



Universidade Federal do Espírito Santo
Centro Universitário Norte do Espírito Santo
Curso de Bacharelado em Matemática Industrial

Thiago Silva Mantegazini

**UMA CONDIÇÃO SEQUENCIAL DE
OTIMALIDADE EM OTIMIZAÇÃO NÃO LINEAR
E SUA APLICAÇÃO A UM MÉTODO
LAGRANGIANO AUMENTADO**

São Mateus

2021

Thiago Silva Mantegazini

**UMA CONDIÇÃO SEQUENCIAL DE
OTIMALIDADE EM OTIMIZAÇÃO NÃO LINEAR
E SUA APLICAÇÃO A UM MÉTODO
LAGRANGIANO AUMENTADO**

Trabalho submetido ao Colegiado do Curso de Bacharelado em Matemática Industrial da UFES (Campus São Mateus), como requisito parcial para a obtenção do grau de Bacharel em Matemática.

São Mateus

2021

Thiago Silva Mantegazini

**UMA CONDIÇÃO SEQUENCIAL DE
OTIMALIDADE EM OTIMIZAÇÃO NÃO LINEAR
E SUA APLICAÇÃO A UM MÉTODO
LAGRANGIANO AUMENTADO**

Trabalho submetido ao Colegiado do Curso de Bacharelado em Matemática Industrial da UFES (Campus São Mateus), como requisito parcial para a obtenção do grau de Bacharel em Matemática.

Aprovada em 13 de Maio de 2021.

Comissão Examinadora

Prof. Leonardo Delarmelina Secchin
Universidade Federal do Espírito Santo
Orientador

Prof. André Renato Sales Amaral
Universidade Federal do Espírito Santo

Prof. Isaac Pinheiro dos Santos
Universidade Federal do Espírito Santo

Resumo

No trabalho em questão, consideramos o problema de programação não linear geral, bem como a ferramenta essencial para resolvê-lo: as condições de Karush-Kuhn-Tucker (KKT). Na prática numérica, métodos iterativos de otimização testam a validade das condições KKT apenas de maneira aproximada. Essa característica é inerente à inexatidão dos cálculos feitos por computador. A fim de descrever teoricamente essa inexatidão, foram definidas na literatura formas de aproximar KKT. Essas aproximações são conhecidas como condições sequenciais de otimalidade. Neste trabalho, estudamos os chamados pontos Aproximadamente KKT (AKKT). Argumentamos que pontos AKKT são de certa forma candidatos razoáveis à solução de um problema de otimização com restrições, além de estarem associados naturalmente à alguns métodos populares na literatura, como por exemplo, o método do Lagrangiano Aumentado.

Palavras-chave: Karush-Kuhn-Tucker, condições sequenciais de otimalidade, Lagrangiano aumentado.

Lista de Símbolos

\mathbb{R}	conjunto dos números reais.
\mathbb{R}_+	conjunto dos números reais não negativos.
∇r	gradiente de r .
$\ \mathbf{u}\ $	norma Euclideana de \mathbf{u} .
$\ \mathbf{u}\ _\infty$	norma do máximo de $\mathbf{u} \in \mathbb{R}^n$, isto é, $\ \mathbf{u}\ _\infty = \max\{ u_1 , \dots, u_n \}$.
a_+	$\max\{a, 0\}$.
$\mathbf{g}(\mathbf{x})_+$	$(g_1(\mathbf{x})_+, \dots, g_p(\mathbf{x})_+) \in \mathbb{R}_+^p$.

Sumário

1	Introdução	6
1.1	Objetivo do trabalho	7
2	Revisão de conceitos	8
2.1	O problema de programação não linear (PNL)	9
3	Condição de Karush-Kuhn-Tucker	11
3.1	Penalização externa	11
3.2	Condições de Karush-Kuhn-Tucker (KKT)	14
4	Condições sequenciais de otimalidade	20
4.1	A condição Aproximadamente KKT (AKKT)	20
4.2	Da condição sequencial para KKT	26
4.3	Outras condições sequenciais de otimalidade na literatura	28
5	Algoritmos: convergência global via condições sequenciais de otimalidade	29
5.1	Lagrangiano Aumentado	29
5.2	Notas	36
	Conclusões	37
	Referências Bibliográficas	38

1 Introdução

Processos em diversos campos da ciência requerem a resolução de problemas envolvendo minimização de funções com restrições. Métodos práticos para resolução de tais problemas são, em geral, iterativos. Decidir se um dado ponto é solução de um problema de otimização não linear geral pode ser uma tarefa difícil, talvez impraticável. Logo, condições que fornecem candidatos à solução são usadas, inclusive como critério de parada dos algoritmos. Para tanto, as principais ferramentas são as condições de Karush-Kuhn-Tucker (KKT) [1]. Essas condições generalizam o fato de, em otimização sem restrições, minimizadores possuem gradientes nulos. Essa última condição é comumente vista em cursos de cálculo.

As equações matemáticas que modelam processos são na maioria das vezes difíceis de serem manipuladas algebricamente. Daí a necessidade em conhecer métodos de solução computacional. Identificar um bom candidato a solução, nesses casos, é muito difícil. Isso requer do método parar quando um sistema de equações e inequações que caracteriza de forma razoável os minimizadores locais do problema sejam satisfeitas apenas aproximadamente. De fato, vários métodos testam se o candidato à solução satisfazem as condições de Karush-Kuhn-Tucker (KKT), que serão apresentadas no próximo capítulo, de forma aproximada. Veja a discussão em [2, 3]. No entanto, há diferenças entre uma solução ser de fato KKT e satisfazer KKT aproximadamente. Felizmente, esta prática computacional pode ser justificada teoricamente. Neste caso, ainda que um minimizador local não seja KKT, ele pode sempre ser aproximado por uma sequência de pontos que cumprem as condições KKT de maneira aproximada. Uma das formas de aproximar KKT é descrita pela chamada condição *Aproximadamente KKT* (AKKT) [4], que veremos neste trabalho. Existem outras formas de aproximar KKT além de AKKT, mas não serão tratadas com detalhes neste trabalho. Essas condições são conhecidas como *condições sequenciais de otimalidade*, pois, ao contrário de KKT, caracterizam minimizadores como pontos de acumulação de sequências.

Diante do exposto, no segundo capítulo, apresentamos o problema de programação não

linear e uma revisão de conceitos básicos sobre problemas de otimização não linear gerais. Apresentamos também exemplos e resultados importantes acerca desses problemas.

No terceiro capítulo apresentamos o método da penalidade externa, assim como as condições KKT. Ambos são importantes para a construção teórica da condição AKKT.

Condições sequências de otimalidade são tratadas no quarto capítulo, abordando essencialmente a condição AKKT e as justificativas teóricas que a conectam à KKT. Além disso, outras condições sequências de otimalidade na literatura são citadas.

Por fim, no quinto capítulo, citamos alguns métodos que geram naturalmente pontos AKKT. Mais especificamente, apresentaremos com detalhes o método do Lagrangiano Aumentado com salvaguardas definido em [5] (veja também [3]).

1.1 **Objetivo do trabalho**

Apresentar um estudo da condição sequencial de otimalidade AKKT e a sua aplicação a um método Lagrangiano Aumentado. Sobretudo, fazer uma releitura do texto “*Otimização prática usando o Lagrangiano Aumentado*” escrito por José Mario Martínez.

2 Revisão de conceitos

O problema de programação não linear (PNL) consiste em

$$\begin{aligned} & \text{minimizar } f(\mathbf{x}) \\ & \text{sujeito a } h_i(\mathbf{x}) = 0, i = 1, 2, \dots, m, \\ & \quad g_i(\mathbf{x}) \leq 0, i = 1, 2, \dots, p, \end{aligned} \tag{P}$$

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

Dizemos que um ponto $\hat{\mathbf{x}} \in \mathbb{R}^n$ é viável quando $\hat{\mathbf{x}} \in \Omega = \{\mathbf{x} \in \mathbb{R}^n | h_i(\mathbf{x}) = 0, i = 1, 2, \dots, m, g_i(\mathbf{x}) \leq 0, i = 1, 2, \dots, p\}$. O conjunto Ω é chamado de *conjunto viável*.

Definição 2.1. *Um ponto viável \mathbf{x}^* é minimizador local quando existe uma vizinhança V de \mathbf{x}^* tal que $f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in V \cap \Omega$. Quando $V = \mathbb{R}^n$ dizemos que \mathbf{x}^* é minimizador global.*

Deve-se ressaltar que não é prático avaliar numericamente f em toda a vizinhança de um ponto. Por isso, a definição acima não é adequada como critério de otimalidade algoritmos. Nesse sentido, o que se busca são candidatos à minimizador que cumpram propriedades numericamente verificáveis. Usaremos de propriedades analíticas do problema para caracterizar minimizadores locais de uma forma conveniente (veremos mais a frente as condições KKT). Quaisquer condições satisfeitas por minimizadores são chamadas de condições necessárias de otimalidade ou, apenas, *condições de otimalidade*. O ideal é utilizarmos condições que sejam fáceis de serem verificadas na prática e que descrevam bem candidatos à solução do problema.

No que segue, relembremos alguns conceitos básicos de análise no \mathbb{R}^n utilizados no trabalho. O leitor pode consultar ainda o capítulo 1 de [1]. Uma referência completa sobre o assunto é o livro de E. L. Lima [6].

Definição 2.2. *Dizemos que $\mathbf{x} \in \mathbb{R}^n$ é um ponto de acumulação do conjunto $X \subset \mathbb{R}^n$ quando toda vizinhança de \mathbf{x} contém algum ponto de X diferente de \mathbf{x} .*

Note que um ponto de acumulação de X pode pertencer a X ou não.

Segue da definição acima que um ponto de acumulação \mathbf{x} pode ser caracterizado como limite de uma sequência de pontos $\mathbf{x}^k \in X$, isto é, $\lim_{k \rightarrow \infty} \mathbf{x}^k = \mathbf{x}$. O *fecho* do conjunto $X \subset \mathbb{R}^n$ é o conjunto $\overline{X} = X \cup \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \text{ é ponto de acumulação de } X\}$. Dizemos que um conjunto $F \subset \mathbb{R}^n$ é *fechado* quando $F = \overline{F}$. Um conjunto $L \subset \mathbb{R}^n$ é limitado quando existe $M > 0$ tal que $\|x\| \leq M$ para todo $x \in L$.

Definição 2.3. Um conjunto $X \in \mathbb{R}^n$ é dito compacto quando é limitado e fechado.

2.1 O problema de programação não linear (PNL)

O problema de PNL é o problema central considerado no trabalho, que consiste em

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

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\mathbf{h} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ e $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^p$ são funções continuamente diferenciáveis. Podemos escrever $\mathbf{h} = (h_1, h_2, \dots, h_m)$, onde $h_i : \mathbb{R}^n \rightarrow \mathbb{R}$ para todo $i = 1, \dots, m$. O mesmo para \mathbf{g} .

A Figura 2.1 ilustra minimizadores de uma função f de uma variável. Repare que nem todo minimizador local é global. Podemos ter ainda pontos de sela, caso em que também a derivada de f é nula (ponto estacionário).

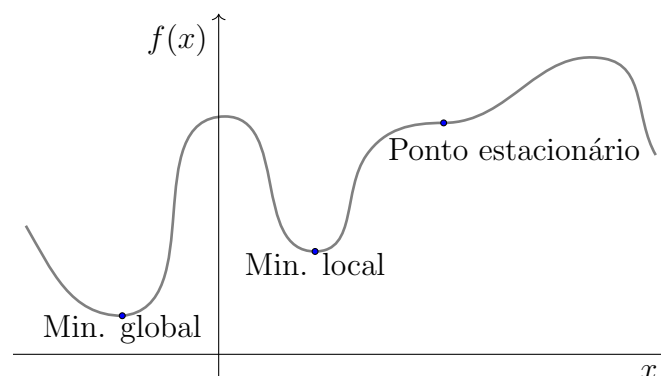


Figura 1: Minimizadores locais e globais.

O teorema abaixo garante a existência de um minimizador.

Teorema 2.1 (de Weierstrass). Se $\Omega \in \mathbb{R}^n$ é compacto e não vazio e $f : \Omega \rightarrow \mathbb{R}$ contínua, então f admite minimizador global sobre Ω .

Demonstração. Ver Teorema 1.3.1 de [7]. \square

Como as funções das restrições do PNL são contínuas, o conjunto viável é fechado. De fato, tomando uma sequência convergente de pontos viáveis $(\mathbf{x}^k) \rightarrow \mathbf{x}^*$, isto é, $\mathbf{h}(\mathbf{x}^k) = \mathbf{0}$ e $\mathbf{g}(\mathbf{x}^k) \leq \mathbf{0}$ para todo k , note que, pela continuidade de \mathbf{h} e \mathbf{g} em \mathbf{x}^* , temos $h_i(\mathbf{x}^*) = \lim h_i(\mathbf{x}^k) = 0$ para todo i e $g_j(\mathbf{x}^*) = \lim g_j(\mathbf{x}^k) \leq \lim 0 = 0$ para todo j .

Na prática, uma forma de garantir limitação do conjunto viável é intersectá-lo com uma caixa. Isto é, adicionar as restrições $\ell_i \leq x_i \leq u_i$ para todo $i = 1, \dots, n$. Note que, porém, tais restrições podem ser incorporadas nas restrições de desigualdade $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$, e logo não as escrevemos explicitamente na definição de Ω . Notamos que, na prática numérica, caixas são impostas mesmo porque a representação de números no computador tem limites inferiores e superiores.

Quando Ω não é compacto, pode não existir minimizador ainda que o conjunto viável seja limitado.

Exemplo 2.1. A função $f : \Omega \rightarrow \mathbb{R}$ definida por

$$f(x) = x^2$$

não admite minimizador caso tenhamos $\Omega = (0, 1]$. Veja que sempre existe $x \in \Omega$ tal que $f(x) < f(\bar{x})$ para um $\bar{x} \in \Omega$ arbitrário. \square

Exemplo 2.2. Agora, eis um exemplo onde $f : \Omega \rightarrow \mathbb{R}$ definida por

$$f(x) = \frac{1}{x}$$

não admite minimizador caso tenhamos $\Omega = (0, \infty)$ não limitado. Análogo ao exemplo anterior, existe sempre $x \in \Omega$ tal que $f(x) < f(\bar{x})$ para algum $\bar{x} \in \Omega$ arbitrário. \square

A caracterização de pontos estacionários de problemas com restrições recaem nas condições de Karush-Kuhn-Tucker (KKT), que apresentaremos à frente. Porém, veremos que existem minimizadores que não são KKT. Felizmente, é possível sempre caracterizar qualquer minimizador por sequências de pontos aproximadamente KKT (AKKT). Uma importante ferramenta que garante isto é o método da penalização externa. Tal método não é considerado bom para ser usado na prática numérica mas possui propriedades analíticas que iluminam provas de resultados importantes em otimização e será introduzido apenas para a construção da demonstração do teorema de AKKT. É o que veremos a seguir.

3 Condição de Karush-Kuhn-Tucker

Neste capítulo apresentaremos a principal ferramenta para caracterizar minimizadores de (P), as chamadas condições de Karush-Kuhn-Tucker (KKT). Antes, apresentamos o método de penalização externa que será usado na prova das condições AKKT.

3.1 Penalização externa

A estratégia da penalização externa consiste em eliminar as restrições “difíceis” $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ e $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ de (P) e incluí-las na função objetivo, reduzindo o conjunto de restrições à $\mathbf{x} \in \Omega$, onde Ω é um conjunto compacto (na prática $x_i \in [\ell_i, u_i]$ para todo $i = 1, \dots, n$), o qual, configuram as restrições “fáceis”.

O método da penalidade externa consiste em computar um minimizador global \mathbf{x}^k de uma sequência de subproblemas da forma

$$\text{minimizar} \quad f(\mathbf{x}) + \frac{\rho_k}{2} [\|\mathbf{h}(\mathbf{x})\|^2 + \|\max\{\mathbf{g}(\mathbf{x}), \mathbf{0}\}\|^2] \quad (3.1)$$

$$\text{sujeito a} \quad \mathbf{x} \in \Omega \quad (3.2)$$

onde $0 < \rho_k \rightarrow \infty$. O termo ρ é chamado de parâmetro de penalização e pode ser aumentado a cada iteração multiplicando-o por um escalar $\gamma > 1$. A teoria permite inicializar ρ_0 como qualquer escalar positivo. Na prática, tomamos ρ_0 considerando os dados do problema [3].

Chamaremos de medida de inviabilidade a função

$$\phi(\mathbf{x}) = \|\mathbf{h}(\mathbf{x})\|^2 + \|\max\{\mathbf{g}(\mathbf{x}), \mathbf{0}\}\|^2.$$

Observe que $\phi(\mathbf{x}) = 0$ se e somente se $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ e $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$, ou seja, em (3.1) buscamos minimizar a não viabilidade de \mathbf{x} .

Eventualmente, o parâmetro ρ_k precisa ser muito grande para obtermos viabilidade suficiente do iterando \mathbf{x}^k . De qualquer modo, o método de penalização externa (3.1)–(3.2) fornece o resultado teórico a seguir, que será útil para nosso propósito.

Teorema 3.1 (Teorema 2.4 de [3]). *Suponhamos que o conjunto viável do problema (P) é não vazio e a sequência de pontos $\{\mathbf{x}^k\}$ fornecida pelo método de penalidade externa (3.1)–(3.2) admite um ponto de acumulação \mathbf{x}^* . Então, \mathbf{x}^* é um minimizador global de (P).*

Demonstração. Tome $K \subset \mathbb{N}$ o conjunto de infinitos índices tal que

$$\lim_{k \in K} \mathbf{x}^k = \mathbf{x}^*.$$

Suponha que \mathbf{x}^* não é viável. Assim existe $\mathbf{z} \in \Omega$ tal que $\phi(\mathbf{x}^*) > \phi(\mathbf{z}) = 0$, ou melhor, $\|\mathbf{h}(\mathbf{x}^*)\|^2 + \|\max\{\mathbf{g}(\mathbf{x}^*), \mathbf{0}\}\|^2 > \|\mathbf{h}(\mathbf{z})\|^2 + \|\max\{\mathbf{g}(\mathbf{z}), \mathbf{0}\}\|^2$. Vamos denotar $\mathbf{g}(\mathbf{x})_+ = \max\{\mathbf{g}(\mathbf{x}), \mathbf{0}\}$. Como \mathbf{h} e \mathbf{g} são contínuas, existe $c > 0$ tal que

$$\|\mathbf{h}(\mathbf{x}^k)\|^2 + \|\mathbf{g}(\mathbf{x}^k)_+\|^2 > \|\mathbf{h}(\mathbf{z})\|^2 + \|\mathbf{g}(\mathbf{z})_+\|^2 + c$$

para todo $k \in K$ suficientemente grande. Logo, para esses índices k ,

$$f(\mathbf{x}^k) + \frac{\rho_k}{2} [\|\mathbf{h}(\mathbf{x}^k)\|^2 + \|\mathbf{g}(\mathbf{x}^k)_+\|^2] > f(\mathbf{x}^k) + \frac{\rho_k}{2} [\|\mathbf{h}(\mathbf{z})\|^2 + \|\mathbf{g}(\mathbf{z})_+\|^2] + \frac{\rho_k c}{2}.$$

Somando $f(\mathbf{z}) - f(\mathbf{z}) = 0$ no lado direito da expressão acima obtemos

$$f(\mathbf{x}^k) + \frac{\rho_k}{2} [\|\mathbf{h}(\mathbf{x}^k)\|^2 + \|\mathbf{g}(\mathbf{x}^k)_+\|^2] > f(\mathbf{z}) + \frac{\rho_k}{2} [\|\mathbf{h}(\mathbf{z})\|^2 + \|\mathbf{g}(\mathbf{z})_+\|^2] + \frac{\rho_k c}{2} + f(\mathbf{x}^k) - f(\mathbf{z})$$

e como $\rho_k \rightarrow \infty$, segue que

$$\lim_{k \in K} \left(f(\mathbf{x}^k) - f(\mathbf{z}) + \frac{\rho_k c}{2} \right) = \infty.$$

Logo,

$$f(\mathbf{x}^k) + \frac{\rho_k}{2} [\|\mathbf{h}(\mathbf{x}^k)\|^2 + \|\mathbf{g}(\mathbf{x}^k)_+\|^2] > f(\mathbf{z}) + \frac{\rho_k}{2} [\|\mathbf{h}(\mathbf{z})\|^2 + \|\mathbf{g}(\mathbf{z})_+\|^2],$$

contrariando o fato de que \mathbf{x}^k é um minimizador global do subproblema definido por ρ_k . O motivo desta contradição foi supor que \mathbf{x}^* não era viável. Assim, \mathbf{x}^* é viável e devemos ter, pela construção do método, \mathbf{x}^k minimizador global de cada subproblema, isto é,

$$f(\mathbf{x}^k) + \frac{\rho_k}{2} [\|\mathbf{h}(\mathbf{x}^k)\|^2 + \|\mathbf{g}(\mathbf{x}^k)_+\|^2] \leq f(\mathbf{z}) + \frac{\rho_k}{2} [\|\mathbf{h}(\mathbf{z})\|^2 + \|\mathbf{g}(\mathbf{z})_+\|^2].$$

Como $\mathbf{z} \in \Omega$, $\mathbf{h}(\mathbf{z}) = \mathbf{0}$ e $\mathbf{g}(\mathbf{z}) \leq \mathbf{0}$, assim

$$f(\mathbf{x}^k) + \frac{\rho_k}{2} [\|\mathbf{h}(\mathbf{x}^k)\|^2 + \|\mathbf{g}(\mathbf{x}^k)_+\|^2] \leq f(\mathbf{z}).$$

Sabendo que $\mathbf{x}^* \in \Omega$ e $(\mathbf{x}^k) \rightarrow \mathbf{x}^*$ sobre K , podemos concluir que

$$f(\mathbf{x}^*) \leq f(\mathbf{z}).$$

Ou seja, \mathbf{x}^* é minimizador global. □

Exemplo 3.1. Considere o problema

$$\text{minimizar } f(x) = x \quad \text{sujeito a } g(x) = -x \leq 0.$$

O subproblema consistirá em resolver

$$\text{minimizar } x + \frac{\rho_k}{2} |\max\{-x, 0\}|^2 \quad \text{sujeito a } x \in \mathbb{R}.$$

Observe que se $x \geq 0$ então $\max\{-x, 0\} = 0$ e se $x < 0$ então $\max\{-x, 0\} = -x$. Geometricamente, no subproblema onde $x \geq 0$ teremos uma função afim e onde $x < 0$ teremos uma parábola. Vamos resolver o subproblema onde $x < 0$ que consiste em minimizar $x + \frac{\rho_k}{2} x^2$. Derivando a função objetivo e igualando a zero, teremos

$$1 + \rho_k x^k = 0 \Rightarrow x^k = \frac{-1}{\rho_k},$$

ou seja, realmente $x^k < 0$ e quando $\rho_k \rightarrow \infty$ a solução se aproxima de zero, a qual, é a solução do problema original.

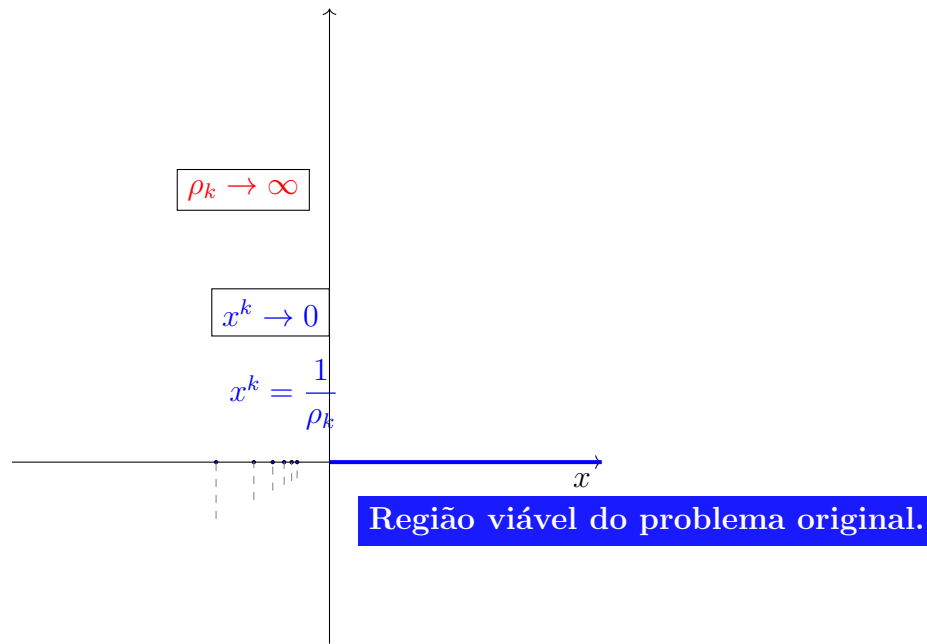


Figura 2: Convergência do método da penalidade externa.

□

Grosso modo, determinar que um ponto viável \mathbf{x}^* é minimizador local ou global de f requisitaria avaliar as funções objetivo e das restrições em um conjunto denso, o que é impraticável. Além disso, nas implementações numéricas é difícil garantir minimizadores, mesmo locais, no PNL geral. Neste caso, o que tentamos fazer é buscar pontos limites de algoritmos que satisfaçam propriedades de um minimizador (condições necessárias de otimalidade). A ferramenta que guia essa busca são as condições KKT.

3.2 Condições de Karush-Kuhn-Tucker (KKT)

Tido como uma das principais ferramentas na programação não linear, as condições de Karush-Kuhn-Tucker nos garantem a descoberta de pontos estacionários e, no caso de problemas convexos, minimizadores globais.

Diremos que g_i é uma restrição ativa em \mathbf{x} quando $g_i(\mathbf{x}) = 0$ e que o conjunto $I = \{i \in \{1, \dots, p\} | g_i(\mathbf{x}^*) = 0\}$ contém todos os índices onde g_i é restrição ativa em \mathbf{x}^* .

Assim, as condições KKT estabelecem que se \mathbf{x}^* é minimizador local e $\{\nabla h_i(\mathbf{x}^*) | i = 1, \dots, m\} \cup \{\nabla g_i(\mathbf{x}^*) | i \in I\}$ é um conjunto linearmente independente, então existem

$\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\mu} \in \mathbb{R}^p$, onde $\boldsymbol{\mu} \geq 0$ tais que

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \nabla h_i(\mathbf{x}^*) + \sum_{i \in I} \mu_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}.$$

Outra forma de dizer que o ponto viável \mathbf{x}^* é KKT é que devemos ter $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\boldsymbol{\mu} \in \mathbb{R}^p$ tais que

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \nabla h_i(\mathbf{x}^*) + \sum_{i=1}^p \mu_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}, \quad (3.3)$$

$$\text{Para todo } i = 1, \dots, p, \quad \mu_i g_i(\mathbf{x}^*) = 0, \quad (3.4)$$

$$\text{Para todo } i = 1, \dots, p, \quad \mu_i \geq 0, \quad (3.5)$$

$$\text{Para todo } i = 1, \dots, m, \quad \lambda_i \quad \text{livre.} \quad (3.6)$$

As condições KKT generalizam as conhecidas condições de Lagrange para restrições de igualdade vistas em cursos de cálculo. Por isso, os escalares λ_i e μ_j são chamados *multiplicadores de Lagrange*, ou simplesmente *multiplicadores*.

A condição (3.3) diz que $-\nabla f(\mathbf{x}^*)$ é combinação linear dos gradientes das restrições (veja o Exemplo 3.2 para uma visão geométrica). A condição (3.4) é chamada de *complementaridade*. Nela, exige-se $\mu = 0$ caso tenhamos $g(\mathbf{x}^*) \neq 0$. As inequações em (3.5) dizem que μ_i são não negativos.

Exemplo 3.2. [1] Vamos verificar que a solução $\mathbf{x}^* = (1, 1)$ do seguinte problema é um ponto KKT:

$$\begin{aligned} \text{minimizar} \quad & f(\mathbf{x}) = (x_1 - 2)^2 + (x_2 - 1)^2 \\ \text{sujeito a} \quad & g_1(\mathbf{x}) = x_1 + x_2 - 2 \leq 0 \\ & g_2(\mathbf{x}) = x_1^2 - x_2 \leq 0. \end{aligned}$$

Pelas condições KKT, para este problema devemos ter

$$\begin{aligned} \nabla f(\mathbf{x}^*) + \mu_1 \nabla g_1(\mathbf{x}^*) + \mu_2 \nabla g_2(\mathbf{x}^*) &= \mathbf{0}, \\ \mu_1 g_1(\mathbf{x}^*) &= 0, \quad \mu_2 g_2(\mathbf{x}^*) = 0, \\ \mu_1 &\geq 0, \quad \mu_2 \geq 0. \end{aligned}$$

Isto é,

$$\begin{aligned} \begin{bmatrix} 2(x_1 - 2) \\ 2(x_2 - 1) \end{bmatrix} + \mu_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \mu_2 \begin{bmatrix} 2x_2 \\ -1 \end{bmatrix} &= \mathbf{0}, \\ \mu_1(x_1 + x_2 - 2) &= 0, \quad \mu_2(x_1^2 - x_2) = 0, \\ \mu_1 &\geq 0, \quad \mu_2 \geq 0. \end{aligned}$$

Note que o ponto $\mathbf{x}^* = (1, 1)$ é viável (nas restrições obtemos $g_1(\mathbf{x}^*) = 0$ e $g_2(\mathbf{x}^*) = 0$) e também cumpre as condições KKT com os multiplicadores $\mu_1 = \mu_2 = 2/3 > 0$.

$$\begin{aligned} \begin{bmatrix} 2(1 - 2) \\ 2(1 - 1) \end{bmatrix} + \mu_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \mu_2 \begin{bmatrix} 2 \\ -1 \end{bmatrix} &= \mathbf{0} \\ \mu_1(1 + 1 - 2) = 0, \quad \mu_2(1^2 - 1) = 0 &\Rightarrow \quad \mu_1 = \mu_2 = \frac{2}{3}. \\ \mu_1 &\geq 0, \quad \mu_2 \geq 0 \end{aligned}$$

Na Figura 3, observe a direção do vetor negativo do gradiente de f . Isso quer dizer que os valores das curvas de nível da função objetivo (circunferências cinzas) diminuem à medida que aproxima de seu centro. Assim, convém escolher como solução (ponto \mathbf{x}^*) um ponto mais a direita dentro da região viável (região azul). Ou seja, a curva de nível de menor valor que cruza a região viável descreve a solução do PNL.

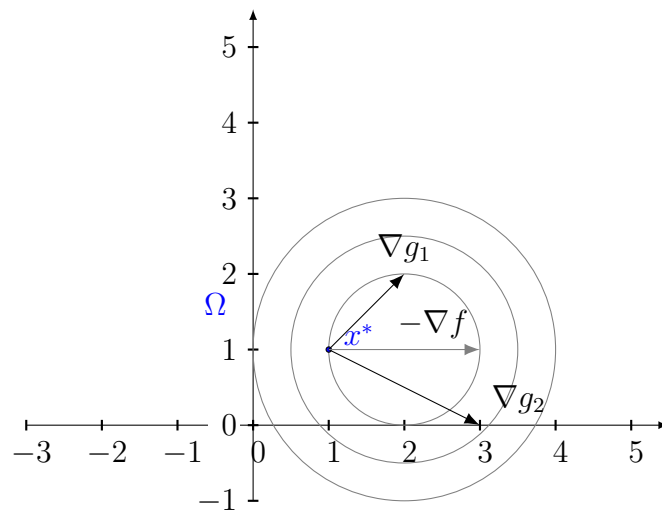


Figura 3: Figura para o Exemplo 3.2.

□

Exemplo 3.3. [1] Considere o seguinte PNL:

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

Neste exemplo, $\mathbf{x}^* = (0, 0)$ é o minimizador do problema, mas vamos mostrar que \mathbf{x}^* não satisfaz as condições KKT. Observe que

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \nabla g_1(\mathbf{x}) = \begin{bmatrix} -3x_1^2 \\ 1 \end{bmatrix} \text{ e } \nabla g_2(\mathbf{x}) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

Assim

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

Deste modo, as condições KKT não valem pois

$$\begin{aligned} \nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \nabla h_i(\mathbf{x}^*) + \sum_{i \in I} \mu_i \nabla g_i(\mathbf{x}^*) &= \\ \nabla f(\mathbf{x}^*) + \mu_1 \nabla g_1(\mathbf{x}^*) + \mu_2 \nabla g_2(\mathbf{x}^*) &= \\ (1, 0) + \mu_1(0, 1) + \mu_2(0, -1) &= (1, \mu_1 - \mu_2) \neq \mathbf{0}. \end{aligned}$$

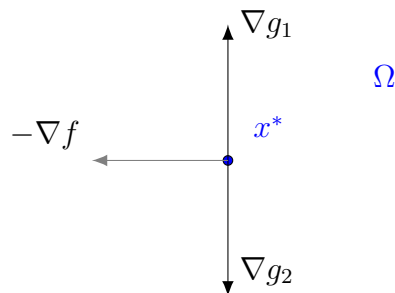


Figura 4: Figura para o Exemplo 3.3

□

Como vimos, as condições KKT relacionam a função objetivo e as restrições do

problema de forma consideravelmente simples e, além de serem fáceis de verificar na prática, selecionam bem os candidatos a minimizador do problema. No entanto, como vimos no Exemplo 3.3, KKT não é uma condição de otimalidade. Mesmo assim, podemos relaciona-lo junto ao que chamamos de *condições de qualificação*.

Uma condição de qualificação (CQ) é qualquer hipótese que garante que qualquer minimizador local, de qualquer problema, seja um ponto KKT. Logo, se o minimizador não é KKT, então ele não satisfaz alguma CQ. Portanto, podemos dizer que uma condição de otimalidade é a solução satisfazer KKT ou não satisfazer alguma CQ. Isto é, uma condição de otimalidade, associada a KKT, consiste em um ponto satisfazer

KKT ou não-CQ.

A principal CQ utilizada neste trabalho é a independência linear dos gradientes das restrições ativas (LICQ, do inglês *Linear Independence Constraint Qualification*), que será definida a seguir.

Definição 3.1. (LICQ) Dizemos que um ponto $\mathbf{x} \in \Omega$ satisfaz LICQ se $\{\nabla h_i(\mathbf{x}); i = 1, \dots, m\} \cup \{\nabla g_i(\mathbf{x}); i \in I\}$ é um conjunto linearmente independente.

Na Figura 4 é possível observar que os gradientes das restrições ativas são linearmente dependentes e, portanto, não satisfazem LICQ.

Apesar do minimizador no Exemplo 3.3 não satisfazer exatamente as condições KKT, elas podem sempre ser aproximadas em um certo sentido. Veja que, se no exemplo anterior consideramos a sequência $(\mathbf{x}^k) = (1/\sqrt[3]{k}, 1/k) > 0$ e as sequências de multiplicadores definidas por $\mu_1^k = 1/3x_1^2$ e $\mu_2^k = 1/3x_1^2$, vamos obter

$$g_1(\mathbf{x}^k) = -\left(1/\sqrt[3]{k}\right)^3 + 1/k = 0 \quad \text{e} \quad g_2(\mathbf{x}^k) = -1/k < 0$$

verificando que (\mathbf{x}^k) é uma sequência de pontos viáveis. Além disso,

$$\min\{\mu_1^k, -g_1(\mathbf{x}^k)\} \leq \varepsilon_k \quad \text{e} \quad \min\{\mu_2^k, -g_2(\mathbf{x}^k)\} \leq \varepsilon_k$$

onde $\varepsilon_k > 0$, comprovando que (\mathbf{x}^k) cumpre a complementaridade das condições aproximadamente KKT (mais detalhada no próximo capítulo). Por fim,

$$\begin{aligned} \nabla f(\mathbf{x}^k) + \mu_1 \nabla g_1(\mathbf{x}^k) + \mu_2 \nabla g_2(\mathbf{x}^k) &= \\ (1, 0) + 1/3(x_1^k)^2(-3(x_1^k)^2, 1) + 1/3(x_1^k)^2(0, -1) &= \\ (1 - 1, 1/3(x_1^k)^2 - 1/3(x_1^k)^2) &= (0, 0). \end{aligned}$$

Ou seja, existe uma sequência $(\mathbf{x}^k) \rightarrow \mathbf{0}$ que “aproxima” KKT. Assim, o limite $\mathbf{x}^* = \mathbf{0}$ da sequência (\mathbf{x}^k) , será chamado de ponto aproximadamente KKT (AKKT). Essa condição é o que chamamos de uma condição sequencial de otimalidade, que veremos com detalhes no próximo capítulo.

4 Condições sequenciais de otimalidade

Na prática, são utilizados métodos iterativos para resolver problemas de otimização onde, a cada iteração, deve-se averiguar se o iterando é solução (ponto KKT) do problema. Em geral, é difícil, ou mesmo impossível, atestar otimalidade de maneira exata. O que se faz é interromper a execução do algoritmo quando alguma condição de otimalidade é aproximadamente satisfeita. As condições sequenciais de otimalidade nos fornecem as justificativas teóricas que formalizam esta prática.

4.1 A condição Aproximadamente KKT (AKKT)

No fim do capítulo anterior vimos que, apesar da solução da PNL não satisfazer KKT, seu minimizador pode ser aproximado por uma sequência. Os elementos dessa sequência cumpre aproximadamente as condições KKT, as quais serão apresentadas a seguir.

Dizemos que \mathbf{x}^* é ponto AKKT se existem sequências $(\mathbf{x}^k) \rightarrow \mathbf{x}^*$, $(\lambda^k) \subset \mathbb{R}^m$, $(\mu^k) \subset \mathbb{R}^p$ e $\varepsilon_k \rightarrow 0$ tais que, para todo k ,

$$\|\nabla f(\mathbf{x}^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(\mathbf{x}^k) + \sum_{i=1}^p \mu_i^k \nabla g_i(\mathbf{x}^k)\| \leq \varepsilon_k \quad (4.1)$$

$$\text{Para todo } i = 1, \dots, p, \quad |\min\{\mu_i^k, -g_i(\mathbf{x}^k)\}| \leq \varepsilon_k \quad (4.2)$$

$$\text{Para todo } i = 1, \dots, p, \quad \mu_i^k \geq 0. \quad (4.3)$$

Em qualquer vizinhança de um minimizador local podem ser encontrados pontos que satisfazem as inequações acima, para qualquer ε_k positivo, como demonstraremos no Corolário 4.1 à frente.

A expressão (4.2) é uma versão aproximada da complementaridade (3.4) das condições KKT. De fato, dado que $\mu_i \geq 0$, $\min\{\mu_i, -g_i(\mathbf{x})\} = 0$ é equivalente à $(\mu_i = 0 \text{ e } g_i(\mathbf{x}) \leq 0)$ ou $g_i(\mathbf{x}) = 0$. Há outras formas de representá-la, por exemplo, exigindo $|g_i(\mathbf{x}^k)\mu_i^k| \leq$

ε_k para todo i . Exigir que o produto vá para zero é mais do que (4.2). De fato, o produto está presente em uma condição sequencial mais forte que AKKT, conhecida como AKKT complementar (CAKKT) [8]. No entanto, o mínimo presente em (4.2) é mais numericamente estável que o produto: quando $g_i(\mathbf{x}^k) \rightarrow 0$, não importa quão rápido μ_i^k vai para infinito, o mínimo sempre tenderá a zero; enquanto que para o produto não temos essa garantia (por exemplo, se $-g_i(\mathbf{x}^k) = 1/k$ e $\mu_i^k = k^2$).

O seguinte teorema nos mostra que AKKT é uma propriedade dos minimizadores que independe de hipóteses adicionais, como necessita KKT.

Teorema 4.1 (Teorema 4.1 de [3]). *Seja \mathbf{x}^* minimizador local do problema (P). Então existem seqüências $(\mathbf{x}^k) \subset \mathbb{R}^n$, $(\lambda^k) \subset \mathbb{R}^m$, $(\mu^k) \subset \mathbb{R}_+^p$, tais que*

1. $\lim_{k \rightarrow \infty} \mathbf{x}^k = \mathbf{x}^*$.
2. $\lim_{k \rightarrow \infty} \nabla f(\mathbf{x}^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(\mathbf{x}^k) + \sum_{i=1}^p \mu_i^k \nabla g_i(\mathbf{x}^k) = \mathbf{0}$.
3. $\lim_{k \rightarrow \infty} \|\mathbf{h}(\mathbf{x}^k)\| = \lim_{k \rightarrow \infty} \|\mathbf{g}(\mathbf{x}^k)_+\| = 0$.
4. Para todo $i = 1, \dots, p$ e para todo $k \in \mathbb{N}$, $g_i(\mathbf{x}^k) < 0 \Rightarrow \mu_i^k = 0$. Em particular, $\lim_{k \rightarrow \infty} |\min\{\mu_i^k, -g_i(\mathbf{x}^k)\}| = 0$ para todo i .

Demonstração. Se \mathbf{x}^* é minimizador local de (P), então existe uma vizinhança $\|\mathbf{x} - \mathbf{x}^*\| \leq \delta$ tal que \mathbf{x}^* é minimizador global para algum $\delta > 0$. Logo, \mathbf{x}^* é a única solução global do seguinte problema:

$$\text{minimizar } f(\mathbf{x}) + \frac{1}{2} \|\mathbf{x} - \mathbf{x}^*\| \quad \text{sujeito a } \mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \|\mathbf{x} - \mathbf{x}^*\| \leq \delta. \quad (4.4)$$

Aplicando o Método de Penalidade Externa em (4.4), em cada subproblema devemos minimizar uma função que é contínua num conjunto compacto:

$$\begin{aligned} &\text{minimizar } f(\mathbf{x}) + \frac{1}{2} \|\mathbf{x} - \mathbf{x}^*\|^2 + \frac{\rho_k}{2} (\|\mathbf{h}(\mathbf{x})\|^2 + \|\mathbf{g}(\mathbf{x})_+\|^2) \\ &\text{sujeito a } \|\mathbf{x} - \mathbf{x}^*\| \leq \delta. \end{aligned}$$

Portanto, existe solução global \mathbf{x}^k para todo $k \in \mathbb{N}$.

O Teorema (3.1), relativo à convergência do método de penalidade externa, indica que

$$\lim_{k \rightarrow \infty} \mathbf{x}^k = \mathbf{x}^*.$$

Assim, tendo $\|\mathbf{x}^k - \mathbf{x}^*\| \leq \delta$ para k suficientemente grande, podemos dizer que \mathbf{x}^k é um ponto interior do subproblema e o gradiente da função objetivo deste subproblema deve ser zero. Ou seja, para todo $k \in \mathbb{N}$ suficientemente grande,

$$\nabla f(\mathbf{x}^k) + \nabla \mathbf{h}(\mathbf{x}^k)[\rho_k \mathbf{h}(\mathbf{x}^k)] + \nabla \mathbf{g}(\mathbf{x}^k)[\rho_k \mathbf{g}(\mathbf{x}^k)_+] + (\mathbf{x}^k - \mathbf{x}^*) = \mathbf{0}.$$

Definindo $\boldsymbol{\lambda}^k = \rho_k \mathbf{h}(\mathbf{x}^k)$ e $\boldsymbol{\mu}^k = \rho_k \mathbf{g}(\mathbf{x}^k)_+$ e sabendo que $\lim_{k \rightarrow \infty} \|\mathbf{x}^k - \mathbf{x}^*\| = 0$, vamos obter

$$\lim_{k \rightarrow \infty} \nabla f(\mathbf{x}^k) + \nabla \mathbf{h}(\mathbf{x}^k) \boldsymbol{\lambda}^k + \nabla \mathbf{g}(\mathbf{x}^k) \boldsymbol{\mu}^k = \mathbf{0}.$$

Com relação à complementariedade, basta observar que $\boldsymbol{\mu}^k = \rho_k \mathbf{g}(\mathbf{x}^k)_+$ fornece $\mu_i^k = 0$ se $g_i(\mathbf{x}^k) < 0$. \square

Corolário 4.1. *Suponhamos as hipóteses do Teorema 4.1. Seja $\varepsilon > 0$. Então, existem $\mathbf{x} \in \mathbb{R}^m$, $\lambda_i \in \mathbb{R}$ para $i = 1, 2, \dots, m$ e $\mu_i \in \mathbb{R}_+$ para $i = 1, 2, \dots, p$, tais que*

1. $\|\mathbf{x} - \mathbf{x}^*\| \leq \varepsilon$,
2. $\|\nabla f(\mathbf{x}) + \sum_{i=1}^m \lambda_i \nabla h_i(\mathbf{x}) + \sum_{i=1}^p \mu_i \nabla g_i(\mathbf{x})\| \leq \varepsilon$,
3. $\|\mathbf{h}(\mathbf{x})\| \leq \varepsilon$, $\|\mathbf{g}(\mathbf{x})_+\| \leq \varepsilon$,
4. Para todo $i = 1, \dots, p$, $|\min\{\mu_i, -g_i(\mathbf{x})\}| \leq \varepsilon$.

Demonstração. Segue imediatamente do Teorema 4.1. \square

Assim, podemos dizer que \mathbf{x}^* viável é um ponto AKKT se, para todo $\varepsilon > 0$, existem $\bar{\mathbf{x}}$, $\boldsymbol{\lambda}$ e $\boldsymbol{\mu} \geq \mathbf{0}$ satisfazendo os itens do Corolário 4.1. Ademais, o Corolário 4.1 nos diz que todo minimizador local é AKKT e, portanto, aproximável por pontos $\bar{\mathbf{x}}$.

Exemplo 4.1. Neste exemplo, apresentamos um problema simples onde AKKT é bom, isto é, consegue descrever corretamente minimizadores em que KKT falha:

$$\text{minimizar } x \quad \text{sujeito a } x^2 \leq 0.$$

Note que a única solução $x^* = 0$ é o único ponto viável.

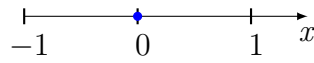


Figura 5: Figura para o exemplo 4.1

Porém, não existe μ que permita x^* satisfazer as condições KKT pois

$$x' + \mu(x^2)' = 1 + \mu(2x)$$

e $x^* = 0$ fornece

$$1 + \mu(2 \cdot 0) \neq 0.$$

O que pode ser feito é se aproximar tais condições. Veja que se $x < 0$ então $\mu = -1/2x > 0$ resolve

$$x' + \mu(x^2)' = 1 + \mu(2x) = 1 + \left(-\frac{1}{2x}\right)(2x) = 0.$$

Assim, x^* é AKKT com as sequências definidas por $x_k = -1/k$ e $\mu_k = k/2$, $k \geq 2$, pois neste caso $|\min\{\mu_k, -g(x_k)\}| = 1/k^2 \rightarrow 0$. \square

Exemplo 4.2. [9] O exemplo a seguir ilustra um caso onde AKKT pode ser ruim, isto é, não caracterizar minimizadores bem. No seguinte exemplo, há muitos pontos AKKT que não são minimizadores:

$$\begin{aligned} \text{minimizar} \quad & f(\mathbf{x}) = (x_1 - 1)^2 + (x_2 - 1)^2 \\ \text{sujeito a} \quad & g_1(\mathbf{x}) = x_1 x_2 \leq 0 \\ & g_2(\mathbf{x}) = -x_1 \leq 0 \\ & g_3(\mathbf{x}) = -x_2 \leq 0. \end{aligned}$$

Pela interpretação geométrica do problema, na figura abaixo, podemos ver que os pontos (1,0) e (0,1) são os únicos minimizadores. Apesar disso, vamos mostrar que para esse problema, todos os pontos viáveis são AKKT!

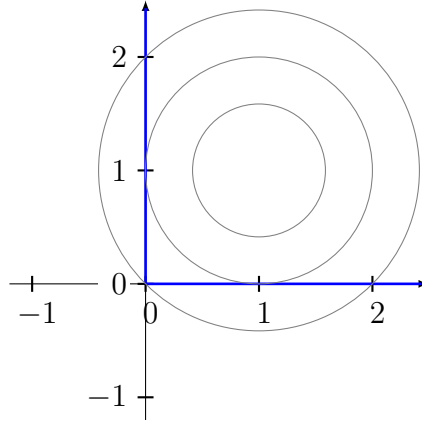


Figura 6: Figura para o Exemplo 4.2

Seja $\bar{\mathbf{x}}$ qualquer ponto viável. Como há simetria no problema, vamos admitir, sem perda de generalidade, que $\bar{x}_2 = 0$. Assim, se $\bar{x}_1 > 0$ então as seqüências definidas por $\mathbf{x}^k = (\bar{x}_1, 2(1 - \bar{x}_1)/k)$, $\mu_1^k = k$, $\mu_2^k = 0$ e $\mu_3^k = k\bar{x}_1 - 2$ para todo $k \geq 2/\bar{x}_1$ fornecem $\bar{\mathbf{x}}$ como um ponto AKKT. De fato, é fácil ver que $\mathbf{x}^k \rightarrow \bar{\mathbf{x}}$ e $|\min\{\mu_i^k, -g_i(\mathbf{x}^k)\}| \rightarrow 0$ para $i = 1, 2, 3$, e ainda

$$\begin{aligned} \nabla f(\mathbf{x}^k) + \mu_1^k \nabla g_1(\mathbf{x}^k) + \mu_2^k \nabla g_2(\mathbf{x}^k) + \mu_3^k \nabla g_3(\mathbf{x}^k) &= \\ \begin{bmatrix} 2(x_1 - 1) \\ 2(x_2 - 1) \end{bmatrix} + k \begin{bmatrix} x_2 \\ x_1 \end{bmatrix} + 0 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + (k\bar{x}_1 - 2) \begin{bmatrix} 0 \\ -1 \end{bmatrix} &= \\ \begin{bmatrix} 2(x_1 - 1) + kx_2 \\ 2(x_2 - 1) + kx_1 - (k\bar{x}_1 - 2) \end{bmatrix} = \begin{bmatrix} 2(\bar{x}_1 - 1) + k2\left(\frac{1 - \bar{x}_1}{k}\right) \\ 2\left(2\left(\frac{1 - \bar{x}_1}{k}\right) - 1\right) + k\bar{x}_1 - k\bar{x}_1 + 2 \end{bmatrix} &= \\ \begin{bmatrix} 2(\bar{x}_1 - 1) - 2(\bar{x}_1 - 1) \\ 4\left(\frac{1 - \bar{x}_1}{k}\right) - 2 + 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 4\left(\frac{1 - \bar{x}_1}{k}\right) \end{bmatrix} = \begin{bmatrix} 0 \\ 4\bar{x}_2 \end{bmatrix} &= \mathbf{0}. \end{aligned}$$

Agora, se $\bar{x}_1 = 0$ então as seqüências definidas por $\mathbf{x}^k = (2/k, 2/k)$, $\mu_1^k = k$, $\mu_2^k = 0$ e $\mu_3^k = 0$ para todo $k \geq 1$ também nos retornam $\bar{\mathbf{x}}$ um ponto AKKT. Novamente, $\mathbf{x}^k \rightarrow \bar{\mathbf{x}}$, $|\min\{\mu_i^k, -g_i(\mathbf{x}^k)\}| \rightarrow 0$ para $i = 1, 2, 3$, e

$$\begin{aligned}
& \nabla f(\mathbf{x}^k) + \mu_1^k \nabla g_1(\mathbf{x}^k) + \mu_2^k \nabla g_2(\mathbf{x}^k) + \mu_3^k \nabla g_3(\mathbf{x}^k) = \\
& \begin{bmatrix} 2(x_1 - 1) \\ 2(x_2 - 1) \end{bmatrix} + k \begin{bmatrix} x_2 \\ x_1 \end{bmatrix} + 0 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \\
& \begin{bmatrix} 2(x_1 - 1) + kx_2 \\ 2(x_2 - 1) + kx_1 \end{bmatrix} = \begin{bmatrix} 2x_1 - 2 + k2/k \\ 2x_2 - 2 + k2/k \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} = \begin{bmatrix} 2\bar{x}_1 \\ 2\bar{x}_2 \end{bmatrix} = \mathbf{0}.
\end{aligned}$$

□

O problema do exemplo anterior contém restrições do tipo $\mathbf{a}(\mathbf{x}) \geq \mathbf{0}$, $\mathbf{b}(\mathbf{x}) \geq \mathbf{0}$, $\mathbf{a}(\mathbf{x})^T \mathbf{b}(\mathbf{x}) \leq 0$, que são conhecidas como *restrições de complementaridade*. Este tipo de restrições têm alto grau de degeneração, e portanto é de se esperar que AKKT não seja adequado. De fato, problemas com esse tipo de restrições exigem uma teoria adequada própria, assim como condições sequenciais específicas [10].

Exemplo 4.3. (Observação 3.1 de [11]) Considere o seguinte problema convexo:

$$\begin{aligned}
& \text{minimizar} && f(\mathbf{x}) = x_1 \\
& \text{sujeito a} && h(\mathbf{x}) = x_2 = 0 \\
& && g(\mathbf{x}) = \left[\sqrt{(x_1^2 - 1)^2 + 4x_2^2} + x_1^2 - 1 \right]^2 \leq 0.
\end{aligned}$$

Sabemos que em problemas convexos KKT é suficiente até mesmo para otimalidade global [1]. Apesar disso, vamos mostrar que para este exemplo AKKT não é.

Repare que conjunto viável é $[-1, 1] \times \{0\}$ e o ponto $(-1, 0)$ é o único minimizador. De fato, pela restrição de igualdade devemos ter $x_2 = 0$. Substituindo na restrição de desigualdade, obtemos $\sqrt{(x_1^2 - 1)^2 + 4x_2^2} + x_1^2 - 1 = 0$, que implica $x_1 \in [-1, 1]$. Afirmamos que a sequência $\mathbf{x}^k = (-1/k, 1/k)$ com sequências de multiplicadores adequadas satisfazem as condições AKKT. Evidentemente, $\{\mathbf{x}^k\}$ não converge para o minimizador $(-1, 0)$.

Os gradientes das funções neste problema são dados por

$$\begin{aligned}
\nabla f(\mathbf{x}) &= \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \nabla h(\mathbf{x}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{ e} \\
\nabla g(\mathbf{x}) &= \begin{bmatrix} t(\mathbf{x}) \left(\frac{2x_1(x_1^2 - 1)}{\sqrt{(x_1^2 - 1)^2 + 4x_2^2}} + 2x_1 \right) \\ t(\mathbf{x}) \left(\frac{4x_2}{\sqrt{(x_1^2 - 1)^2 + 4x_2^2}} \right) \end{bmatrix} = \begin{bmatrix} t(\mathbf{x})a(\mathbf{x}) \\ t(\mathbf{x})b(\mathbf{x}) \end{bmatrix}
\end{aligned}$$

onde $t(\mathbf{x}) = 2 \left(\sqrt{(x_1^2 - 1)^2 + 4x_2^2} + x_1^2 - 1 \right)$. Assim, a substituição por \mathbf{x}^k fornece

$$\nabla f(\mathbf{x}^k) + \lambda^k \nabla h(\mathbf{x}^k) + \mu^k \nabla g(\mathbf{x}^k) = \begin{bmatrix} \frac{2t(\mathbf{x}^k)}{k} \left(\frac{\left(1 - \frac{1}{k^2}\right)}{\sqrt{\left(1 - \frac{1}{k^2}\right)^2 + \frac{4}{k^2}}} - 1 \right) \\ t(\mathbf{x}^k) \left(\frac{\frac{4}{k}}{\sqrt{\left(1 - \frac{1}{k^2}\right)^2 + \frac{4}{k^2}}} \right) \end{bmatrix} + \lambda^k \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \mu^k \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Veja que para todo k temos $g(\mathbf{x}^k) \geq 0$, e devemos ter $\mu^k > 0$. Como $t(-1/k, 1/k) > 0$ e

$$\frac{\left(1 - \frac{1}{k^2}\right)}{\sqrt{\left(1 - \frac{1}{k^2}\right)^2 + \frac{4}{k^2}}} < 1,$$

então $t(\mathbf{x}^k)a(\mathbf{x}^k) < 0$. Logo, podemos definir $\mu^k = -[t(\mathbf{x}^k)a(\mathbf{x}^k)]^{-1} > 0$. Definindo $\lambda^k = -\mu^k t(\mathbf{x}^k)b(\mathbf{x}^k)$ segue que

$$\nabla f(\mathbf{x}^k) + \lambda^k \nabla h(\mathbf{x}^k) + \mu^k \nabla g(\mathbf{x}^k) = \mathbf{0}$$

para todo k . Além disso $\mathbf{x}^k \rightarrow (0, 0)$, $h(\mathbf{x}^k) \rightarrow 0$ e $g(\mathbf{x}^k) \rightarrow 0$. Portanto, $(0, 0)$ é AKKT mas não é solução do problema. \square

4.2 Da condição sequencial para KKT

AKKT é uma versão aproximada de KKT, que, como vimos, diferentemente de KKT, é satisfeita sobre qualquer minimizador de (P) independentemente de hipóteses adicionais.

É interessante, no entanto, conectar AKKT com o conceito pontual KKT. Em primeiro lugar, é trivial ver que todo ponto KKT é AKKT (basta tomar sequências constantes iguais a $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$). A recíproca nem sempre vale, como vimos no Exemplo 4.1. Ou seja, nem todo ponto AKKT é KKT. Nos perguntamos então sob quais condições (de regularidade) um ponto AKKT é também KKT, isto é, quais condições CQ (condição de qualificação) satisfazem “AKKT \wedge CQ \Rightarrow KKT”. Note que uma tal CQ é uma condição de qualificação pois se \mathbf{x}^* é minimizador e satisfaz CQ, então pelo Teorema 4.1, \mathbf{x}^* é KKT.

Uma dessas condições é a independência linear dos gradientes das restrições ativas (LICQ). O teorema a seguir mostra que sob LICQ, um ponto AKKT é de fato KKT. Apontando assim, a conexão entre condições sequencial e pontual.

Teorema 4.2. *Se \mathbf{x}^* é um ponto AKKT de (P) que satisfaz LICQ então \mathbf{x}^* é KKT.*

Demonstração. Sejam as sequências $(\mathbf{x}^k) \subset \mathbb{R}^n$, $(\boldsymbol{\lambda}^k) \subset \mathbb{R}^m$, $(\boldsymbol{\mu}^k) \subset \mathbb{R}_+^p$, tais que

- $\lim_{k \rightarrow \infty} \mathbf{x}^k = \mathbf{x}^*$,
- $\lim_{k \rightarrow \infty} \nabla f(\mathbf{x}^k) + \sum_{i=1}^m \lambda_i^k \nabla h_i(\mathbf{x}^k) + \sum_{i=1}^p \mu_i^k \nabla g_i(\mathbf{x}^k) = 0$,
- $\lim_{k \rightarrow \infty} \|h(\mathbf{x}^k)\| = \lim_{k \rightarrow \infty} \|g(\mathbf{x}^k)_+\| = 0$,
- Para todo $i = 1, \dots, p$, $\lim_{k \rightarrow \infty} |\min\{\mu_i^k, -g_i(\mathbf{x}^k)\}| = 0$.

Vamos definir $M_k = \|(1, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)\|_\infty$ para todo k e considerar dois casos:

Caso 1: Se $\{M_k\}$ possui subsequência limitada, então $(\boldsymbol{\lambda}^k)$ e $(\boldsymbol{\mu}^k)$ possuem subsequências convergentes. Tomando tais subsequências e passando ao limite concluímos que \mathbf{x}^* é KKT independentemente de LICQ.

Caso 2: Se $\{M_k\}$ não possui subsequência limitada, dividindo o segundo limite por M_k ,

$$\lim_{k \rightarrow \infty} \frac{\nabla f(\mathbf{x}^k)}{M_k} + \sum_{i=1}^m \frac{\lambda_i^k}{M_k} \nabla h_i(\mathbf{x}^k) + \sum_{i=1}^p \frac{\mu_i^k}{M_k} \nabla g_i(\mathbf{x}^k) = 0,$$

teremos

$$\lim_{k \rightarrow \infty} \sum_{i=1}^m \frac{\lambda_i^k}{M_k} \nabla h_i(\mathbf{x}^k) + \sum_{i=1}^p \frac{\mu_i^k}{M_k} \nabla g_i(\mathbf{x}^k) = 0.$$

Observe que (λ_i^k/M_k) e (μ_j^k/M_k) são limitadas para todos i, j e, pela definição de M_k , pelo menos uma dessas sequências não converge à zero. Também, note que as restrições inativas não aparecem em $\sum_{i=1}^p \frac{\mu_i^k}{M_k} \nabla g_i(\mathbf{x}^k)$ para todo k suficientemente grande, devido ao limite da complementaridade. Logo, a expressão anterior leva à uma combinação linear nula e não trivial dos gradientes das restrições ativas, contradizendo LICQ. \square

Uma condição menos exigente (na verdade, a menos exigente possível em certo sentido) é a Condição de Cone Continuidade (CCP, do inglês *Cone Continuity Property*) [12]. Cabe ressaltar que CCP sempre vale quando as restrições são lineares. Pontos que cumprem uma CQ menos exigente nos dão uma condição de otimalidade mais forte, o que aumenta as chances de tais pontos serem minimizadores.

Como vimos, a maioria dos algoritmos desenvolvidos para resolver um PNL são iterativos. Tais implementações devem incluir critérios de parada que testam na iteração atual se estamos próximos à uma solução. O procedimento feito é justificado teoricamente porque, como provamos no Teorema 4.1, todo minimizador local é um ponto AKKT. Assim, as condições (4.1)–(4.2) da definição de AKKT servem como critério de parada para algoritmos. É o caso do método de Lagrangiano aumentado que veremos no próximo capítulo.

4.3 Outras condições sequenciais de otimalidade na literatura

Existem na literatura outras condições sequenciais, mais fortes que AKKT. Geralmente, elas são descrições inexatas das condições KKT, como a própria condição AKKT. Dentre elas, destacamos AKKT positiva (PAKKT, do inglês *Positive AKKT*) [9] e AKKT complementar (CAKKT, do inglês *Complementary AKKT*) [8]. Uma outra condição que segue esse mesmo princípio é a condição AKKT forte (SAKKT, do inglês *Strong AKKT*). Basicamente, todas essas condições diferem entre si na forma como aproximam a complementaridade. Veja [2] para detalhes.

5 Algoritmos: convergência global via condições sequenciais de otimalidade

As condições sequenciais de otimalidade têm sido usadas no estudo da convergência de algoritmos práticos, na medida em que se demonstra que eles geram pontos AKKT. Veja por exemplo [4, 8, 9, 12–17]. Assim, se supormos que os pontos gerados por tais métodos satisfazem, por exemplo, LICQ, então estes pontos também são KKT. De qualquer forma, independentemente da validade de condições de qualificação, a própria condição sequencial AKKT é um atestado de qualidade do ponto gerado.

5.1 Lagrangiano Aumentado

Vimos que o Método da Penalidade Externa é pouco utilizado na prática. Porém, o seu conceito serve como uma ideia para muitos outros métodos, proveniente de alguns ajustes. Por exemplo, o parâmetro ρ_k pode aumentar muito e causar instabilidades numéricas. Podemos tentar evitar que ρ aumente demais, aumentando-o somente caso seja necessário. Desse modo, podemos aumentar ρ_k se a medida de inviabilidade avaliada em \mathbf{x}^k não for menor que a medida de inviabilidade avaliada em \mathbf{x}^{k-1} , em outras palavras, se na iteração corrente $\|\mathbf{h}(\mathbf{x}^k)\|^2 + \|\mathbf{g}(\mathbf{x}^k)_+\|^2 < \|\mathbf{h}(\mathbf{x}^{k-1})\|^2 + \|\mathbf{g}(\mathbf{x}^{k-1})_+\|^2$ então não aumentamos o valor de ρ_k .

Além disso, sempre nos aproximamos da solução do problema por meio de uma sequência de pontos não viáveis. Isso pode causar uma certa insegurança a respeito da viabilidade do limite da sequência gerada e, além do mais, pontos não viáveis como resultado, na prática, não têm utilidade alguma. Podemos mitigar esse fato fazendo um pequeno “deslocamento” da região viável, para que haja a chance de alguns elementos da sequência $\{\mathbf{x}^k\}$ gerada pelo método estarem, de fato, na região viável. Portanto, ao invés de considerarmos a medida de inviabilidade dada por $\|\mathbf{h}(\mathbf{x}^k)\|^2 + \|\mathbf{g}(\mathbf{x}^k)_+\|^2$, vamos

denota-la com desvio em relação a vetores \mathbf{v}^k e \mathbf{w}^k , isto é, a nova medida de inviabilidade será dada por $\|\mathbf{h}(\mathbf{x}^k) - \mathbf{v}^k\|^2 + \|(\mathbf{g}(\mathbf{x}^k) - \mathbf{w}^k)_+\|^2$.

Entretanto, não é conveniente que \mathbf{v}^k e \mathbf{w}^k sejam grandes demais, do contrário teríamos um grande deslocamento da região viável e estaríamos resolvendo um problema diferente do original. O ideal é que $\mathbf{v}^k \rightarrow \mathbf{0}$ e $\mathbf{w}^k \rightarrow \mathbf{0}$. Assim, podemos aproveitar do fato de que $\rho_k \rightarrow \infty$ e definir

$$\mathbf{v}^k = -\frac{\bar{\boldsymbol{\lambda}}^k}{\rho_k} \quad \text{e} \quad \mathbf{w}^k = -\frac{\bar{\boldsymbol{\mu}}^k}{\rho_k},$$

onde $\bar{\lambda}_i^1 \in [\lambda_{min}, \lambda_{max}]$, para todo $i = 1, \dots, m$ e $\bar{\mu}_i^1 \in [0, \mu_{max}]$, para todo $i = 1, \dots, p$.

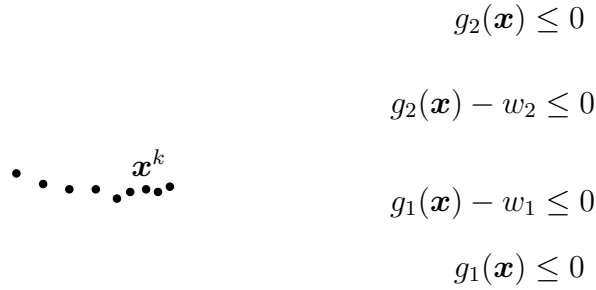


Figura 7: A figura representa a região viável original e a do subproblema, que é deslocada por $\mathbf{w}^k = -\boldsymbol{\mu}^k/\rho_k$, $\rho_k > 0$. A minimização no subproblema ocorre em relação ao problema deslocado (restrições com \mathbf{w}). O iterando \mathbf{x}^k , soluções do subproblema, está próximo à região hachurada. Note que, em relação ao problema original, sem deslocamento, \mathbf{x}^k é viável.

É com essas ideias que obtemos o Método do Lagrangiano Aumentado. A função Lagrangiano aumentado é definida por

$$L_\rho(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \frac{\rho}{2} \left\{ \sum_{i=1}^m \left[h_i(\mathbf{x}) + \frac{\lambda_i}{\rho} \right]^2 + \sum_{i=1}^p \left[g_i(\mathbf{x}) + \frac{\mu_i}{\rho} \right]^2 \right\}_+.$$

Na iteração k do método, para dados $\rho_k > 0$, $\bar{\boldsymbol{\mu}}^k$ e $\bar{\boldsymbol{\lambda}}^k$, obtemos um candidato \mathbf{x}^k a minimizador da função acima (em \mathbf{x}), isto é,

$$\begin{aligned} \nabla_{\mathbf{x}} L_{\rho_k}(\mathbf{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k) &= \mathbf{0} \\ \Rightarrow \nabla f(\mathbf{x}^k) + \sum_{i=1}^m \rho_k \left(h_i(\mathbf{x}^k) + \frac{\bar{\lambda}_i^k}{\rho_k} \right) \nabla h_i(\mathbf{x}^k) + \sum_{i=1}^p \rho_k \left(g_i(\mathbf{x}^k) + \frac{\bar{\mu}_i^k}{\rho_k} \right)_+ \nabla g_i(\mathbf{x}^k) &= \mathbf{0} \end{aligned}$$

$$\Rightarrow \nabla f(\mathbf{x}^k) + \sum_{i=1}^m [\bar{\lambda}_i^k + \rho_k h_i(\mathbf{x}^k)] \nabla h_i(\mathbf{x}^k) + \sum_{i=1}^p [\bar{\mu}_i^k + \rho_k g_i(\mathbf{x}^k)]_+ \nabla g_i(\mathbf{x}^k) = \mathbf{0}.$$

Os termos $\bar{\lambda}_i^k + \rho_k h_i(\mathbf{x}^k)$ e $\bar{\mu}_i^k + \rho_k g_i(\mathbf{x}^k)$ acima fazem o papel de multiplicadores de Lagrange de (P). Logo, na aplicação do método podemos atualizar as estimativas de multiplicadores $\bar{\lambda}^{k+1}$ e $\bar{\mu}^{k+1}$ da próxima iteração da seguinte forma: calculamos

$$\lambda_i^{k+1} = \bar{\lambda}_i^k + \rho_k h_i(\mathbf{x}^k), \forall i = 1, \dots, m \quad , \quad \mu_i^{k+1} = \bar{\mu}_i^k + \rho_k g_i(\mathbf{x}^k), \forall i = 1, \dots, p$$

e fazemos $\bar{\lambda}_i^{k+1}$ e $\bar{\mu}_i^{k+1}$ iguais às projeções de λ_i^{k+1} e μ_i^{k+1} nos intervalos $[\lambda_{min}, \lambda_{max}]$ e $[0, \mu_{max}]$, respectivamente. Assim, $\bar{\lambda}$ e $\bar{\mu}$ imitam os multiplicadores da solução de (P) (multiplicadores projetados). O fato de utilizarmos projeções tais estimativas, que são limitadas, é conhecido como estratégia de *salvaguardas*. Assim, enunciaremos o método Lagrangiano aumentado com salvaguardas. Por simplicidade, omitiremos o termo “salvaguardas” daqui para frente.

No seguinte exemplo pode-se observar a relevância dos ajustes incluídos no método da Penalidade Externa. Em seguida, apresentamos o método do Lagrangiano aumentado no Algoritmo 1.

Exemplo 5.1. Vamos retomar o problema do Exemplo 3.1,

$$\text{minimizar } f(x) = x \quad \text{sujeito a } g(x) = -x \leq 0.$$

Vimos que o Método da Penalidade Externa (sem deslocamento) deve executar “infinitas” iterações para atingir a solução $x^* = 0$. Ao aplicar o método com deslocamento o subproblema consistirá em resolver

$$\text{minimizar } x + \frac{\rho_k}{2} \left(\frac{\bar{\mu}^k}{\rho_k} - x \right)_+^2 \quad \text{sujeito a } x \in \mathbb{R}.$$

Sejam dados $\bar{\mu}_1 = 0, \rho_1 = 1$ e $x_0 = -2$. Suponha que $x \leq 0$. Na primeira iteração ($k = 1$) devemos resolver o subproblema:

$$\text{minimizar } x + \frac{\rho_1}{2} \left(\frac{\bar{\mu}_1}{\rho_1} - x \right)_+^2 = x + \frac{1}{2} x^2 \quad \text{sujeito a } x \in \mathbb{R}.$$

Assim, derivando a função objetivo do subproblema e igualando a zero obtemos

$$1 + x = 0 \Rightarrow x = -1$$

Logo, $x_1 = -1$. Note que $x_1 < 0$, o que é compatível com a hipótese. Vamos prosseguir com o método atualizando os parâmetros da nova iteração.

Novo multiplicador projetado:

$$\bar{\mu}_2 = (\bar{\mu}_1 - \rho_1 x_1)_+ = 1.$$

Novo penalizador: note que a viabilidade melhorou pois $(-x_1)_+ = 1 < 2 = (-x_0)_+$. Então manteremos o valor

$$\rho_2 = \rho_1 = 1.$$

Desta forma, supondo $x \leq 1$, avançamos para a próxima iteração ($k = 2$), onde o subproblema consistirá em resolver:

$$\text{minimizar } x + \frac{\rho_2}{2} \left(\frac{\bar{\mu}_2}{\rho_2} - x \right)_+^2 = x + \frac{1}{2} (1 - x)^2 \quad \text{sujeito a } x \in \mathbb{R}.$$

Novamente, derivando a função objetivo e igualando a zero, temos

$$1 - (1 - x) = 0 \Rightarrow x = 0$$

Portanto $x_2 = 0$ e chegamos na solução com apenas 2 iterações. Ressalta-se que já conhecemos a solução do problema, mas o método em si, não conhece. Assim, as execuções do método devem parar após verificar que $x_2 = 0$ é um ponto KKT. \square

Algoritmo 1 Lagrangiano Aumentado

Entrada: $\mathbf{x}^0 \in \mathbb{R}^n$ um ponto inicial arbitrário, $\tau \in [0, 1)$, $\gamma > 1$, $-\infty < \lambda_{min} < \lambda_{max} < \infty$, $\mu_{max} > 0$, $\rho_1 > 0$, $\bar{\lambda}_i^1 \in [\lambda_{min}, \lambda_{max}] \forall i = 1, \dots, m$, $\bar{\mu}_i^1 \in [0, \mu_{max}] \forall i = 1, \dots, p$, uma sequência $(\varepsilon_k) \subset \mathbb{R}_{++}$ com $\lim_k \varepsilon_k = 0$.

Saída: \mathbf{x}^k .

Passo 1: *Inicialização.*

$k \leftarrow 1$. Definir $V^0 = \mathbf{g}(\mathbf{x}^0)_+$.

Passo 2: *Resolução do subproblema*

Calcular $\mathbf{x}^k \in \mathbb{R}^n$ satisfazendo

$$\|\nabla_x L_{\rho_k}(\mathbf{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k)\| \leq \varepsilon_k$$

Passo 3: *Estimar multiplicadores*

Para todo $i = 1, \dots, m$, calcular

$$\lambda_i^{k+1} = \bar{\lambda}_i^k + \rho_k h_i(\mathbf{x}^k) \quad \text{e} \quad \bar{\lambda}_i^{k+1} \in [\lambda_{min}, \lambda_{max}].$$

Para todo $i = 1, \dots, p$, calcular

$$\mu_i^{k+1} = \max\{0, \bar{\mu}_i^k + \rho_k g_i(\mathbf{x}^k)\}, \quad \bar{\mu}_i^{k+1} \in [0, \mu_{max}] \quad \text{e} \quad V_i^k = \max\left\{g_i(\mathbf{x}^k), -\frac{\bar{\mu}_i^k}{\rho_k}\right\}.$$

Passo 4: Atualizar parâmetro de penalidade

Se

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

definir

$$\rho_{k+1} = \rho_k.$$

Senão, definir

$$\rho_{k+1} = \gamma \rho_k.$$

Passo 5: Nova iteração

Atualizar $k \leftarrow k + 1$. Voltar a Passo 2.

O teorema a seguir garante que o método do Lagrangiano Aumentado gera pontos AKKT de (P).

Teorema 5.1. *Seja $\{\mathbf{x}^k\}$ a sequência gerada pelo Algoritmo 1.*

1. *Se \mathbf{x}^* é um ponto de acumulação de $\{\mathbf{x}^k\}$ viável, então \mathbf{x}^* é ponto AKKT do problema original (P);*
2. *Se \mathbf{x}^* é um ponto de acumulação de $\{\mathbf{x}^k\}$ (não necessariamente viável), então \mathbf{x}^* é ponto KKT do problema de inviabilidade*

$$\min_{\mathbf{x}} \|\mathbf{h}(\mathbf{x})\|_2^2 + \|\mathbf{g}(\mathbf{x})_+\|_2^2. \quad (5.1)$$

Demonstração. Para a demonstração dos dois itens, consideremos dois casos:

Caso 1: $\{\rho_k\}$ é limitada. Neste caso, a partir de certa iteração ocorre

$$\max\{\|\mathbf{h}(\mathbf{x}^{k-1})\|_\infty, \|V^{k-1}\|_\infty\} \leq \tau \max\{\|\mathbf{h}(\mathbf{x}^k)\|_\infty, \|V^k\|_\infty\}, \quad \text{onde } 0 \leq \tau < 1,$$

isto é, o teste de viabilidade deu certo para todo k suficientemente grande. Assim,

$$\lim_{k \rightarrow \infty} \|\mathbf{h}(\mathbf{x}^k)\|_\infty = \mathbf{0} \quad \text{e} \quad \lim_{k \rightarrow \infty} \|V^k\|_\infty = \mathbf{0}.$$

Ou seja,

$$\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$$

e se $g_j(\mathbf{x}^*) > 0$ então

$$g_j(\mathbf{x}^k) \geq \frac{g_j(\mathbf{x}^*)}{2} > 0.$$

Assim

$$V_j^k \geq g_j(\mathbf{x}^k) \geq \frac{g_j(\mathbf{x}^*)}{2} > 0,$$

o que é um absurdo, pois desta forma teríamos $V_j^k \not\rightarrow 0$. Portanto,

$$g_j(\mathbf{x}^*) \leq 0 \quad \text{para todo } j = 1, \dots, p$$

e \mathbf{x}^* é um ponto viável, independentemente da hipótese do item 1. Além disso, a sequência $\{\rho_k\}$ e as sequências de multiplicadores projetados $\{\bar{\boldsymbol{\lambda}}^{k+1}\}$ e $\{\bar{\boldsymbol{\mu}}^{k+1}\}$, do Passo 3, são limitadas. Portanto, admitem subsequências convergentes. Ademais, note que no Passo 2

$$\|\nabla_x L_{\rho_k}(\mathbf{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k)\| \leq \varepsilon_k,$$

e note também que $V_j^k \rightarrow 0$ garante que tenhamos complementariedade no limite, isto é, para todo $j = 1, \dots, p$

$$|\min\{\mu_j^k, -g_j(\mathbf{x}^k)\}| \rightarrow 0.$$

Concluimos assim que \mathbf{x}^* é ponto AKKT de (P). Este ponto também será KKT do problema de inviabilidade

$$\min_{\mathbf{x}} \|\mathbf{h}(\mathbf{x})\|_2^2 + \|\mathbf{g}(\mathbf{x})_+\|_2^2$$

pois ele é viável, e logo minimizador global deste problema irrestrito.

Caso 2: $\{\rho_k\}$ é ilimitada. A partir da aplicação do método, sabemos que

$$\|\nabla_x L_{\rho_k}(\mathbf{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k)\| \leq \varepsilon_k.$$

Agora, para que tenhamos AKKT precisamos mostrar que

$$|\min\{\mu_i^k, -g_i(\mathbf{x}^k)\}| \leq \varepsilon_k,$$

onde $\varepsilon_k \rightarrow 0$. Se $g_j(\mathbf{x}^k) \rightarrow 0$, de fato, temos a inequação acima. Se considerarmos que $g_j(\mathbf{x}^*) < 0$, como $\{\bar{\boldsymbol{\mu}}^k\}$ é limitada e como $g_j(\mathbf{x}^{k-1}) \leq \frac{g_j(\mathbf{x}^*)}{2} < 0$ para todo k grande

então

$$\bar{\mu}_j^{k-1} + \rho_{k-1}g_j(\mathbf{x}^{k-1}) \leq \bar{\mu}_j^{k-1} + \frac{\rho_{k-1}}{2}g_j(\mathbf{x}^*) \rightarrow -\infty.$$

Logo, $\mu_j^k = [\bar{\mu}_j^{k-1} + \rho_k g_j(\mathbf{x}^{k-1})]_+ = 0$ para todo k grande, concluindo que \mathbf{x}^* é um ponto AKKT de (P).

Resta mostrar que \mathbf{x}^* é ponto KKT do problema de inviabilidade. Para isso, tomemos

$$\nabla_x L_{\rho_k}(\mathbf{x}^k, \bar{\boldsymbol{\mu}}^k, \bar{\boldsymbol{\lambda}}^k) = \delta_k$$

onde $\lim_{k \rightarrow \infty} \|\delta^k\| = 0$. Dividindo por ρ_k obteremos

$$\frac{\nabla f(\mathbf{x}^k)}{\rho_k} + \sum_{i=1}^m \left[\frac{\bar{\lambda}_i^k}{\rho_k} + h_i(\mathbf{x}^k) \right] \nabla h_i(\mathbf{x}^k) + \sum_{i=1}^p \left[\frac{\bar{\mu}_i^k}{\rho_k} + g_i(\mathbf{x}^k) \right]_+ \nabla g_i(\mathbf{x}^k) = \frac{\delta_k}{\rho_k}.$$

Passando ao limite e depois multiplicando os termos por 2, vem

$$\sum_{i=1}^m 2[h_i(\mathbf{x}^*)] \nabla h_i(\mathbf{x}^*) + \sum_{i=1}^p 2[g_i(\mathbf{x}^*)]_+ \nabla g_i(\mathbf{x}^*) = 0.$$

Ou seja, o gradiente de

$$\sum_{i=1}^m [h_i(\mathbf{x})]^2 + \sum_{i=1}^p [g_i(\mathbf{x})]_+^2 = \|\mathbf{h}(\mathbf{x})\|_2^2 + \|\mathbf{g}(\mathbf{x})\|_2^2$$

se anula em \mathbf{x}^* . Portanto \mathbf{x}^* é um ponto KKT do problema de inviabilidade. \square

O teorema anterior diz que o Algoritmo 1 gera pontos AKKT. Com isso, sua convergência a pontos KKT fica automaticamente estabelecida com a condição de independência linear dos gradientes das restrições ativas, pela aplicação direta do Teorema 4.2.

Corolário 5.1. *Se \mathbf{x}^* é ponto de acumulação viável da sequência gerada pelo Algoritmo 1 e satisfaz LICQ, então \mathbf{x}^* é ponto KKT de (P).*

Como comentado na seção 4.2, existem outras condições de qualificação menos exigentes que LICQ associadas à AKKT. A menos exigente possível, conforme comentamos no capítulo anterior, é a condição de cone continuidade (CCP). Como o Teorema 5.1, item 1, conecta a convergência do método à condição AKKT, fica também estabelecida a convergência do Algoritmo 1 com a condição CCP. Este resultado é muito mais forte que o Corolário 4.1. Porém, neste trabalho, não abordamos a condição CCP. Para maiores detalhes, consulte [17].

Existe a chance dos pontos gerados pelo Algoritmo (1) não serem viáveis. Basicamente, o Teorema 5.1 item 2 diz que, se \mathbf{x}^* não é um ponto viável, então é provavelmente um minimizador da inviabilidade, ou seja, \mathbf{x}^* é um bom candidato à solução do problema de inviabilidade (5.1), que visa minimizar a violação das restrições. Isso é especialmente útil quando estamos tentando resolver um problema inviável: o algoritmo retornará, com grande chance, aquele ponto com menor inviabilidade possível.

5.2 Notas

As condições sequenciais não estão associadas apenas ao método do Lagrangiano Aumentado. Outros métodos também geram pontos AKKT. Dentre eles podemos mencionar os métodos de programação quadrática sequencial desenvolvido em [18] e de pontos interiores apresentado em [19].

Apesar do grande potencial das condições sequenciais em generalizar o estudo da convergência de diferentes métodos, como demonstrado nas referências deste capítulo, nem todo algoritmo gera uma sequência/ponto AKKT. Recentemente, demonstrou-se que o método de Newton-Lagrange pode falhar em detectar convergência usando as medidas de proximidade à solução tipo AKKT [20]; por outro lado, não se sabe se sua versão “estabilizada” – muito usada na prática – gera AKKT. Ainda assim, condições sequenciais seguem atraindo a atenção da comunidade científica. Recentemente, condições sequenciais foram definidas para problemas particulares, como programação matemática com restrições de complementaridade, problemas multiobjetivo e equilíbrio de Nash. Mais ainda, estão sendo levadas à contextos gerais, como programação cônica e otimização em espaços de Banach.

Conclusões

Neste trabalho falamos especificamente da condição sequencial de otimalidade conhecida como Aproximadamente KKT (AKKT). Tal condição é comumente utilizada na prática numérica como forma de obter as condições KKT de forma aproximada. Sua aplicação também possui justificativa teórica e funciona bem para muitos problemas. Vimos também que todo minimizador local é um ponto AKKT e que um ponto AKKT sob a condição de qualificação LICQ também é um ponto KKT. Há também outras condições sequências na literatura, que não foram abordados neste trabalho. Um dos métodos mais conhecidos que geram pontos AKKT é método do Lagrangiano Aumentado definido em [5]. Falamos sobre a sua construção e argumentamos que ele gera naturalmente pontos AKKT.

Diante disto, podem ser estudados também a aplicação das condições AKKT em outros métodos de otimização, outras condições sequenciais de otimalidade (por exemplo, as que foram citadas no trabalho), a aplicação de tais condições em distintos problemas, os resultados ao implementar e utilizar como critério de parada diferentes condições sequencias de otimalidade e também uma comparação na aplicação em diferentes métodos de otimização.

O leitor interessado no tema das condições sequenciais pode consultar o livro de Martínez [3], o artigo [9] e suas referências.

Referências Bibliográficas

- 1 RIBEIRO, A. A.; KARAS, E. W. *Otimização contínua: Aspectos teóricos e computacionais*. [S.l.]: Cengage Learning, 2014.
- 2 RAMOS, J. A. *Tópicos em Condição de Otimalidade para Otimização não Linear*. Tese (Doutorado) — IME, São Paulo, jan. 2016. In Portuguese.
- 3 MARTÍNEZ, J. M. *Otimização prática usando o Lagrangiano aumentado*. [S.l.], 2009. Disponível em: <<http://www.ime.unicamp.br/~martinez/lagraum.pdf>>.
- 4 ANDREANI, R.; HAESER, G.; MARTÍNEZ, J. M. On sequential optimality conditions for smooth constrained optimization. *Optimization*, v. 60, n. 5, p. 627–641, 2011.
- 5 ANDREANI, R. et al. On augmented Lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization*, v. 18, n. 4, p. 1286–1309, 2007.
- 6 LIMA, E. L. *Curso de Análise vol. 2*. 8. ed. Rio de Janeiro: IMPA, 2005.
- 7 MARTÍNEZ, J. M.; SANTOS, S. A. *Métodos computacionais de otimização*. [S.l.], 2020. No prelo. Disponível em: <http://www.ime.unicamp.br/~sandra/MT601/handouts/MCDO_set2020.pdf>.
- 8 ANDREANI, R.; MARTÍNEZ, J. M.; SVAITER, B. F. A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization*, v. 20, n. 6, p. 3533–3554, 2010.
- 9 ANDREANI, R. et al. A sequential optimality condition related to the quasi-normality constraint qualification and its algorithmic consequences. *SIAM Journal on Optimization*, v. 29, n. 1, p. 743–766, 2019.
- 10 ANDREANI, R. et al. New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences. *SIAM Journal on Optimization*, v. 29, n. 4, p. 3201–3230, 2019.
- 11 MARTÍNEZ, J. M.; SVAITER, B. F. A practical optimality condition without constraint qualifications for nonlinear programming. *Journal of Optimization Theory and Applications*, v. 118, n. 1, p. 117–133, 2003. ISSN 1573–2878.
- 12 ANDREANI, R. et al. Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Mathematics of Operations Research*, v. 43, n. 3, p. 693–717, 2018.
- 13 ANDREANI, R. et al. *On the convergence of augmented Lagrangian strategies for nonlinear programming*. [S.l.], 2020. Disponível em: <http://www.optimization-online.org/DB_HTML/2020/03/7701.html>.

- 14 ANDREANI, R. et al. A second-order sequential optimality condition associated to the convergence of optimization algorithms. *IMA Journal of Numerical Analysis*, v. 37, n. 4, p. 1902–1929, 2017.
- 15 HAESER, G. A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms. *Computational Optimization and Applications*, v. 70, n. 2, p. 615–639, 2018.
- 16 ANDREANI, R. et al. On second-order optimality conditions in nonlinear optimization. *Optimization Methods and Software*, v. 32, n. 1, p. 22–38, 2017.
- 17 ANDREANI, R. et al. A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization*, v. 26, n. 1, p. 96–110, 2016.
- 18 QI, L.; WEI, Z. On the constant positive linear dependence condition and its application to SQP methods. *SIAM Journal on Optimization*, v. 10, n. 4, p. 963–981, 2000.
- 19 CHEN, L.; GOLDFARB, D. Interior-point ℓ_2 -penalty methods for nonlinear programming with strong global convergence properties. *Mathematical Programming*, v. 108, n. 1, p. 1–36, 2006.
- 20 ANDREANI, R.; MARTÍNEZ, J. M.; SANTOS, L. T. Newton’s method may fail to recognize proximity to optimal points in constrained optimization. *Mathematical Programming*, v. 160, n. 1, p. 547–555, 2016.