invés da minimização aproximada como em [10]. Obtemos que os pontos limites são estacionários para o problema de minimizar a inviabilidade (supondo MF) e, sendo o ponto limite viável, temos sob a CPLD que os pontos limites são estacionários para o problema original com um conjunto de restrições adicionais, lineares, satisfeitas pelo minimizador global.

Tratamos também do algoritmo proposto em [13], onde os autores propõem um método primal-dual para a solução do subproblema do método de penalidade interna, resolvendo o sistema KKT de uma sequência de subproblemas penalizados. Obtivemos alguns resultados substituindo MF por CPLD.

4.1 Algoritmo de Birgin, Floudas e Martínez

Considere o problema

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeito a} & x \in X = \{x \in \mathbb{R}^n | h(x) = 0, g(x) \leq 0\} \\ & x \in \Omega = \{x \in \mathbb{R}^n | a^{\min} \leq x \leq a^{\max}\} \end{array}$$

Onde $f\colon \mathbb{R}^n \to \mathbb{R}$, $h\colon \mathbb{R}^n \to \mathbb{R}^m$, $g\colon \mathbb{R}^n \to \mathbb{R}^p$ são continuamente diferenciáveis em um aberto contendo Ω.

Seja $P_k = \{x \in \mathbb{R}^n | (a_i^k)^{\mathrm{T}} x \leq b_i^k, i = 1, \ldots, r_k\}$ e suponha que o minimizador global z do problema é tal que $z \in P_k \quad \forall k$.

Definimos a função Lagrangiano Aumentado como

$$L_\rho(x, \lambda, \mu) = f(x) + \frac{\rho}{2} \left\{ \sum_{i=1}^{m} \left[h_i(x) + \frac{\lambda_i}{\rho} \right]^2 + \sum_{i=1}^{p} \left[\max\left(0, g_i(x) + \frac{\mu_i}{\rho}\right) \right]^2 \right\}$$

Lema 4.1. *$L_\rho(x, \lambda, \mu)$ é continuamvente diferenciável com respeito a x.*

Demonstração: É consequência imediata do Lema 2.1. □

Assim, o gradiente de $L_\rho(x, \lambda, \mu)$ com respeito a x é dado por

$$\nabla L_\rho(x, \lambda, \mu) = \nabla f(x) + \sum_{i=1}^{m} (\lambda_i + \rho h_i(x)) \nabla h_i(x) + \sum_{i=1}^{p} \max(0, \mu_i + \rho g_i(x)) \nabla g_i(x).$$

O algoritmo em [10] consiste em encontrar iterandos x^k tais que

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

para todo $x \in \Omega \cap P_k$ com $\varepsilon_k \to 0$, deste modo, os autores mostraram que todo ponto limite é solução global do problema original.

Consideraremos uma versão relaxada, onde ao invés de x^k ser um minimizador global aproximado do Lagrangiano, x^k será aproximadamente KKT para o problema de minimizar $L_{\rho_k}(x, \lambda^k, \mu^k)$, sujeito a $x \in \Omega \cap P_k$.

O nosso primeiro resultado mostra que todo ponto limite do algoritmo é estacionário para um problema de minimzar a inviabilidade, desde que $\Omega \cap P$ satisfaça MF, onde P é o limite do conjunto de restrições P_k (a definição precisa de P será dada nas proposições a seguir). No caso do ponto limite ser viável, o segundo resultado que obtivemos é que todo ponto limite viável é KKT para o problema de minimizar $f(x)$ sujeito a $x \in X \cap \Omega \cap P$, desde que este problema satisfaça a CPLD, com P satisfazendo MF.

Os Passos 2 e 3 do algoritmo abaixo para determinar o parâmetro de penalidade ρ_{k+1} e λ_{k+1}, μ_{k+1} são feitos como no artigo original.

A idéia deste algoritmo é construir as restrições P_k de maneira a evitar minimizadores locais, que não são globais. Assim, espera-se que a solução do problema original incluindo as restrição em P seja o minimizador global z do problema original. Veja os detalhes em [10].

Algoritmo (Birgin-Floudas-Martínez)

Parâmetros: $\lambda_{\min} < \lambda_{\max}, \mu_{\max} > 0, \gamma > 1, 0 < \tau < 1$, uma sequência $\{\varepsilon_k\}$ com $\varepsilon_k \geq 0, \varepsilon_k \to 0, \bar{\lambda}_i^1 \in [\lambda_{\min}, \lambda_{\max}] \forall i = 1, \ldots, m, \bar{\mu}_i^1 \in [0, \mu_{\max}] \forall i = 1, \ldots, p, \rho_1 > 0, k \leftarrow 1$.

Passo 1: Encontre $x^k \in \mathbb{R}^n, u^k \in \mathbb{R}^n, v^k \in \mathbb{R}^n, w^k \in \mathbb{R}^{r_k}$ tal que

$$\delta_k = \nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) + \sum_{i=1}^{n} u_i^k(-e_i) + \sum_{i=1}^{n} v_i^k e_i + \sum_{i=1}^{r_k} w_i^k a_i^k$$

$$||\delta_k|| \leq \varepsilon_k$$

$$u_i^k \geq 0 \forall i = 1, \ldots, n, \quad v_i^k \geq 0 \forall i = 1, \ldots, n, \quad w_i^k \geq 0 \forall i = 1, \ldots, r_k$$

$$a_i^{\min} - \varepsilon_k \leq x_i^k \leq a_i^{\max} + \varepsilon_k \quad \forall i = 1, \ldots, n$$

$$(a_i^k)^{\mathrm{T}} x^k - b_i^k \leq \varepsilon_k \quad \forall i = 1, \ldots, r_k$$

$$a_i^{\min} - x_i^k < -\varepsilon_k \Rightarrow u_i^k = 0$$

$$x_i^k - a_i^{\max} < -\varepsilon_k \Rightarrow v_i^k = 0$$

$$(a_i^k)^{\mathrm{T}} x^k - b_i^k < -\varepsilon_k \Rightarrow w_i^k = 0$$

Passo 2: Defina $V_i^k = \max\left\{g_i(x^k), -\frac{\bar{\mu}_i^k}{\rho_k}\right\} \quad \forall i = 1, \ldots, p.$
Se $\max\{||h(x^k)||_\infty, ||V^k||_\infty\} \leq \tau \max\{||h(x^{k-1})||_\infty, ||V^{k-1}||_\infty\}$
então $\rho_{k+1} = \rho_k$ senão $\rho_{k+1} = \gamma\rho_k$.

Passo 3: Calcule $\bar{\lambda}_i^{k+1} \in [\lambda_{\min}, \lambda_{\max}] \forall i = 1, \ldots, m, \quad \bar{\mu}_i^{k+1} \in [0, \mu_{\max}] \forall i = 1, \ldots, p, \quad k \leftarrow k+1$ e volte para o Passo 1.

Proposição 4.2. *Suponha que* $\{r_k\}, \{a_i^k\}, \{b_i^k\}$ *são limitadas para cada* i*, logo existe* $K_0 \underset{\infty}{\subset} \mathbb{N}$ *tal que* $r_k = r \quad \forall k \in K_0$, $\lim_{k \in K_0} a_i^k = a_i$ *e* $\lim_{k \in K_0} b_i^k = b_i$. *Defina* $P = \{x \in \mathbb{R}^n | a_i^T x \leq b_i, i = 1, \ldots, r\}$. *Seja* x^* *um ponto limite da sequência* $\{x^k\}$ *gerada pelo algoritmo BFM, então* $x^* \in \Omega \cap P$. *Se, além disso,* x^* *é um ponto Mangasarian-Fromovitz de* $\Omega \cap P$ *então* x^* *é um ponto KKT do problema de minimizar* $||h(x)||_2^2 + ||\max(0, g(x))||_2^2$ *sujeito a* $x \in \Omega \cap P$.

Demonstração:

Considere $K_1 \underset{\infty}{\subset} K_0$ tal que $\lim_{k \in K_1} x^k = x^*$, como $a_i^{\min} - \varepsilon_k \leq x_i^k \leq a_i^{\max} + \varepsilon_k \quad \forall k, \forall i = 1, \ldots, n$ e $\varepsilon_k \to 0$, temos $a_i^{\min} \leq x_i^* \leq a_i^{\max} \quad \forall i = 1, \ldots, n$. Do mesmo modo, como $(a_i^k)^{\mathrm{T}} x^k - b_i^k \leq \varepsilon_k$ temos $a_i^{\mathrm{T}} x^* \leq b_i$ pela continuidade de $(z, w) \mapsto z^{\mathrm{T}} w$. Assim $x^* \in \Omega \cap P$.

Caso 1) Se $\{\rho_k\}$ é limitada temos $||h(x^k)||_\infty \to 0$ e $||V^k||_\infty \to 0$. Assim $h(x^*) = 0$ e supondo $g_i(x^*) > 0$ temos $g_i(x^k) > 0 \quad \forall k > k'$, como $-\frac{\bar{\mu}_i^k}{\rho_k} \leq 0$ temos $V_i^k = \max\{g_i(x^k), -\frac{\bar{\mu}_i^k}{\rho_k}\} = g_i(x^k) \quad \forall k > k'$. Logo $V_i^k \to g_i(x^*) = 0$, o que é uma contradição. Logo $g_i(x^*) \leq 0 \quad \forall i = 1, \ldots, p$.

Caso 2) Se $\{\rho_k\}$ é não limitada, então $\rho_k \to +\infty$. Da definição do algoritmo temos

$$\begin{aligned} \delta_k &= \nabla f(x^k) + \sum_{i=1}^{m} \left(\bar{\lambda}_i^k + \rho_k h_i(x^k)\right) \nabla h_i(x^k) + \sum_{i=1}^{p} \max\left(0, \bar{\mu}_i^k + \rho_k g_i(x^k)\right) \nabla g_i(x^k) \\ &\quad + \sum_{i=1}^{n} u_i^k(-e_i) + \sum_{i=1}^{n} v_i^k e_i + \sum_{i=1}^{r_k} w_i^k a_i^k \end{aligned}$$

com $u_i^k \geq 0, v_i^k \geq 0, w_i^k \geq 0$ e $||\delta_k|| \to 0$. Se $a_i^{\min} - x_i^* < 0$, temos $a_i^{\min} - x_i^k < -\varepsilon_k \quad \forall k > k_1$, assim $u_i^k = 0 \quad \forall k > k_1$. Analogamente $x_i^* - a_i^{\max} < 0 \Rightarrow v_i^k = 0 \quad \forall k > k_2$ e $a_i^{\mathrm{T}} x^* - b_i <$

$0 \Rightarrow w_i^k = 0 \quad \forall k > k_3$.

Seja $\mathrm{I}_1 = \{i \in \{1,\ldots,n\} | a_i^{\min} = x_i^*\}, \mathrm{I}_2 = \{i \in \{1,\ldots,n\} | x_i^* = a_i^{\max}\}, \mathrm{I}_3 = \{i \in \{1,\ldots,r\} | a_i^{\mathrm{T}} x^* - b_i = 0\}$. Assim $\forall k \in K_1, k > \max(k_1,k_2,k_3)$ temos

$$\begin{aligned} \delta_k &= \nabla f(x^k) + \sum_{i=1}^{m} \left(\bar{\lambda}_i^k + \rho_k h_i(x^k)\right) \nabla h_i(x^k) + \sum_{i=1}^{p} \max\left(0, \bar{\mu}_i^k + \rho_k g_i(x^k)\right) \nabla g_i(x^k) \\ &\quad + \sum_{i \in \mathrm{I}_1} u_i^k(-e_i) + \sum_{i \in \mathrm{I}_2} v_i^k e_i + \sum_{i \in \mathrm{I}_3} w_i^k a_i^k \end{aligned} \tag{4.1}$$

Dividindo (4.1) por ρ_k temos

$$\begin{aligned} \frac{\delta_k}{\rho_k} &= \frac{\nabla f(x^k)}{\rho_k} + \sum_{i=1}^{m} \left(\frac{\bar{\lambda}_i^k}{\rho_k} + h_i(x^k)\right) \nabla h_i(x^k) + \sum_{i=1}^{p} \max\left(0, \frac{\bar{\mu}_i^k}{\rho_k} + g_i(x^k)\right) \nabla g_i(x^k) \\ &\quad + \sum_{i \in \mathrm{I}_1} \frac{u_i^k}{\rho_k}(-e_i) + \sum_{i \in \mathrm{I}_2} \frac{v_i^k}{\rho_k} e_i + \sum_{i \in \mathrm{I}_3} \frac{w_i^k}{\rho_k} a_i^k \end{aligned} \tag{4.2}$$

Seja $S_k = \left|\left|\frac{(u^k, v^k)}{\rho_k}\right|\right|_\infty$ e $W_k = \left|\left|\frac{w^k}{\rho_k}\right|\right|_\infty$.

Caso 2.1) Se $\{S_k\}$ é limitada e $\{W_k\}$ é limitada, considere $K_2 \underset{\infty}{\subset} K_1$ tal que $\lim_{k \in K_2} \frac{(u^k, v^k)}{\rho_k} = (\hat{u}, \hat{v})$ e $\lim_{k \in K_2} \frac{w^k}{\rho_k} = \hat{w}$ com $\hat{u} \geq 0, \hat{v} \geq 0, \hat{w} \geq 0$. Tomando o limite de (4.2) em K_2 obtemos

$$\begin{aligned} 0 &= \sum_{i=1}^{m} h_i(x^*) \nabla h_i(x^*) + \sum_{i=1}^{p} \max\left(0, g_i(x^*)\right) \nabla g_i(x^*) \\ &\quad + \sum_{i \in \mathrm{I}_1} \hat{u}_i(-e_i) + \sum_{i \in \mathrm{I}_2} \hat{v}_i e_i + \sum_{i \in \mathrm{I}_3} \hat{w}_i a_i \end{aligned}$$

Assim x^* é um ponto KKT do problema de minimizar $||h(x)||_2^2 + ||\max(0, g(x))||_2^2$ sujeito a $x \in \Omega \cap P$.

Caso 2.2) Se $\{S_k\}$ é limitada e $\{W_k\}$ é não limitada, considere $K_2 \underset{\infty}{\subset} K_1$ tal que $\lim_{k \in K_2} \frac{(u^k, v^k)}{\rho_k} = (\hat{u}, \hat{v})$ e $\lim_{k \in K_2} W_k = +\infty$, assim existe $K_3 \underset{\infty}{\subset} K_2$ tal que $\lim_{k \in K_3} \frac{w^k/\rho_k}{W_k} = \hat{w}$ com $\hat{u} \geq 0, \hat{v} \geq 0$, $\hat{w} \geq 0$ e $\hat{w} \neq 0$ pois pelo menos uma coordenada é igual a 1. Dividindo

(4.2) por W_k e tomando o limite em K_3 obtemos

$$0 = \sum_{i \in \mathrm{I}_3} \hat{w}_i a_i$$

Contradizendo o fato de x^* ser MF de $\Omega \cap P$.

Caso 2.3) Se $\{S_k\}$ é não limitada e $\{W_k\}$ é não limitada, considere $K_2 \underset{\infty}{\subset} K_1$ tal que $\lim_{k \in K_2} S_k = +\infty$ e $\lim_{k \in K_2} W^k = +\infty$.

Caso 2.3.1) Se $\left\{\dfrac{W_k}{S_k}\right\}$ é não limitada, considere $K_3 \underset{\infty}{\subset} K_2$ tal que $\lim_{k \in K_3} \dfrac{W_k}{S_k} = +\infty$ e $\lim_{k \in K_3} \dfrac{w^k/\rho_k}{W_k} = \hat{w}$ com $\hat{w} \geq 0$ e $\hat{w} \neq 0$ pois pelo menos uma coordenada é igual a 1. Dividindo (4.2) por W_k e tomando o limite em K_3 obtemos

$$0 = \sum_{i \in \mathrm{I}_3} \hat{w}_i a_i$$

Contradizendo o fato de x^* ser MF de $\Omega \cap P$.

Caso 2.3.2) Se $\left\{\dfrac{W_k}{S_k}\right\}$ é limitada, considere $K_3 \underset{\infty}{\subset} K_2$ tal que $\lim_{k \in K_3} \dfrac{(u^k, v^k)/\rho_k}{S_k} = (\hat{u}, \hat{v})$, $\lim_{k \in K_3} \dfrac{w_k/\rho_k}{S_k} = \hat{w}$ com $\hat{u} \geq 0, \hat{v} \geq 0$, $\hat{w} \geq 0$ e $(\hat{u}, \hat{v}, \hat{w}) \neq 0$ pois pelo menos uma coordenada de $(\hat{u}, \hat{v})$ é igual a 1. Dividindo (4.2) por S_k e tomando o limite em K_3 obtemos

$$0 = \sum_{i \in \mathrm{I}_1} \hat{u}_i(-e_i) + \sum_{i \in \mathrm{I}_2} \hat{v}_i e_i + \sum_{i \in \mathrm{I}_3} \hat{w}_i a_i$$

Contradizendo o fato de x^* ser MF de $\Omega \cap P$. □

Lema 4.3. *Seja x^* um ponto limite da sequência $\{x^k\}$ gerada pelo algoritmo BFM e $K \underset{\infty}{\subset} \mathbb{N}$ tal que $\lim_{k \in K} x^k = x^*$. Se $g_i(x^*) < 0$ então existe $k' \in \mathbb{N}$ tal que $\bar{\mu}_i^k + \rho_k g_i(x^k) < 0 \quad \forall k \in K, k > k'$.*

Demonstração:

Como $g_i(x^*) < 0$, temos $g_i(x^k) < c < 0 \quad \forall k \in K, k > k'$. Se $\rho_k \to +\infty$, como $\{\bar{\mu}_i^k\}$ é limitada temos $\bar{\mu}_i^k + \rho_k g_i(x^k) < 0 \quad \forall k \in K, k > k'$. Se $\{\rho_k\}$ é limitada então

$V_i^k = \max\left\{g_i(x^k), -\dfrac{\bar{\mu}_i^k}{\rho_k}\right\} \to 0$, assim $\lim_{k\in K}\bar{\mu}_i^k = 0$, logo $\bar{\mu}_i^k + \rho_k g_i(x^k) < 0 \quad \forall k \in K, k > k'$. $\square$

Proposição 4.4. *Suponha que* $\{r_k\}, \{a_i^k\}, \{b_i^k\}$ *são limitadas para cada* i, *logo existe* $K_0 \underset{\infty}{\subset} \mathbb{N}$ *tal que* $r_k = r \quad \forall k \in K_0$, $\lim\limits_{k\in K_0} a_i^k = a_i$ *e* $\lim\limits_{k\in K_0} b_i^k = b_i$. *Defina* $P = \{x \in \mathbb{R}^n | a_i^T x \leq b_i, i = 1, \ldots, r\}$. *Seja* x^* *um ponto limite da sequência* $\{x^k\}$ *gerada pelo algoritmo BFM e suponha* $x^* \in X$. *Se* x^* *é um ponto Mangasarian-Fromovitz de* P *e* x^* *satisfaz a CPLD de* $X \cap \Omega \cap P$, *então* x^* *é um ponto KKT do problema de minimizar* $f(x)$ *sujeito a* $x \in X \cap \Omega \cap P$.

Demonstração:

Considere $K_1 \underset{\infty}{\subset} K_0$ tal que $\lim\limits_{k\in K_1} x^k = x^*$. Pela definição do algoritmo temos

$$\begin{aligned}\delta_k \;=\; & \nabla f(x^k) + \sum_{i=1}^{m}\left(\bar{\lambda}_i^k + \rho_k h_i(x^k)\right)\nabla h_i(x^k) + \sum_{i=1}^{p}\max\left(0, \bar{\mu}_i^k + \rho_k g_i(x^k)\right)\nabla g_i(x^k)\\ & + \sum_{i=1}^{n} u_i^k(-e_i) + \sum_{i=1}^{n} v_i^k e_i + \sum_{i=1}^{r_k} w_i^k a_i^k\end{aligned}$$

com $||\delta_k|| \to 0$. Se $a_i^{\min} - x_i^* < 0$, temos $a_i^{\min} - x_i^k < -\varepsilon_k \quad \forall k \in K_1, k > k_1$, assim $u_i^k = 0 \quad \forall k \in K_1, k > k_1$. Analogamente $x_i^* - a_i^{\max} < 0 \Rightarrow v_i^k = 0 \quad \forall k \in K_1, k > k_2$ e $a_i^{\mathrm{T}} x^* - b_i < 0 \Rightarrow w_i^k = 0 \quad \forall k \in K_1, k > k_3$. Pelo Lema 4.3, $g_i(x^*) < 0 \Rightarrow \bar{\mu}_i^k = 0 \quad \forall k \in K_1, k > k'$. Definindo $\lambda_i^{k+1} = \bar{\lambda}_i^k + \rho_k h_i(x^k)$ e $\mu^{k+1} = \max\left(0, \bar{\mu}_i^k + \rho_k g_i(x^k)\right)$ temos

$$\begin{aligned}\delta_k \;=\; & \nabla f(x^k) + \sum_{i=1}^{m}\lambda_i^{k+1}\nabla h_i(x^k) + \sum_{i=1}^{p}\mu_i^{k+1}\nabla g_i(x^k)\\ & + \sum_{i|a_i^{\min}=x_i^*} u_i^k(-e_i) + \sum_{i|x_i^*=a_i^{\max}} v_i^k e_i + \sum_{i|a_i^{\mathrm{T}}x^*-b_i=0} w_i^k a_i^k\end{aligned}$$

Com $\mu_i^{k+1} \geq 0, u_i^k \geq 0, v_i^k \geq 0, w_i^k \geq 0$.

Pelo Lema 1.23 de Carathéodory, existem $\mathrm{I}_1^k \subset \{1, \ldots, m\}, \mathrm{I}_2^k \subset \{i \in \{1, \ldots, p\} | g_i(x^*) = 0\}, \mathrm{I}_3^k \subset \{i \in \{1, \ldots, n\} | a_i^{\min} = x_i^*\}, \mathrm{I}_4^k \subset \{i \in \{1, \ldots, n\} | x_i^* = a_i^{\max}\}, \mathrm{I}_5^k \subset \{i \in \{1, \ldots, r\} | a_i^{\mathrm{T}} x^* - b_i = 0\}, \hat{\lambda}_i^k, \hat{\mu}_i^k \geq 0, \hat{u}_i^k \geq 0, \hat{v}_i^k \geq 0, \hat{w}_i^k \geq 0$, tais que

$$
\begin{aligned}
\delta_k &= \nabla f(x^k) + \sum_{i\in \mathrm{I}_1^k} \hat{\lambda}_i^k \nabla h_i(x^k) + \sum_{i\in \mathrm{I}_2^k} \hat{\mu}_i^k \nabla g_i(x^k) \\
&\quad + \sum_{i\in \mathrm{I}_3^k} \hat{u}_i^k(-e_i) + \sum_{i\in \mathrm{I}_4^k} \hat{v}_i^k e_i + \sum_{i\in \mathrm{I}_5^k} \hat{w}_i^k a_i^k
\end{aligned}
$$

e $\{\nabla h_i(x^k)\}_{i\in \mathrm{I}_1^k} \cup \{\nabla g_i(x^k)\}_{i\in \mathrm{I}_2^k} \cup \{-e_i\}_{i\in \mathrm{I}_3^k} \cup \{e_i\}_{i\in \mathrm{I}_4^k} \cup \{a_i^k\}_{i\in \mathrm{I}_5^k}$ linearmente independente.

Como o número de índices é finito, existe $K_2 \underset{\infty}{\subset} K_1$ tal que $\mathrm{I}_1^k = \mathrm{I}_1, \mathrm{I}_2^k = \mathrm{I}_2, \mathrm{I}_3^k = \mathrm{I}_3, \mathrm{I}_4^k = \mathrm{I}_4, \mathrm{I}_5^k = \mathrm{I}_5$. Logo para $k \in K_2$ temos

$$
\begin{aligned}
\delta_k &= \nabla f(x^k) + \sum_{i\in \mathrm{I}_1} \hat{\lambda}_i^k \nabla h_i(x^k) + \sum_{i\in \mathrm{I}_2} \hat{\mu}_i^k \nabla g_i(x^k) \\
&\quad + \sum_{i\in \mathrm{I}_3} \hat{u}_i^k(-e_i) + \sum_{i\in \mathrm{I}_4} \hat{v}_i^k e_i + \sum_{i\in \mathrm{I}_5} \hat{w}_i^k a_i^k
\end{aligned} \tag{4.3}
$$

e $\{\nabla h_i(x^k)\}_{i\in \mathrm{I}_1} \cup \{\nabla g_i(x^k)\}_{i\in \mathrm{I}_2} \cup \{-e_i\}_{i\in \mathrm{I}_3} \cup \{e_i\}_{i\in \mathrm{I}_4} \cup \{a_i^k\}_{i\in \mathrm{I}_5}$ linearmente independente.

Seja $S_k = \max\Big\{\max\{|\hat{\lambda}_i^k|, i \in \mathrm{I}_1\}, \max\{\hat{\mu}_i^k, i \in \mathrm{I}_2\}, \max\{\hat{u}_i^k, i \in \mathrm{I}_3\}, \max\{\hat{v}_i^k, i \in \mathrm{I}_4\}\Big\}$ e $W_k = \max\{\hat{w}_i^k, i \in \mathrm{I}_5\}$.

Caso 1) Se $\{S_k\}$ é limitada e $\{W_k\}$ é limitada, considere $K_3 \underset{\infty}{\subset} K_2$ tal que $\lim_{k\in K_3} (\hat{\lambda}^k, \hat{\mu}^k, \hat{u}^k, \hat{v}^k) = (\hat{\lambda}, \hat{\mu}, \hat{u}, \hat{v})$ e $\lim_{k\in K_3} \hat{w}^k = \hat{w}$ com $\hat{\mu} \geq 0, \hat{u} \geq 0, \hat{v} \geq 0, \hat{w} \geq 0$. Tomando o limite de (4.3) em K_3 obtemos

$$
\begin{aligned}
0 &= \nabla f(x^*) + \sum_{i\in \mathrm{I}_1} \hat{\lambda}_i \nabla h_i(x^*) + \sum_{i\in \mathrm{I}_2} \hat{\mu}_i \nabla g_i(x^*) \\
&\quad + \sum_{i\in \mathrm{I}_3} \hat{u}_i(-e_i) + \sum_{i\in \mathrm{I}_4} \hat{v}_i e_i + \sum_{i\in \mathrm{I}_5} \hat{w}_i a_i^k
\end{aligned}
$$

Assim x^* é um ponto KKT do problema de minimizar $f(x)$ sujeito a $x \in X \cap \Omega \cap P$.

Caso 2) Se $\{S_k\}$ é limitada e $\{W_k\}$ é não limitada, considere $K_3 \underset{\infty}{\subset} K_2$ tal que $\lim_{k\in K_3} W_k = +\infty$, assim existe $K_4 \underset{\infty}{\subset} K_3$ tal que $\lim_{k\in K_4} \dfrac{w^k}{W_k} = \hat{w}$ com $\hat{w} \geq 0$ e $\hat{w} \neq 0$ pois pelo menos uma coordenada é igual a 1. Dividindo (4.3) por W_k e tomando o limite em K_4 obtemos

$$0 = \sum_{i \in \mathrm{I}_5} \hat{w}_i a_i$$

Contradizendo o fato de x^* ser MF de P.

Caso 3) Se $\{S_k\}$ é não limitada e $\{W_k\}$ é não limitada, considere $K_3 \underset{\infty}{\subset} K_2$ tal que $\lim_{k \in K_3} S_k = +\infty$ e $\lim_{k \in K_3} W^k = +\infty$.

Caso 3.1) Se $\left\{\dfrac{W_k}{S_k}\right\}$ é não limitada, considere $K_4 \underset{\infty}{\subset} K_3$ tal que $\lim_{k \in K_4} \dfrac{S_k}{W_k} = 0$ e $\lim_{k \in K_4} \dfrac{w^k}{W_k} = \hat{w}$ com $\hat{w} \geq 0$ e $\hat{w} \neq 0$ pois pelo menos uma coordenada é igual a 1. Dividindo (4.3) por W_k e tomando o limite em K_4 obtemos

$$0 = \sum_{i \in \mathrm{I}_5} \hat{w}_i a_i$$

Contradizendo o fato de x^* ser MF de P.

Caso 3.2) Se $\left\{\dfrac{W_k}{S_k}\right\}$ é limitada, considere $K_4 \underset{\infty}{\subset} K_3$ tal que $\lim_{k \in K_4} \dfrac{(\hat{\lambda}^k, \hat{\mu}^k, \hat{u}^k, \hat{v}^k)}{S_k} = (\hat{\lambda}, \hat{\mu}, \hat{u}, \hat{v})$, $\lim_{k \in K_3} \dfrac{w_k}{S_k} = \hat{w}$ com $\hat{\mu} \geq 0, \hat{u} \geq 0, \hat{v} \geq 0$, $\hat{w} \geq 0$ e $(\hat{\lambda}, \hat{\mu}, \hat{u}, \hat{v}, \hat{w}) \neq 0$ pois pelo menos uma coordenada de $(\hat{\lambda}, \hat{\mu}, \hat{u}, \hat{v})$ em módulo é igual a 1. Dividindo (4.3) por S_k e tomando o limite em K_4 obtemos

$$\begin{aligned} 0 \;=\; & \sum_{i \in \mathrm{I}_1} \hat{\lambda}_i \nabla h_i(x^*) + \sum_{i \in \mathrm{I}_2} \hat{\mu}_i \nabla g_i(x^*) \\ & + \sum_{i \in \mathrm{I}_3} \hat{u}_i(-e_i) + \sum_{i \in \mathrm{I}_4} \hat{v}_i e_i + \sum_{i \in \mathrm{I}_5} \hat{w}_i a_i^k \end{aligned}$$

logo $(\{\nabla h_i(x^*)\}_{i \in \mathrm{I}_1}, \{\nabla g_i(x^*)\}_{i \in \mathrm{I}_2} \cup \{-e_i\}_{i \in \mathrm{I}_3} \cup \{e_i\}_{i \in \mathrm{I}_4} \cup \{a_i\}_{i \in \mathrm{I}_5})$ é PLD.
Mas $\left(\{\nabla h_i(x^k)\}_{i \in \mathrm{I}_1}, \{\nabla g_i(x^k)\}_{i \in \mathrm{I}_2} \cup \{-e_i\}_{i \in \mathrm{I}_3} \cup \{e_i\}_{i \in \mathrm{I}_4} \cup \{a_i\}_{i \in \mathrm{I}_5}\right)$ é PLI para todo $k \in K_4$, contradizendo o fato de x^* satisfazer a CPLD de $X \cap \Omega \cap P$. $\square$

4.2 Algoritmo de Chen e Goldfarb

Considere o problema

$$\begin{array}{ll} \text{Minimizar} & f(x) \\ \text{Sujeito a} & c(x) \geq 0 \\ & g(x) = 0, \end{array}$$

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

Considere a aplicação de penalidade interna a este problema, isto é, definimos o problema FP_μ abaixo

$$\begin{array}{ll} \text{Minimizar} & \varphi_\mu(x) = f(x) - \mu \sum_{i=1}^{m} \log(c_i(x)) \\ \text{Sujeito a} & c(x) > 0 \\ & g(x) = 0 \end{array},$$

e vamos resolvê-lo para uma sequência decrescente de parâmetros $\mu > 0$ tendendo a zero.

O algoritmo proposto em [13] consiste em resolver aproximadamente o subproblema FP_μ penalizando as restrições de igualdade. Isto é, para μ fixo, uma solução aproximada de FP_μ é obtida resolvendo o problema $l_2\mathrm{FP}_\mu$ abaixo

$$\begin{array}{ll} \text{Minimizar} & \Phi_{\mu,r}(x) = \varphi_\mu(x) + r||g(x)|| \\ \text{Sujeito a} & c(x) > 0 \end{array},$$

para parâmetros de penalidade r arbitrariamente grandes.

Nos pontos onde $\Phi_{\mu,r}(x)$ é diferenciável, isto é, $||g(x)|| > 0$, o sistema KKT do problema $l_2\mathrm{FP}_\mu$ pode ser escrito como

$$
\begin{aligned}
\nabla_x L(x,u,v) &= 0,\\
C(x)u &= \mu e,\\
c_i(x) > 0, u_i > 0, i = 1,\ldots,m&,\\
g_i(x) - v_i\delta_x = 0, i = 1,\ldots,p&,
\end{aligned}
$$

onde $C(x) = \mathrm{diag}(c(x))$, $e = (1,\ldots,1)$, $L(x,u,v) = f(x) - u^{\mathtt{T}}c(x) + v^{\mathtt{T}}g(x)$, $\delta_x = \frac{||g(x)||}{r}$.

O algoritmo de Chen e Goldfarb consiste em aplicar o método de Newton ao sistema acima, considerando δ_x constante.

Seja $\mathcal{H}$ uma aproximação para a Hessiana do Lagrangiano $\nabla^2_{xx}L(x,u,v)$, cada iteração do método de Newton (ou quase-Newton) consiste em resolver o sistema linear

$$
\begin{pmatrix} \mathcal{H} & -\nabla c(x) & \nabla g(x) \\ \mathcal{U}\nabla c(x)^{\mathtt{T}} & C(x) & 0 \\ \nabla g(x)^{\mathtt{T}} & 0 & -\delta_x \mathtt{I} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta u \\ \Delta v \end{pmatrix} = \begin{pmatrix} -\nabla f(x) + \nabla c(x)u - \nabla g(x)v \\ -C(x)u + \mu e \\ \delta_x v - g(x) \end{pmatrix},
$$

onde $\mathcal{U} = \mathrm{diag}(u)$ e $\mathtt{I}$ é a matriz identidade.

Definindo $\lambda = u + \Delta u, y = v + \Delta v$, podemos escrever

$$
\begin{pmatrix} \mathcal{H} & -\nabla c(x) & \nabla g(x) \\ \mathcal{U}\nabla c(x)^{\mathtt{T}} & C(x) & 0 \\ \nabla g(x)^{\mathtt{T}} & 0 & -\delta_x \mathtt{I} \end{pmatrix} \begin{pmatrix} \Delta x \\ \lambda \\ y \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ \mu e \\ -g(x) \end{pmatrix}, \tag{4.4}
$$

resolvendo para λ, obtemos $\lambda = C(x)^{-1}(\mu e - \mathcal{U}\nabla c(x)^{\mathtt{T}}\delta_x)$, reduzindo o sistema a

$$
M \begin{pmatrix} \Delta x \\ y \end{pmatrix} = \begin{pmatrix} -\nabla f(x) + \mu\nabla c(x)^{\mathtt{T}}C(x)^{-1}e \\ -g(x) \end{pmatrix}, \tag{4.5}
$$

onde

$$
M = \begin{pmatrix} \hat{\mathcal{H}} & \nabla g(x) \\ \nabla g(x)^{\mathtt{T}} & \delta_x \mathtt{I} \end{pmatrix}
$$

e

$$\hat{\mathcal{H}} = \mathcal{H} + \nabla c(x) C(x)^{-1} \mathcal{U} \nabla c(x)^{\mathtt{T}}.$$

A seguir daremos a definição precisa do algoritmo. Em [13], os autores mostram que sob MF para o problema original, os pontos limites são KKT para FP_{μ}, caso o parâmetro de penalidade r_k é limitado. Se o parâmetro de penalidade tende a infinito, existe um ponto limite que é Fritz-John para um certo problema de minimizar a inviabilidade, caso valha MF para o problema original.

Em [14], os autores mostram convergência superlinear de uma variação deste algoritmo sob LICQ, complementaridade estrita e uma condição suficiente de segunda ordem.

Fixado $\mu > 0$, o algoritmo a seguir é utilizado para a resolução de FP_{μ}. O algoritmo para resolver o problema original é obtido resolvendo FP_{μ} para uma sequência de parâmetros $\mu \to 0$, com $\varepsilon_{\mu} \to 0$ e utilizando a solução $(x, u, r, \mathcal{H})$ de FP_{μ} como aproximação inicial para o problema com o novo μ.

Algoritmo (Chen-Goldfarb)

Passo 0: **Inicialização**

Parâmetros $\varepsilon_{\mu} > 0, \sigma \in \left(0, \frac{1}{2}\right), \chi > 1, \kappa_1 \in (0,1), \kappa_2 > 1, \pi_{\mu} = \max\{\mu, 0.1\}, \nu > 0, 0 < \gamma_{min} < 1 < \gamma_{max}$.

Dados $x^0 \in \mathcal{F}^0 = \{x \in \mathbb{R}^n | c(x) > 0\}, u^0 > 0, r_0 > 0, \mathcal{H}^0 \in \mathbb{R}^{n \times n}$.

$k := 0$.

Passo 1: **Direção de busca**

Modificar $\mathcal{H}^k$, se necessário, para que valha a condição C-5

$$\begin{cases} d^{\mathtt{T}}\tilde{\mathcal{H}}^k d \geq \nu||d||^2, \forall d \neq 0 & \text{se } ||g(x^k)|| > 0 \\ d^{\mathtt{T}}\tilde{\mathcal{H}}^k d \geq \nu||d||^2, \forall d \neq 0, \nabla g(x^k)^{\mathtt{T}} d = 0 & \text{se } ||g(x^k)|| = 0 \end{cases},$$

onde

$$\tilde{\mathcal{H}}^k = \begin{cases} \hat{\mathcal{H}}^k & \text{se } ||g(x^k)|| = 0 \\ \hat{\mathcal{H}}^k + \frac{1}{\delta_{x^k}} \nabla g(x^k)^{\mathtt{T}} \nabla g(x^k) & \text{se } ||g(x^k)|| > 0 \end{cases},$$

com $\delta_{x^k} = \dfrac{||g(x^k)||}{r_k}, \hat{\mathcal{H}}^k = \mathcal{H}^k + \nabla c(x^k) C(x^k)^{-1} \mathcal{U}^k \nabla c(x^k)^{\mathtt{T}}, \mathcal{U}^k = \text{diag}(u^k), C(x^k) = \text{diag}(c(x^k))$.

Calcular $(\Delta x^k, \lambda^k, y^k)$ sendo $(\Delta x^k, y^k)$ solução do sistema (4.5) com índices k apropriados, e $\lambda^k = C(x^k)^{-1}(\mu e - \mathcal{U}^k \nabla c(x^k)^{\mathtt{T}} \delta_{x^k})$.

Passo 2: **Teste de parada**

Parar se $||\mathcal{R}_\mu(x^k, \lambda^k, y^k)|| \leq \varepsilon_\mu$ e $\lambda^k \geq -\varepsilon_\mu e$, onde

$$\mathcal{R}_\mu(x^k, \lambda^k, y^k) = \begin{pmatrix} \nabla_x L(x^k, \lambda^k, y^k) \\ C(x^k)\lambda^k - \mu e \\ g(x^k) \end{pmatrix}.$$

Passo 3: **Atualização do parâmetro de penalidade**

Se valem

(C-1): $||g(x^k)|| > 0$
(C-2): $||\Delta x^k|| \leq \pi_\mu$
(C-3): $\kappa_1 \mu e \leq C(x^k)\lambda^k \leq \kappa_2 \mu e$
(C-4): $||\bar{w}^k|| < \pi_\mu, \bar{w}^k = y^k - \dfrac{r_k}{||g(x^k)||} g(x^k)$

então faça

$r_{k+1} := \chi r_k$
$(x^{k+1}, u^{k+1}, \mathcal{H}^{k+1}) := (x^k, u^k, \mathcal{H}^k)$

$k := k + 1$
e volte para o Passo 1.

Passo 4: **Busca linear**

incializar com $t_k = 1$ e dividir por 2, se necessário, até que valha

T-1: $c(x^k + t_k \Delta x^k) > 0$
T-2: $\Phi_{\mu,r_k}(x^k + t_k \Delta x^k) - \Phi_{\mu,r_k}(x^k) \leq -\sigma t_k (\Delta x^k)^{\mathtt{T}} \tilde{\mathcal{H}} \Delta x^k$,

onde $\Phi_{\mu,r} = \varphi_\mu(x) + r||g(x)||$ e $\varphi_\mu(x) = f(x) - \mu \sum_{i=1}^{m} \log(c_i(x))$.

Passo 5: **Atualização**

Defina u_i^{k+1} a projeção de λ_i^k no intervalo $[\mu \frac{\gamma_{\min}}{c_i(x^k)}, \mu \frac{\gamma_{\max}}{c_i(x^k)}]$, para cada i.
$x^{k+1} := x^k + t_k \Delta x^k$
$r_{k+1} := r_k$
Calcular a nova Hessiana do lagrangiano ou sua estimativa $\mathcal{H}^{k+1}$.
$k := k + 1$
voltar para o Passo 1.

Mostraremos alguns resultados presentes no artigo [13], utilizando sempre que possível a condição CPLD ao invés de MF.

O lema a seguir foi provado sem assumir nenhuma condição de qualificação. No artigo os autores utilizam MF para obter este resultado. O caso em que $\delta_x = 0$ pode ser encontrado em [39], mas o incluímos aqui por completude.

Lema 4.5. *Se $x \in \mathbb{R}^n, c(x) > 0, \mu > 0, r > 0, \mu > 0$ e vale C-5 então o sistema linear (4.5) possui solução.*

Demonstração: Se $\delta_x > 0$, por C-5, $\hat{\mathcal{H}} + \frac{1}{\delta_x} \nabla g(x) \nabla g(x)^{\mathtt{T}}$ é definida positiva e pelo Lema 3.1 em [13] M é não singular e o sistema admite solução (única).

Se $\delta_x = 0$, temos $g(x) = 0$ pela definição de δ_x. Assim o sistema (4.5) se resume a

$$\begin{pmatrix} \hat{\mathcal{H}} & \nabla g(x) \\ \nabla g(x)^{\mathtt{T}} & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ y \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix},$$

com $b = -\nabla f(x) + \mu \nabla c(x)^{\mathtt{T}} C(x)^{-1} e$.

Ou seja,

$$\begin{aligned} \hat{\mathcal{H}}\Delta x + \nabla g(x) y &= b \\ \nabla g(x)^{\mathtt{T}} \Delta x &= 0 \end{aligned}$$

Se $\mathrm{Ker}(\nabla g(x)^{\mathtt{T}}) = \{0\}$ então $\Delta x = 0$ e o sistema linear se reduz a $\nabla g(x) y = b$.

Como $\mathrm{Im}(\nabla g(x)) = \mathrm{Ker}(\nabla g(x)^{\mathtt{T}})^{\perp} = \mathbb{R}^n$, então $b \in \mathrm{Im}(\nabla g(x))$ e portanto o sistema admite solução.

Se $\dim(\mathrm{Ker}(\nabla g(x)^{\mathtt{T}})) = r > 0$, seja $Z = [z_1, \ldots, z_r]$ uma matriz cujas colunas formam uma base para $\mathrm{Ker}(\nabla g(x)^{\mathtt{T}})$. Defina $\Delta x = Z(Z^{\mathtt{T}}\hat{\mathcal{H}}Z)^{-1}Z^{\mathtt{T}}b$, já que por C-5, $\hat{\mathcal{H}}$ é definida positiva em $\mathrm{Ker}(\nabla g(x)^{\mathtt{T}})$ e portanto $Z^{\mathtt{T}}\hat{\mathcal{H}}Z$ é invertível. Assim, $\Delta x \in \mathrm{Ker}(\nabla g(x)^{\mathtt{T}})$, portanto $\nabla g(x)^{\mathtt{T}}\Delta x = 0$.

Para mostrarmos que existe solução y para $\hat{\mathcal{H}}\Delta x + \nabla g(x) y = b$, mostremos que $b - \hat{\mathcal{H}}\Delta x \in \mathrm{Ker}(\nabla g(x)^{\mathtt{T}})^{\perp} = \mathrm{Im}(\nabla g(x))$. Para isso, basta observar que $Z^{\mathtt{T}}(b - \hat{\mathcal{H}}\Delta x) = Z^{\mathtt{T}}b - (Z^{\mathtt{T}}\hat{\mathcal{H}}Z)(Z^{\mathtt{T}}\hat{\mathcal{H}}Z)^{-1}Z^{\mathtt{T}}b = 0$, assim, $z_i^{\mathtt{T}}(b - \hat{\mathcal{H}}\Delta x) = 0$ para todo $i = 1, \ldots, r$ e o resultado segue. $\square$

Vamos assumir que a sequência de Hessianas $\{\mathcal{H}^k\}$ é limitada.

Proposição 4.6. *Se o parâmetro de penalidade $r_k \to +\infty$ e se x^* é um ponto limite do* *algoritmo com $||g(x^*)|| > 0$, tal que vale a CPLD para*

$$\begin{array}{ll} \textit{Minimizar} & ||g(x)||^2 \\ \textit{Sujeito a} & c(x) \geq 0, \end{array}$$

então x^* é KKT para este problema.

Demonstração: Vamos considerar uma subsequência $\{x^k\}$ tal que $x^k \to x^*$ e r_k é aumentado para todo k, assim, valem as condições C-1 até C-4, portanto, utilizando o sistema linear (4.4) com índices k apropriados temos

$$-\mathcal{H}^k \Delta x^k + \sum_{i=1}^{m} \lambda_i^k \nabla c_i(x^k) - \sum_{i=1}^{p} (r_k \frac{g_i(x^k)}{||g(x^k)||} + \bar{w}_i^k) \nabla g_i(x^k) = \nabla f(x^k).$$

Por C-3, $0 < \kappa_1 \mu \leq c_i(x^k)\lambda_i^k$, portanto $\lambda_i^k > 0$ já que $c_i(x^k) > 0$. Pelo Lema 1.23 de Carathéodory, existe $\mathtt{I}_k \subset \{1, \ldots, m\}, \bar{\lambda}_i^k > 0$ tais que

$$\sum_{i \in \mathtt{I}_k} \bar{\lambda}_i^k \nabla c_i(x^k) - \sum_{i=1}^{p} \frac{r_k}{||g(x^k)||} g_i(x^k) \nabla g_i(x^k) = \nabla f(x^k) + \mathcal{H}^k \Delta x^k + \sum_{i=1}^{p} \bar{w}_i^k \nabla g_i(x^k), \quad (4.6)$$

com $\{\nabla c_i(x^k)\}_{i \in \mathtt{I}_k}$ LI.

Vamos tomar uma subsequência tal que $\mathtt{I}_k = \mathtt{I}$. Como $\lambda_i^k \leq \dfrac{\kappa_2 \mu}{c_i(x^k)}$, pela limitação dos multiplicadores dados pelo Lema de Carathéodory temos $\bar{\lambda}_i^k \leq \dfrac{2^{m-1}\kappa_2 \mu}{c_i(x^k)}$, portanto $0 < c_i(x^k)\bar{\lambda}_i^k \leq 2^{m-1}\kappa_2 \mu$.

Se $\left\{\dfrac{\bar{\lambda}^k}{r_k}\right\}$ é limitada considere uma subsequência tal que $\dfrac{\bar{\lambda}^k}{r_k} \to \lambda'$. Dividindo (4.6) por r_k e tomando o limite, observando que $\{\Delta x^k\}$ e $\{\bar{w}^k\}$ são limitadas por C-2 e C-4, temos

$$\sum_{i=1}^{p} \frac{g_i(x^*)}{||g(x^*)||} \nabla g_i(x^*) - \sum_{i \in \mathtt{I}} \lambda_i' \nabla c_i(x^*) = 0$$

e $0 < c_i(x^k)\dfrac{\bar{\lambda}_i^k}{r_k} \leq \dfrac{2^{m-1}\kappa_2 \mu}{r_k} \Rightarrow c_i(x^*)\lambda_i' = 0$, portanto x^* é KKT para minimizar

$||g(x)||^2$ sujeito a $c(x) \geq 0$.

Caso $\dfrac{\bar{\lambda}^k}{r_k} \to +\infty$, dividindo (4.6) $||\bar{\lambda}^k||_\infty$ e tomando o limite para uma subsequência tal que $\dfrac{\bar{\lambda}^k}{||\bar{\lambda}^k||_\infty} \to \bar{\lambda} \geq 0, \bar{\lambda} \neq 0$ temos

$$\sum_{i \in \mathtt{I}} \bar{\lambda}_i \nabla c_i(x^*) = 0,$$

com $0 < c_i(x^k) \dfrac{\bar{\lambda}_i^k}{||\bar{\lambda}^k||_\infty} \leq \dfrac{2^{m-1}\kappa_2 \mu}{||\bar{\lambda}^k||_\infty} \Rightarrow c_i(x^*)\bar{\lambda}_i = 0.$

Excluindo de $\mathtt{I}$ os índices com $\bar{\lambda}_i = 0$, temos $\mathtt{I} \subset \mathtt{I}(x^*)$ e não vale CPLD com respeito às restrições de desigualdade. □

Agora vamos obter um resultado quanto ao algoritmo para resolver o problema original. Vamos considerar uma sequência de números positivos $\mu_k \to 0$ e $\varepsilon_k \to 0$, e aplicamos o algoritmo de Chen e Goldfarb para resolver FP_μ, obtendo (x^k, λ^k, y^k) satisfazendo o critério de parada do Passo 2, $||\mathcal{R}_{\mu_k}(x^k, \lambda^k, y^k)|| \leq \varepsilon_k$.

Proposição 4.7. *Se o problema original satisfaz CPLD, e o algoritmo interno pra resolver FP_{μ_k} está bem definido, então todo ponto limite é KKT.*

Demonstração: Seja x^* um ponto limite e considere $\{x^k\}$ gerada pelo algoritmo com $x^k \to x^*$. Pelo critério de parada do passo 2, temos

$$\nabla f(x^k) - \sum_{i=1}^{m} \lambda_i^k \nabla c_i(x^k) + \sum_{i=1}^{p} y_i^k \nabla g_i(x^k) = \delta_1^k \tag{4.7}$$

$$C(x^k)\lambda^k - \mu_k e = \delta_2^k \tag{4.8}$$

$$g(x^k) = \delta_3^k, \tag{4.9}$$

com $||(\delta_1^k, \delta_2^k, \delta_3^k)|| \leq \varepsilon_k$ e $\lambda_i^k \geq -\varepsilon_k$.

Pelo Lema 1.23 de Carathéodory, existem $\bar{\lambda}_i^k$, $\bar{y}_i^k$, $\mathtt{I}_k \subset \{1, \ldots, m\}, \mathtt{J}_k \subset \{1, \ldots, p\}$ (vamos tomar uma subsequência tal que $\mathtt{I}_k = \mathtt{I}$ e $\mathtt{J}_k = \mathtt{J}$) tais que

$$\nabla f(x^k) - \sum_{i\in\mathtt{I}} \bar{\lambda}_i^k \nabla c_i(x^k) + \sum_{i\in\mathtt{J}} \bar{y}_i^k \nabla g_i(x^k) = \delta_1^k,$$

com $\{\nabla c_i(x^k)\}_{i\in\mathtt{I}} \cup \{\nabla g_i(x^k)\}_{i\in\mathtt{J}}$ LI e $|\bar{\lambda}_i^k| \leq 2^{m-1}|\lambda_i^k|$, portanto $\bar{\lambda}_i^k \geq -2^{m-1}\varepsilon_k$. Defina $\alpha_k = ||(\bar{\lambda}^k, \bar{y}^k)||_\infty$.

Se $\{\alpha_k\}$ é limitada, considere uma subsequência tal que $(\bar{\lambda}^k, \bar{y}^k) \to (\bar{\lambda}, \bar{y})$. Como $\varepsilon_k \to 0$, tomando o limite em (4.7) temos

$$\nabla f(x^*) - \sum_{i\in\mathtt{I}} \bar{\lambda}_i \nabla c_i(x^*) + \sum_{i\in\mathtt{J}} \bar{y}_i \nabla g_i(x^*).$$

Como $\bar{\lambda}_i^k \geq -2^{m-1}\varepsilon_k$ temos $\bar{\lambda} \geq 0$, além disso, de (4.8) temos $|\bar{\lambda}_i^k| \leq 2^{m-1}|\lambda_i^k| \leq \dfrac{|[\delta_2^k]_i + \mu_k|}{c_i(x^k)} \Rightarrow \bar{\lambda} c_i(x^*) = 0$. Por (4.9) temos $g(x^*) = 0$, portanto x^* é KKT para o problema original.

Se $\alpha_k \to +\infty$ considere uma subsequência tal que $\left(\dfrac{\bar{\lambda}_i^k}{\alpha_k}, \dfrac{\bar{y}_i^k}{\alpha_k}\right) \to (\hat{\lambda}, \hat{y}) \neq 0$, além disso, como $\dfrac{\bar{\lambda}_i^k}{\alpha_k} \geq -\dfrac{2^{m-1}\varepsilon_k}{\alpha_k} \to 0$, temos $\hat{\lambda} \geq 0$.

Dividindo (4.7) por α_k e tomando o limite temos

$$\sum_{i\in\mathtt{I}} \hat{\lambda}_i \nabla c_i(x^*) - \sum_{i\in\mathtt{J}} \hat{y}_i \nabla g_i(x^*) = 0,$$

com $|\bar{\lambda}_i^k| \leq 2^{m-1}|\lambda_i^k| \leq 2^{m-1}\dfrac{|[\delta_2^k]_i + \mu_k|}{c_i(x^k)}$, assim, multiplicando esta desigualdade por $\dfrac{c_i(x^k)}{\alpha_k}$ e tomando o limite temos $c_i(x^*)\hat{\lambda}_i = 0$, portanto removendo de $\mathtt{I}$ os índices tal que $\hat{\lambda}_i = 0$, temos $\mathtt{I} \subset \mathtt{I}(x^*)$, o que contradiz a CPLD. $\square$

Apêndice A
Desigualdades generalizadas

Queremos descobrir quais são as propriedades de um conjunto arbitrário $K \subset X$, onde X é um espaço de Banach real, tal que as definições abaixo façam sentido quando $x, y \in X$

$$x \leq_K y \Leftrightarrow y - x \in K$$

e

$$x <_K y \Leftrightarrow y - x \in \text{int}K,$$

onde int$K$ é o interior de K. Primeiramente, a relação $\leq_K$ deve ser uma relação de ordem parcial, isto é, reflexiva, anti-simétrica e transitiva.

Vejamos que se K é um cone, isto é, $tx \in K$ sempre que $x \in K$ e $t \geq 0$, temos que $0 \in K$, de modo que temos a reflexividade $x \leq_K x$ para todo x. Além disso, um cone nos garante propriedades importantes como $x \leq_K y \Rightarrow tx \leq_K ty, t \geq 0$.

Para obtermos a transitividade, isto é, $x \leq_K y$ e $y \leq_K z \Rightarrow x \leq_K z$, devemos ter que o conjunto é fechado com relação a soma de seus elementos, assim, $z-x = (y-x)+(z-y) \in K$. Mas como K é um cone, esta propriedade é equivalente ao fato que $t_1x+t_2y \in K$ sempre que $x, y \in K$ e $t_1 \geq 0, t_2 \geq 0$, o que também é equivalente ao fato do cone ser convexo.

Para obtermos a anti-simetria, isto é, $x \leq_K y$ e $y \leq_K x \Rightarrow x = y$, vamos supor que K não contém linhas, isto é, se $0 \neq x \in K$, então $-x \notin K$. Em outras palavras, $x \in K$ e $-x \in K \Rightarrow x = 0$. Neste caso dizemos que o cone K é pontudo.

Vamos supor também que K é fechado, assim podemos tomar limites nas desigualdades, isto é, se $x^k \leq_K y^k$ com $x^k \to x$ e $y^k \to y$ então $x \leq_K y$. Além disso, vamos supor que int$K \neq \emptyset$ para que a desigualdade estrita $<_K$ faça sentido, neste caso dizemos que K é sólido.

Quando um cone K é convexo, fechado, sólido e pontudo, dizemos que K é um cone próprio.

É possível construir uma teoria muito rica de otimização utilizando desigualdades definidas por cones próprios, sendo possível demonstrar os principais resultados como a existência de multiplicadores de Lagrange, teoremas de alternativas [46], relações com o problema dual, etc. Além disso, muitos problemas clássicos podem ser escritos como problemas de otimização com desigualdades generalizadas, é o caso do problema de controle ótimo, cálculo variacional e otimizaçao semi-definida, onde as restrições incluem que certas matrizes devem ser semi-definida positiva.

Observe que quando $K = \mathbb{R}^p_+$, temos as desigualdades usuais em $\mathbb{R}^p$, a saber

$$x \leq_K y \Leftrightarrow x_i \leq y_i, \forall i = 1, \ldots, p.$$

O leitor interessado pode encontrar desenvolvimentos detalhados deste tópico em [12, 15, 33].

Referências Bibliográficas

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

[2] W. Achtziger and C. Kanzow. Mathematical programs with vanishing constraints: optimality conditions and constraint qualifications. *Mathematical Programming*, 114:69–99, 2007.

[3] 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.

[4] 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.

[5] R. Andreani, C.R. Echagüe, and M.L. Schuverdt. The constant rank condition is a second order constraint qualification. *Optimization Online*, Feb,2009. http://www.optimizationonline.org/DB_HTML/2009/02/2230.html.

[6] R. Andreani and J.M. Martínez. On the solution of mathematical programming problemas with equilibrium constraints. *Mathematical Methods of Operations Research*, 54:345–358, 2001.

[7] R. Andreani, J.M. Martínez, and M.L. Schuverdt. On the relation between constant positive linear dependence condition and quasinormality constraint qualification. *Journal of Optimization Theory and Applications*, 125:473–485, 2005.

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

[9] D.P. Bertsekas. *Nonlinear Programming: 2nd Edition*. Athena Scientific, 1999.

[10] E.G. Birgin, C.A. Floudas, and J.M. Martínez. Global minimization using an augmented lagrangian method with variable lower-level constraints. *Mathematical Programming*, 2008. (Por aparecer).

[11] E.G. Birgin and J.M. Martínez. Local convergence of an inexact-restoration method and numerical experiments. *Journal of Optimization Theory and Applications*, 127:229–247, 2005.

[12] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[13] L. Chen and D. Goldfarb. Interior–point l2–penalty methods for nonlinear programming with strong global convergence. *Mathematical Programming*, 108(1):1–36, 2006.

[14] L. Chen and D. Goldfarb. On the fast local convergence of interior–point l2–penalty methods for nonlinear programming. Technical report, CORC Columbia University, 2006.

[15] B.D. Craven. *Control and Optimization*. Chapman & Hall, 1998.

[16] A.V. Fiacco and G.P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. Wiley, 1968. (reprinted, SIAM, 1990).

[17] A. Fischer and A. Friedlander. A new line search inexact restoration approach for nonlinear programming. *Computational Optimization and Applications*, 2009. (Por aparecer).

[18] R. Fletcher, N.I.M. Gould, S. Leyffer, P.L. Toint, and A. Wachter. Global convergence of a trust–region sqp–filter algorithm for general nonlinear programming. *SIAM Journal on Optimization*, 10(4):963–981, 2000.

[19] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical Programming*, 91(2):239–269, 2002.

[20] R. Fletcher, S. Leyffer, and P.L. Toint. On the global convergence of a filter–sqp algorithm. *SIAM Journal on Optimization*, 13(1):44–59, 2002.

[21] A. Forsgren, P.E. Gill, and M.H. Wright. Interior methods for nonlinear optimization. *SIAM Review*, 44(4):525–597, 2002.

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

[23] F.A.M. Gomes. A sequential quadratic programming algorithm that combines merit function and filter ideas. *Computational & Applied Mathematics*, 26(3):337–379, 2007.

[24] M.A. Gomes-Ruggiero, J.M. Martínez, and S.A. Santos. Spectral projected gradient method with inexact restoration for minimization with nonconvex constraints. *SIAM Journal on Scientific Computing*, 31(3):1628–1652, 2009.

[25] C.C. Gonzaga, E.W. Karas, and M. Vanti. A globally convergent filter method for nonlinear programming. *SIAM Journal on Optimization*, 14(3):646–669, 2003.

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

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

[28] M.R. Hestenes. *Optimization theory: the finite dimensional case*. John Wiley, 1975.

[29] T. Hoheisel and C. Kanzow. First- and second-order optimality conditions for mathematical programs with vanishing constraints. *Applications of Mathematics*, 52:495–514, 2007.

[30] T. Hoheisel and C. Kanzow. Stationary conditions for mathematical programs with vanishing constraints using weak constraint qualifications. *Journal of Mathematical Analysis and Applications*, 337:292–310, 2008.

[31] T. Hoheisel and C. Kanzow. On the Abadie and Guignard constraint qualifications for mathematical programs with vanishing constraints. *Optimization*, 2009. (Por aparecer).

[32] A.F. Izmailov and M.V. Solodov. Mathematical programs with vanishing constraints: optimality conditions, sensitivity, and a relaxation method. *Journal of Optimization Theory and Applications*, 2009. (Por aparecer).

[33] J. Jahn. *Vector Optimization – Theory, Applications and Extensions*. Springer, Berlin, 2004.

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

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

[36] D. Kinderlehrer and G. Stampacchia. *An Introduction to Variational Inequalities and their Applications*. SIAM Classics in Applied Mathematics, 31, 1987.

[37] 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.

[38] G. Lan and R.D.C. Monteiro. Iteration-complexity of first-order penalty methods for convex programming. *Mathematical Programming*, 2008. (submetido).

[39] D.G. Luenberger and Y. Ye. *Linear and Nonlinear Programming.* Springer, third edition edition, 2007.

[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. Inexact restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. *Journal of Optimization Theory and Applications*, 111:39–58, 2001.

[42] J.M. Martínez. *Otimização prática usando o lagrangiano aumentado (Opúsculo).* 2006. www.ime.unicamp.br/~martinez/lagraum.pdf.

[43] J.M. Martínez and E.A. Pilotta. Inexact restoration algorithms for constrained optimization. *Journal of Optimization Theory and Applications*, 104:135–163, 2000.

[44] 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.

[45] J.M. Martínez and B.F. Svaiter. A practical optimality condition without constraint qualifications for nonlinear programming. *Journal of Optimization Theory and Applications*, 118:117–133, 2003.

[46] T.S. Motzkin. Two consequences of the transposition theorem of linear inequalities. *Econometrica*, 19:184–185, 1951.

[47] Y. Nesterov. *Introductory Lectures on Convex Optimization.* Kluwer, 2004.

[48] Y. Nesterov. Gradient methods for minimizing composite objective functions. Technical Report 2007076, CORE Discussion Papers, 2007.

[49] A.E. Ozdaglar. *Pseudonormality and a Lagrange Multiplier Theory for Constrained Optimization.* PhD thesis, MIT, EECS Department, 2003. http://web.mit.edu/asuman/www/documents/Thesis_final.pdf.

[50] A.E. Ozdaglar and D.P. Bertsekas. The relation between pseudonormality and quasiregularity in constrained optimization. *Optimization Methods and Software*, 19:493–506, 2004.

[51] 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.

[52] S.M. Robinson. Stability theory for systems of inequalities. part I: Linear systems. *SIAM Journal on Numerical Analysis*, 12(5):754–769, 1975.

[53] S.M. Robinson. Stability theory for systems of inequalities. part II: Differentiable nonlinear systems. *SIAM Journal on Numerical Analysis*, 13(4):497–513, 1976.

[54] S.M. Robinson. Generalized equations and their solutions, part I: Basic theory. *Mathematical programming Study*, 10:128–141, 1979.

[55] S.M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5(1):43–62, 1980.

[56] S.M. Robinson. Generalized equations and their solutions, part II: Applications to nonlinear programming. *Mathematical programming Study*, 19:200–221, 1982.

[57] R.T. Rockafellar. The multiplier method of Hestenes and Powell applied to convex programming. *Journal of optimization Theory and Applications*, 12(6):555–562, 1973.

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

[59] M.L. Schuverdt. *Métodos de Lagrangiano aumentado com convergência utilizando a condição de dependência linear positiva constante.* PhD thesis, IMECC-UNICAMP, Departamento de Matemática Aplicada, 2006. http://libdigi.unicamp.br/document/?code=vtls000375801.

[60] C. Shen, W. Xue, and D. Pu. A filter SQP algorithm without a feasibility restoration phase. *Computational & Applied Mathematics*, 2009. (Por aparecer).

[61] M. Slater. Lagrange multipliers revisited: A contribution to non-linear programming. *Cowles Commission Discussion Paper: Mathematics*, 403, 1950.

[62] J.E. Spingarn and R.T. Rockafellar. The generic nature of optimality conditions in nonlinear programming. *Mathematics of Operations Research*, 4(4):425–430, 1979.

[63] D.J. Uherka and A.M. Sergott. On the continuous dependence of the roots of a polynomial on its coefficients. *The American Mathematical Monthly*, 84(5):368–370, 1977.