

Universidade Federal de Juiz de Fora  
Instituto de Ciências Exatas  
Programa de Pós-Graduação em Matemática

**ErasmO Tales Fonseca**

**FDIPA - Algoritmo de Pontos Interiores e Direções Viáveis para Otimização  
Não-Linear Diferenciável: Um Estudo de Parâmetros**

Juiz de Fora

2015

**ErasmO Tales Fonseca**

**FDIPA - Algoritmo de Pontos Interiores e Direções Viáveis para Otimização  
Não-Linear Diferenciável: Um Estudo de Parâmetros**

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal de Juiz de Fora, na área de concentração em Matemática Aplicada, como requisito parcial para obtenção do título de Mestre em Matemática.

Orientador: Wilhelm Passarela Freire

Juiz de Fora

2015

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF  
com os dados fornecidos pelo(a) autor(a)

Fonseca, Erasmo T..

FDIPA - Algoritmo de Pontos Interiores e Direções Viáveis para Otimização Não-Linear Diferenciável: Um Estudo de Parâmetros / Erasmo Tales Fonseca. – 2015.

66 f. : il.

Orientador: Wilhelm Passarela Freire

Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Matemática, 2015.

1. Otimização Não Linear. 2. Método de Pontos Interiores. 3. Algoritmo de direções viáveis.

ErasmO Tales Fonseca

**FDIPA - Algoritmo de Pontos Interiores e Direções Viáveis para Otimização  
Não-Linear Diferenciável: Um Estudo de Parâmetros**

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal de Juiz de Fora, na área de concentração em Matemática Aplicada, como requisito parcial para obtenção do título de Mestre em Matemática.

Aprovada em: 06 de Novembro de 2015.

BANCA EXAMINADORA

---

Prof. Dr. Wilhelm Passarella Freire - Orientador  
Universidade Federal de Juiz de Fora

---

Professor Dr. Sandro Rodrigues Mazorche  
Universidade Federal de Juiz de Fora

---

Professor Dr. Alexandre Rocha Duarte  
Instituto Federal do Sudeste de Minas Gerais

## AGRADECIMENTOS

A Deus por me amparar nos momentos difíceis, por me fortalecer nas dificuldades e por ter sido o meu socorro bem presente na angústia.

Aos meus pais Leonardo Martins Fonseca e Maria José Nonato Silva Fonseca e também aos meus irmãos Herbert, Lenusa e Meire Elem por sempre acreditarem em mim, por sonharem comigo cada sonho que idealizo e por não medirem esforços para que eu possa realizá-los.

Ao meu orientador Wilhelm Passarella Freire pelo apoio e orientação.

A Livia pelos momentos vividos juntos. Obrigado por dividir comigo as angústias e alegrias. Foi bom poder contar com você!

A Taís e Bia pela convivência e por tantas alegrias compartilhadas.

Aos meus amigos Kaká, Sulamita, Letícia e Dayane por terem me ajudado sempre que precisei.

Aos meus irmãos em Cristo pelas orações e pelas palavras de conforto que sempre me direcionaram.

A Tia Gilva e Gilvânia por tantas vezes me fazerem sorrir.

Aos meus colegas Hugo, Jorge, Sandra, Eliza, Mariana, Eduardo, Julio Sangay e Vladimir por sempre me ajudarem.

A Laura, Paula e Roberta pelo carinho e atenção.

Aos professores pelos ensinamentos.

Agradeço também à CAPES pelo apoio financeiro.

Ninguém vence sozinho e sem vocês eu jamais teria chegado até aqui.

Muito obrigado a todos vocês!

Quero implantar no mundo,  
Um pouco mais de Otimização.  
Quero torna-lo mais profundo,  
Para ouvir o meu coração.  
Quero minimizar as dores,  
E restringir todo tipo de guerra.  
Quero um mundo cheio de flores,  
E felicidade em toda Terra.  
Quero maximizar as alegrias,  
Sem restrições quero igualdade.  
Quero encher os meus dias,  
De muita felicidade.  
(ErasmO Tales Fonseca)

## RESUMO

Neste trabalho apresentamos um estudo da influência dos parâmetros de um algoritmo de pontos interiores e direções viáveis para solução de problemas de otimização não linear. Esse algoritmo, denominado FDIPA, tem por objetivo encontrar dentre os pontos de um conjunto definido por restrições de igualdade e/ou desigualdade, aqueles que minimizam uma função diferenciável. O FDIPA baseia-se na resolução de dois sistemas de equações lineares com a mesma matriz de coeficientes, obtidos das condições necessárias de primeira ordem de Karush-Kuhn-Tucker. A partir de um ponto inicial no interior do conjunto viável, o FDIPA gera uma sequência de pontos também interiores ao conjunto. Em cada iteração, uma nova direção de descida é obtida e, em seguida, produz-se uma deflexão da direção de descida no sentido do interior do conjunto viável, de modo a se obter uma nova direção que seja de descida e viável. Realiza-se então uma busca linear para obter um novo ponto interior e garantir a convergência global do método. Uma família de algoritmos pode ser obtida variando-se as regras de atualização dos parâmetros do FDIPA. O estudo apresentado neste trabalho foi feito considerando-se um único algoritmo e com restrições de desigualdade somente. Testes numéricos apontaram para uma escolha de parâmetros que levou a um número menor de iterações na resolução dos problemas teste.

Palavras-chave: Otimização não linear. Método de pontos interiores. Algoritmo de direções viáveis.

## ABSTRACT

This work presents a study on the influence of the parameters of an interior point and feasible directions algorithm for solving non-linear problems. The algorithm, named FDIPA, aims to find among the points of a set defined by equality and/or inequality constraints, those which minimize a differentiable function. The FDIPA is based on two linear systems with the same coefficient matrix, obtained from the Karush-Kuhn-Tucker first order necessary conditions. From a initial point in the interior of the feasible set, FDIPA generates a sequence of points which are also interior to the set. At each iteration, FDIPA produces a descent direction which is deflected towards the interior of the feasible set in order to create a new descent and feasible direction. Then, a linear search is performed to get a new interior point and assure the global convergence of the method. A family of algorithms can be obtained varying the rules used to update the parameters of the FDIPA. The study presented here has been done considering just one particular algorithm and inequality constraints only. Numerical tests pointed to a certain choice of parameters which led to a fewer number of iterations when solving some test problems.

Key-words: Non-linear optimization. Interior point methods. Feasible directions algorithms.



## LISTA DE SÍMBOLOS

$\mathbb{R}$	Conjunto dos número reais.
$\mathbb{R}^n$	Espaço euclidiano n-dimensional.
$x \in \mathbb{R}^n$	Vetor coluna: $x^T = (x_1, x_2, \dots, x_n)$ .
$x^T y$	Produto escalar.
$\mathbb{R}^{m \times n}$	Conjunto das matrizes de ordem $m \times n$ .
$A^T$	Matriz transposta da matriz $A \in \mathbb{R}^{m \times n}$ .
I	Matriz identidade ( $a_{ij} = 0$ se $i \neq j$ e $a_{ii} = 1$ ).
$\nabla f$	Gradiente de $f$ :

$$\nabla^T f(x) = \left( \frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right).$$

$Hf(x)$	Matriz hessiana de $f$ :
---------	--------------------------

$$Hf(x) = \nabla^2 f(x).$$

$Jf(a)$	Matriz jacobiana da função $f : X \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ no ponto $a$
---------	---

$$Jf(a) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \cdots & \frac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \frac{\partial f_m}{\partial x_2}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{bmatrix}$$

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>10</b>
<b>2</b>	<b>CONCEITOS PRELIMINARES . . . . .</b>	<b>13</b>
2.1	SEQUÊNCIAS . . . . .	13
2.2	NOÇÕES DE TOPOLOGIA . . . . .	13
2.3	FUNÇÕES DIFERENCIÁVEIS . . . . .	14
2.4	ÁLGEBRA LINEAR . . . . .	15
2.5	DEFINIÇÕES E RESULTADOS CLÁSSICOS . . . . .	16
<b>3</b>	<b>CONDIÇÕES DE OTIMALIDADE . . . . .</b>	<b>18</b>
3.1	CONDIÇÕES DE OTIMALIDADE PARA PROBLEMAS IRRESTRITOS	18
3.2	CONDIÇÕES DE OTIMALIDADE PARA PROBLEMAS COM RES- TRIÇÕES DE IGUALDADE . . . . .	20
3.3	CONDIÇÕES DE OTIMALIDADE PARA PROBLEMAS COM RES- TRIÇÕES DE IGUALDADE E DESIGUALDADE . . . . .	24
<b>4</b>	<b>ALGORITMOS DE PROGRAMAÇÃO NÃO-LINEAR . . . . .</b>	<b>26</b>
4.1	MÉTODOS DE BUSCA UNIDIRECIONAL . . . . .	28
4.1.1	<b>Busca Unidirecional Exata . . . . .</b>	<b>28</b>
4.1.2	<b>Busca de Armijo . . . . .</b>	<b>29</b>
4.2	MÉTODO DO GRADIENTE . . . . .	30
4.3	MÉTODO DE NEWTON . . . . .	31
4.4	MÉTODOS QUASE-NEWTON . . . . .	32
4.4.1	<b>Algoritmo Básico . . . . .</b>	<b>32</b>
4.4.2	<b>O método DFP . . . . .</b>	<b>33</b>
4.4.3	<b>O método BFGS . . . . .</b>	<b>33</b>
<b>5</b>	<b>ALGORITMO DE PONTOS INTERIORES E DIREÇÕES VIÁ- VEIS PARA OTIMIZAÇÃO NÃO-LINEAR . . . . .</b>	<b>34</b>
5.1	ALGORITMO PARA O PROBLEMA COM RESTRIÇÕES DE DESI- GUALDADE . . . . .	34
5.2	INCLUSÃO DE RESTRIÇÕES DE IGUALDADE . . . . .	40
<b>6</b>	<b>ANÁLISE DA INFLUÊNCIA DE PARAMÊTROS NO FDIPA</b>	<b>42</b>
6.1	INTERVALOS PARA $\xi$ . . . . .	45
6.2	INTERVALOS PARA $\nu$ . . . . .	45
6.3	INTERVALOS PARA $\eta$ . . . . .	46

6.4	INTERVALOS PARA $\varphi$ . . . . .	46
6.5	ANÁLISE DOS PARÂMETROS $\omega$ E $\lambda$ . . . . .	47
6.6	CONCLUSÕES . . . . .	47
7	CONSIDERAÇÕES FINAIS . . . . .	49
	REFERÊNCIAS . . . . .	50
	APÊNDICE A – Problemas Utilizados Para Realização dos Tes- tes . . . . .	52
	APÊNDICE B – Valores Atribuídos aos Parâmetros Para a Realização dos Testes Numéricos . . . . .	66

## 1 INTRODUÇÃO

Todos nós, de alguma maneira, já vivenciamos uma situação de ter que otimizar algo em nosso dia a dia. Quem nunca precisou encurtar um caminho para ganhar tempo ou economizar dinheiro para comprar algo? Todas as vezes que procuramos formas ótimas de aplicarmos nossos recursos, estamos trabalhando com um problema de otimização.

Para o consumidor que deseja adquirir determinado produto, o preço ideal para este produto seria o menor possível, ou seja, seria 0. No entanto, sabemos que na verdade isso é impossível, pois a empresa que produz o produto tem gastos com mão de obra, matéria prima, embalagem, etc. Desta forma, existe um preço mínimo para esta empresa colocar este produto em circulação no mercado. Além deste preço mínimo, com certeza a empresa irá querer ganhar um lucro sobre este produto. O lucro ótimo seria sempre o maior possível, e para isso eles deveriam elevar ao máximo o preço do produto. Mas, se isto acontecer o produto deixa de ser um bem acessível ao consumidor. A otimização trata exatamente desta busca pela melhor alocação de um conjunto limitado de recursos, escolhendo a alternativa que maximize o lucro, ou minimize o custo, dentre todas aquelas que satisfazem um conjunto específico de restrições.

O campo teórico de otimização consiste no estudo de problemas nos quais deseja-se *minimizar* ou *maximizar* funções matemáticas. Muitas vezes, estas funções estão sujeitas a algumas condições as quais chamamos de restrições. A otimização possui aplicações em diversas áreas como economia, estatística, medicina, biologia, localização, transportes e nas engenharias de uma forma geral. Assim, ela ocupa um lugar de destaque no mundo científico.

Em matemática, o estudo de problemas de minimização e maximização teve início a cerca de 25 séculos atrás. Por um longo período, não se descobriu uma maneira uniforme de resolver problemas de otimização. Os primeiros métodos de caráter geral foram desenvolvidos a cerca de 300 anos, na época do desenvolvimento da análise matemática. Desde então, ficou claro que certos problemas de otimização desempenham um papel fundamental em ciências naturais. Em 1744, o cientista francês Pierre Louis Moreau de Maupertius propôs o seu *princípio metafísico*: a natureza sempre opera com a maior economia possível (BOYER, 2003).

A otimização é um assunto muito amplo. Neste trabalho, vamos restringir a nossa atenção principalmente a alguns aspectos da programação não-linear (PNL).

Um problema de programação não-linear é um problema que consiste em encontrar pontos de máximo ou de mínimo de funções não lineares. Na maioria das aplicações, os fenômenos físicos ou econômicos são melhor representados por modelos não lineares. As funções de um problema de PNL podem ser diferenciáveis ou não. Trataremos do caso em que as funções são diferenciáveis.

Um caso particular do problema de PNL é o problema irrestrito (PI) que consiste em minimizar uma função sem restrições. O estudo dos métodos que resolvem problemas irrestritos é de fundamental importância em otimização, porque muitos métodos para resolver problemas de PNL fazem uso dos métodos que resolvem o PI. Em alguns casos é possível encontrar (quando existem) diretamente as soluções de PI. No entanto, quando o número de variáveis ou a complexidade das funções aumentam, torna-se complicado até mesmo encontrar pontos candidatos à solução. Assim, é necessário apelar para métodos numéricos, ou seja, algoritmos.

Um algoritmo, no contexto dos métodos numéricos, é um procedimento iterativo que, a partir de um ponto inicial  $x_0$ , define, conforme determinadas regras, um novo ponto  $x_1$ ; e a partir de  $x_1$ , um ponto  $x_2$ , seguindo as mesmas regras. Desta forma, cria-se uma sequência de pontos  $(x_k)$ . Os algoritmos diferenciam entre si justamente pelas regras que determinam o novo ponto. Normalmente, a sequência  $(x_k)$  gerada por um algoritmo nunca atinge exatamente uma solução, mas converge para um soluções do problema.

Neste trabalho, apresentaremos o algoritmo de pontos interiores e direções viáveis FDIPA, desenvolvido por José Herskovits [9] e faremos uma análise numérica da influência dos parâmetros definidos no algoritmo.

No Capítulo 2 apresentamos algumas definições básicas e alguns resultados de Análise e Álgebra Linear relevantes para este trabalho. As principais referências deste capítulo são [13], [14] e [15].

No Capítulo 3 desenvolvemos os conceitos básicos de otimização. Começamos com algumas situações que garantem a existência de um minimizador e em seguida discutimos as condições de otimalidade para problemas de minimização irrestrita e otimização com restrições. Algumas referências para este assunto são [1], [8], [11], [16], [17] e [18].

No Capítulo 4 descrevemos um modelo geral de algoritmo para minimizar uma função em  $\mathbb{R}^n$  e apresentamos alguns métodos específicos de minimização para otimização irrestrita. Aspectos de convergência global e velocidade de convergência dos métodos não são abordados neste trabalho. As referências para este capítulo são [3] e [12].

No Capítulo 5 descrevemos o algoritmo FDIPA. Esse algoritmo, tem por objetivo encontrar dentre os pontos de um conjunto definido por restrições de igualdade e/ou desigualdade, aqueles que minimizam uma função diferenciável. O FDIPA baseia-se na resolução de dois sistemas de equações lineares com a mesma matriz de coeficientes, obtidos das condições necessárias de primeira ordem de Karush-Kuhn-Tucker. A partir de um ponto inicial no interior do conjunto viável, o FDIPA gera uma sequência de pontos também interiores ao conjunto. Em cada iteração, uma nova direção de descida é obtida e, em seguida, produz-se uma deflexão da direção de descida no sentido do interior do conjunto viável, de modo a se obter uma nova direção que seja de descida e viável. Realiza-se então

uma busca linear para obter um novo ponto interior e garantir a convergência global do método. Uma família de algoritmos pode ser obtida variando-se as regras de atualização dos parâmetros do FDIPA. Este capítulo está baseado em [9].

No Capítulo 6 apresentamos um estudo da influência dos parâmetros do FDIPA na eficiência do algoritmo. O estudo apresentado neste capítulo foi feito considerando-se um único algoritmo e com restrições de desigualdade somente. Testes numéricos apontaram para uma escolha de parâmetros que levaram a um número menor de iterações na resolução dos problemas teste.

Por fim, o Capítulo 7 destina-se a analisar o estudo realizado e apresentar algumas possibilidades de trabalhos futuros.

## 2 CONCEITOS PRELIMINARES

Neste capítulo apresentaremos algumas definições básicas e alguns resultados de Análise e Álgebra Linear relevantes para este trabalho.

### 2.1 SEQUÊNCIAS

**Definição 2.1.1.** Uma **sequência** em  $\mathbb{R}^n$  é uma função  $x : \mathbb{N} \rightarrow \mathbb{R}^n$ , que associa a cada número natural  $n$  um ponto  $x_n \in \mathbb{R}^n$ . Denotaremos uma sequência por  $(x_k)$ .

**Definição 2.1.2.** Diz-se que um ponto  $a \in \mathbb{R}^n$  é o **limite** da sequência  $(x_k)$  quando, para todo  $\epsilon > 0$  dado arbitrariamente, é possível obter  $k_0 \in \mathbb{N}$  tal que  $k > k_0 \Rightarrow |x_k - a| < \epsilon$ .

**Definição 2.1.3.** Uma sequência  $(x_k)$  em  $\mathbb{R}^n$  diz-se **convergente** quando existe  $a = \lim x_k$ .

**Definição 2.1.4.** Uma **subsequência** de  $(x_k)$  é a restrição desta sequência a um subconjunto infinito  $\mathbb{N}' = \{k_1 < \dots < k_m < \dots\} \subset \mathbb{N}$ . Utilizaremos a notação  $(x_{k_m})_{k \in \mathbb{N}}$  para indicar uma subsequência.

**Teorema 2.1.1 (Bolzano-Weierstrass).** Toda sequência limitada em  $\mathbb{R}^n$  possui uma subsequência convergente.

### 2.2 NOÇÕES DE TOPOLOGIA

Seja  $n \in \mathbb{N}$ . O espaço euclidiano  $n$ -dimensional é o produto cartesiano de  $n$  fatores iguais a  $\mathbb{R}$ :  $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}$ . Seus elementos, portanto, são as listas de  $n$  termos reais  $x = (x_1, \dots, x_n)$ . Os elementos de  $\mathbb{R}^n$  são chamados *pontos* e às vezes *vetores*.

**Definição 2.2.1.** Uma **norma** em  $\mathbb{R}^n$  é uma função real  $\| \cdot \| : \mathbb{R}^n \rightarrow \mathbb{R}$  que satisfaz as seguintes propriedades:

1.  $\|x\| \geq 0$ , valendo  $\|x\| = 0$  somente quando  $x = 0$ ;
2.  $\|\alpha \cdot x\| = |\alpha| \cdot \|x\|$ ;
3.  $\|x + y\| \leq \|x\| + \|y\|$ .

Há uma infinidade de normas que podem ser definidas no espaço euclidiano  $\mathbb{R}^n$ . Dentre elas, temos:

- $\|x\| = \sqrt{x_1^2 + \dots + x_n^2}$  (Norma Euclidiana);
- $\|x\|_M = \max\{|x_1|, \dots, |x_n|\}$  (Norma do Máximo);

- $\|x\|_S = |x_1| + \cdots + |x_n|$  (Norma da Soma).

**Definição 2.2.2.** Dados o ponto  $a \in \mathbb{R}^n$  e o número real  $r > 0$ , a **bola aberta** de centro  $a$  e raio  $r$  é o conjunto  $B(a; r)$  dos pontos  $x \in \mathbb{R}^n$  cuja distância ao ponto  $a$  é menor que  $r$ .

**Definição 2.2.3.** Um ponto  $x \in \mathbb{R}^n$  é dito **ponto de fronteira** de um conjunto  $X \subset \mathbb{R}^n$  quando qualquer vizinhança de  $x$  contém algum elemento de  $X$  e algum elemento do complementar de  $X$ . O conjunto dos pontos fronteira de  $X$  é chamado de fronteira de  $X$ .

**Definição 2.2.4.** Um conjunto  $X$  é **fechado** quando contém sua fronteira.

**Definição 2.2.5.** Um conjunto  $X \subset \mathbb{R}^n$  é **limitado** quando quando existe  $k > 0$  tal que  $|x| \leq k$ , para todo  $x \in X$ .

**Definição 2.2.6.** Um conjunto  $X \subset \mathbb{R}^n$  chama-se **compacto** quando é limitado e fechado.

**Definição 2.2.7.** Um ponto  $x \in X \subset \mathbb{R}^n$  é chamado um **ponto interior** de  $X$  quando, para algum  $r > 0$ , tem-se  $B(a; r) \subset X$ .

## 2.3 FUNÇÕES DIFERENCIÁVEIS

**Definição 2.3.1.** Seja  $f : X \rightarrow \mathbb{R}^m$ , definida no conjunto  $x \subset \mathbb{R}^n$ . Dizemos que  $f$  é **contínua** no ponto  $a \in X$  quando, para cada  $\epsilon > 0$  dado, pode-se obter  $\delta > 0$  tal que se  $x \in X$  e  $\|x - a\| < \delta$ , então  $\|f(x) - f(a)\| < \epsilon$ .

Se  $f : X \rightarrow \mathbb{R}^m$  é contínua em todos os pontos do conjunto  $X$  dizemos que  $f$  é uma *aplicação contínua*.

**Observação 2.3.1.** A continuidade de  $f$  no ponto  $a$  independe das normas que se utilizem em  $\mathbb{R}^m$  e  $\mathbb{R}^n$ .

**Definição 2.3.2.** Seja  $f : U \rightarrow \mathbb{R}$  uma função definida no aberto  $U \subset \mathbb{R}^n$ . Para cada  $i = 1, \dots, n$ , a  $i$ -ésima **derivada parcial** de  $f$  no ponto  $a = (a_1, \dots, a_n)^T \in U$  é o número

$$\frac{\partial f}{\partial x_i}(a) = \lim_{t \rightarrow 0} \frac{f(a + te_i) - f(a)}{t}$$

caso este limite exista.

**Definição 2.3.3.** Seja  $f : U \rightarrow \mathbb{R}$  uma função que possui as  $n$  derivadas parciais em todos os pontos do aberto  $U \subset \mathbb{R}^n$ . Ficam então definidas  $n$  funções

$$\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} : U \rightarrow \mathbb{R}.$$

Se estas funções forem contínuas em  $U$ , diremos que  $f$  é uma **função de classe  $C^1$**  e escrevemos  $f \in C^1$ .



**Definição 2.3.4.** Uma função  $f : U \rightarrow \mathbb{R}$  que possui, em cada ponto de  $U$ , todas as derivadas parciais de ordem  $k$ , as quais são funções contínuas em  $U$ , chama-se uma **função de classe  $C^k$** . Escrevemos então  $f \in C^k$ .

**Definição 2.3.5.** Uma função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , diz-se **diferenciável** no ponto  $a \in \mathbb{R}^n$  quando existirem as derivadas parciais  $\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a)$  e, além disso, para todo vetor  $v = (v_1, \dots, v_n)^T \in \mathbb{R}^n$  tivermos

$$f(a+v) - f(a) = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \cdot \alpha_i + r(v), \text{ com } \lim_{|v| \rightarrow 0} \frac{r(v)}{|v|} = 0$$

**Teorema 2.3.1 (Fórmula de Taylor).** Seja  $f : U \rightarrow \mathbb{R}$  de classe  $C^2$  no aberto  $U \subset \mathbb{R}^n$ . Fixando  $a \in U$ , para todo  $v = (\alpha_1, \dots, \alpha_n)^T \in \mathbb{R}^n$  tal que  $a+v \in U$ , escrevemos

$$f(a+v) - f(a) = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \cdot \alpha_i + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} \cdot \alpha_i \alpha_j + r(v)$$

as derivadas sendo calculadas no ponto  $a$ . Então  $\lim_{|v| \rightarrow 0} \frac{r(v)}{|v|^2} = 0$ .

**Teorema 2.3.2 (Regra da Cadeia).** Sejam  $U \subset \mathbb{R}^m$ ,  $V \subset \mathbb{R}^n$  abertos e  $f : U \rightarrow \mathbb{R}^n$ ,  $g : V \rightarrow \mathbb{R}^p$  diferenciáveis no ponto  $a \in U$ ,  $b = f(a) \in V$ , com  $f(U) \subset V$ . Então

$$(g \circ f)'(a) = g'(b) \cdot f'(a) : \mathbb{R}^m \rightarrow \mathbb{R}^p.$$

Resumidamente: a derivada da aplicação composta é a composta das derivadas.

## 2.4 ÁLGEBRA LINEAR

**Definição 2.4.1.** Dizemos que um conjunto de vetores  $V = \{v_1, \dots, v_n\}$  é **linearmente independente (L.I.)** se, e somente, uma igualdade do tipo

$$\alpha_1 v_1 + \dots + \alpha_n v_n = 0$$

com os  $\alpha_i \in \mathbb{R}$ , só for possível para  $\alpha_1 = \dots = \alpha_n = 0$ .

**Definição 2.4.2.** Dizemos que um conjunto de vetores  $V = \{v_1, \dots, v_n\}$  é **linearmente dependente (L.D.)** se, e somente,  $V$  não é L.I..

**Definição 2.4.3.** Uma matriz  $A \in \mathbb{R}^{n \times n}$  diz-se **simétrica** se  $A^T = A$ .

**Definição 2.4.4.** Seja  $A \in \mathbb{R}^{n \times n}$  uma matriz simétrica e  $x \in \mathbb{R}_*^n$ . Dizemos que  $A$  é:

- (a) **Definida positiva** quando  $x^T A x > 0, \forall x \in \mathbb{R}_*^n$ .
- (b) **Semidefinida positiva** quando  $x^T A x \geq 0, \forall x \in \mathbb{R}_*^n$ .
- (c) **Definida negativa** quando  $x^T A x < 0, \forall x \in \mathbb{R}_*^n$ .
- (d) **Semidefinida negativa** quando  $x^T A x \leq 0, \forall x \in \mathbb{R}_*^n$ .
- (e) **Indefinida** se existem  $x$  e  $y$  em  $\mathbb{R}^n$  tais que  $x^T A x > 0$  e  $y^T A y < 0$ .

## 2.5 DEFINIÇÕES E RESULTADOS CLÁSSICOS

**Definição 2.5.1.** Considere uma função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  e  $\bar{x} \in \Omega \subset \mathbb{R}^n$ .

- (a) Dizemos que  $\bar{x}$  é um **mínimo local** de  $f$  em  $\Omega$  quando existe  $\epsilon > 0$ , tal que  $f(\bar{x}) \leq f(x)$  para todo  $x \in \Omega$  tal que  $\|x - \bar{x}\| \leq \epsilon$ .
- (b) Dizemos que  $\bar{x}$  é um **mínimo global** de  $f$  em  $\Omega$  quando  $f(\bar{x}) \leq f(x)$ , para todo  $x \in \Omega$ .
- (c) Dizemos que  $\bar{x}$  é um **máximo local** de  $f$  em  $\Omega$  quando existe  $\epsilon > 0$ , tal que  $f(\bar{x}) \geq f(x)$  para todo  $x \in \Omega$  tal que  $\|x - \bar{x}\| \leq \epsilon$ .
- (d) Dizemos que  $\bar{x}$  é um **máximo global** de  $f$  em  $\Omega$  quando  $f(\bar{x}) \geq f(x)$ , para todo  $x \in \Omega$ .
- (e) Dizemos que  $\bar{x}$  é um **extremo global** de  $f$  em  $\Omega$  se  $\bar{x}$  é um ponto de máximo global ou um ponto de mínimo global de  $f$  em  $\Omega$ .
- (f) Dizemos que  $\bar{x}$  é um **extremo local** de  $f$  em  $\Omega$  se  $\bar{x}$  é um ponto de máximo local ou um ponto de mínimo local de  $f$  em  $\Omega$ .

**Teorema 2.5.1 (Weirstrass).** Toda função  $f : K \subset \mathbb{R}^n \rightarrow \mathbb{R}$  contínua definida no conjunto compacto  $K$  (não-vazio) possui pelo menos um máximo global e pelo menos um mínimo global em  $K$ .

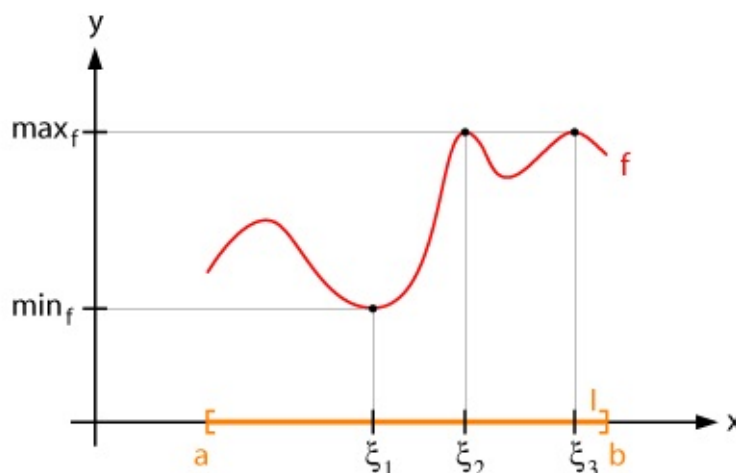


Figura 1 – Ilustração do Teorema de Weirstrass

O gráfico acima representa uma função  $f : [a, b] \rightarrow \mathbb{R}$ . Geometricamente podemos observar que  $f$  é contínua (não apresenta "saltos") e está definida em um conjunto compacto (limitado e fechado). Logo, pelo Teorema de Weirstrass  $f$  possui pelo menos um mínimo

global em  $[a, b]$  (que ocorre em  $\xi_1$ ) e pelo menos um máximo global em  $[a, b]$  (neste caso, dois máximos globais que ocorrem em  $\xi_2$  e  $\xi_3$ ).

### 3 CONDIÇÕES DE OTIMALIDADE

Quando falamos de um problema de otimização, nos referimos a um problema que consiste em encontrar pontos de máximo ou de mínimo de uma função real sobre um conjunto  $\Omega \subset \mathbb{R}^n$ .

Neste trabalho nos ocuparemos de problemas da forma

$$(P) \begin{cases} \min f(x) \\ x \in \Omega, \end{cases} \quad (3.1)$$

no qual assumiremos que a função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  é continuamente diferenciável.

O conjunto  $\Omega$  será chamado *conjunto viável* do problema, os pontos de  $\Omega$  serão chamados *pontos viáveis* e  $f$  será chamada *função objetivo*.

Em geral, o conjunto  $\Omega$  é definido por restrições de igualdades e/ou desigualdades, ou seja

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

onde  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  e  $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$  são funções continuamente diferenciáveis. O problema de otimização pode então ser reescrito como

$$(P) \begin{cases} \min f(x) \\ \text{s. a. } h(x) = 0, \\ g(x) \leq 0. \end{cases} \quad (3.2)$$

Quando  $\Omega = \mathbb{R}^n$ , dizemos que o problema  $(P)$  é *irrestrito* e quando  $\Omega \neq \mathbb{R}^n$  falamos de *otimização com restrições*.

#### 3.1 CONDIÇÕES DE OTIMALIDADE PARA PROBLEMAS IRRESTRITOS

Um problema irrestrito é um problema da forma

$$(P) \begin{cases} \min f(x) \\ x \in \mathbb{R}^n. \end{cases} \quad (3.3)$$

Ao considerar tal problema, desejamos determinar, se possível, se um dado ponto  $\bar{x} \in \mathbb{R}^n$  é ou não um mínimo local ou global da função  $f$ . Para isso, precisamos caracterizar um minimizador de um problema irrestrito. A esta caracterização denominamos *condições de otimalidade*, as quais são estabelecidas nos teoremas a seguir.

**Teorema 3.1.1 (Condições Necessárias de Primeira Ordem).** *Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^1$ . Se  $\bar{x}$  é um minimizador local de  $f$  em  $\mathbb{R}^n$ , então  $\nabla f(\bar{x}) = 0$ .*

**Definição 3.1.1.** Um ponto  $\bar{x}$  que cumpre o Teorema (3.1.1) é dito ponto crítico ou estacionário da função  $f$ .

O Teorema (3.1.1) afirma que todo extremo local de uma função  $f$  necessariamente é um ponto crítico de  $f$ . Contudo, a recíproca desse Teorema é falsa, isto é, nem todo ponto crítico  $\bar{x} \in \mathbb{R}^n$  de uma função  $f$  é um extremo local dessa função. Para um contra-exemplo, considere a função  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , tal que

$$f(x, y) = x^2 - y^2.$$

O ponto  $\bar{x} = (0, 0)$  é um ponto crítico de  $f$ , pois  $\nabla f(0, 0) = (0, 0)$ . Mas não é um extremo local de  $f$  em  $\mathbb{R}^2$ , pois para qualquer bola aberta  $B(0, r)$  de centro em  $(0, 0)$  e raio  $r > 0$ , a função  $f$  assume valores maiores e menores do que  $0 = f(0, 0)$ .

Assim, podemos concluir que o Teorema (3.1.1) funciona como uma espécie de filtro, nos dizendo que entre todos os pontos no interior do conjunto admissível, os candidatos a extremo local da função são os pontos que anulam o gradiente de  $f$ .

**Teorema 3.1.2 (Condições Necessárias de Segunda Ordem).** Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^2$ . Se  $\bar{x}$  é um minimizador local de  $f$  em  $\mathbb{R}^n$ , então a matriz Hessiana de  $f$  no ponto  $\bar{x}$  é semidefinida positiva, isto é

$$d^t \nabla^2 f(\bar{x}) d \geq 0, \quad (3.4)$$

para todo  $d \in \mathbb{R}^n$ .

As condições apresentadas nos teoremas (3.1.1) e (3.1.2) devem ser satisfeitas para toda solução ótima local. No entanto, não são suficientes para que um ponto  $\bar{x}$  seja um minimizador. Por exemplo, a função  $f : \mathbb{R} \rightarrow \mathbb{R}$ , dada por  $f(x) = x^3$ , satisfaz essas condições no ponto  $\bar{x} = 0$  mas este ponto não é minimizador local do problema (3.3). Logo, precisamos de uma condição adicional para garantir que um ponto satisfazendo as condições necessárias de primeira e segunda ordem seja um minimizador local do problema. O teorema a seguir nos dá esta condição.

**Teorema 3.1.3 (Condições Suficientes de Segunda Ordem).** Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^2$ . Se  $\bar{x} \in \mathbb{R}^n$ ,  $\nabla f(\bar{x}) = 0$  e  $\nabla^2 f(\bar{x}) > 0$ , então  $\bar{x}$  é um minimizador local estrito de  $f$  em  $\mathbb{R}^n$ .

**Exemplo 3.1.1.** Descrever os pontos estacionários da função  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  dada por

$$f(x, y) = x^3 + y^3 - 3xy.$$

Para definir os pontos estacionários, resolvemos o seguinte sistema de equações, obtido a partir das condições necessárias de primeira ordem:

$$f'_x = 3x^2 - 3y = 0, \quad f'_y = 3y^2 - 3x = 0$$

As soluções são  $x_1 = (0, 0)$  e  $x_2 = (1, 1)$ .

Para encontrar os minimizadores locais, usamos as condições necessárias de segunda ordem. Assim, temos

$$f''(x_1) = \begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix}, \quad f''(x_2) = \begin{pmatrix} 6 & -3 \\ -3 & 6 \end{pmatrix}.$$

A matriz  $f''(x_1)$  é indefinida. Logo, pelo Teorema(3.1.2),  $x_1$  não é minimizador. Como a matriz  $f''(x_2)$  é definida positiva, pelas condições suficientes de segunda ordem,  $x_2$  é (o único) minimizador local (estrito).

As definições e os resultados envolvendo minimizadores, apresentados nesta seção, podem ser reformulados para maximizadores de forma inteiramente análoga.

### 3.2 CONDIÇÕES DE OTIMALIDADE PARA PROBLEMAS COM RESTRIÇÕES DE IGUALDADE

Consideremos o seguinte problema:

$$(P) \begin{cases} \max 5 - (x_1 - 2)^2 - 2(x_2 - 1)^2 \\ \text{s.a. } x_1 + 4x_2 = 3 \end{cases}$$

Se ignorarmos a restrição, temos a solução  $x_1 = 2$ ,  $x_2 = 1$ , que é muito grande para satisfazer a igualdade. Para que tenhamos a restrição satisfeita, vamos introduzir um fator  $\lambda$  e ajustá-lo de maneira que a restrição se torne tão grande quanto necessitamos. Para isso utilizaremos a função auxiliar

$$L(x_1, x_2, \lambda) = 5 - (x_1 - 2)^2 - 2(x_2 - 1)^2 + \lambda(3 - x_1 - 4x_2).$$

Essa função é chamada de *Lagrangiana* do problema. A idéia principal é ajustar  $\lambda$  de modo que possamos obter um valor que satisfaça a restrição.

- Se  $\lambda = 0$ ,  $\bar{x} = (2, 1)$ .
- Se  $\lambda = 1$ ,  $\bar{x} = \left(\frac{5}{3}, 0\right)$ , que está próximo da restrição.
- Se  $\lambda = \frac{2}{3}$ ,  $\bar{x} = \left(\frac{5}{3}, \frac{1}{3}\right)$ , que satisfaz a restrição.

A partir de agora, exploraremos esta idéia mais formalmente.

Dado o problema de programação não-linear (P), com restrições de igualdade:

$$(P) \begin{cases} \min f(x) \\ \text{s.a. } h(x) = 0, \end{cases} \quad (3.5)$$

onde  $f, h \in C^1$ ,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  e  $m < n$ . Como veremos a seguir, uma solução do problema (3.5) pode ser encontrada utilizando-se a *Lagrangiana*:

$$L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i \nabla h_i(x).$$

**Definição 3.2.1.** *O plano tangente a uma superfície  $S \subseteq \mathbb{R}^n$ , em um ponto  $\bar{x} \in \mathbb{R}^n$ , é o conjunto de vetores de  $\mathbb{R}^n$ , que são tangentes em  $\bar{x}$  a alguma curva diferenciável contida em  $S$  e que passa por  $\bar{x}$ .*

**Definição 3.2.2.** *Um ponto  $\bar{x}$  que satisfaz as equações  $h(x) = 0$  é regular em relação às restrições se, e somente se o conjunto de vetores  $\{\nabla h_1(\bar{x}), \dots, \nabla h_m(\bar{x})\}$  é linearmente independente.*

**Teorema 3.2.1.** *Se  $\bar{x}$  é um ponto regular da superfície  $S \equiv \{x \in \mathbb{R}^n | h(x) = 0\}$ , então o plano tangente  $T$  verifica*

$$T = \{y \in \mathbb{R}^n | Jh(\bar{x})y = 0\}. \quad (3.6)$$

**Teorema 3.2.2 (Condições Necessárias de Primeira Ordem).** *Seja  $\bar{x}$  um minimizador local de (3.5). Suponhamos que  $\bar{x}$  é um ponto regular das restrições. Então, existe  $\bar{\lambda} \in \mathbb{R}^m$  tal que*

$$\nabla f(\bar{x}) = \sum_{i=1}^m \bar{\lambda}_i \nabla h_i(\bar{x}) \quad (3.7)$$

A condição acima representa um sistema de equações em que o número de variáveis nesse sistema é igual ao número de equações. De forma equivalente, este sistema pode ser escrito como

$$\nabla L(\bar{x}, \bar{\lambda}) = 0,$$

onde a derivada é em relação a todas as variáveis da Lagrangiana. O vetor  $\bar{\lambda} \in \mathbb{R}^m$  é chamado o *multiplicador de Lagrange* associado às restrições. Este teorema é também conhecido como *Teorema dos Multiplicadores de Lagrange*, o qual caracteriza pontos estacionários de um problema de otimização.

No sentido geométrico, o Teorema dos Multiplicadores de Lagrange nos diz que o gradiente da função objetivo  $\nabla f(\bar{x})$  pode ser representado por uma combinação linear dos gradientes das restrições, isto é, uma combinação linear de  $\{\nabla h_i(\bar{x}), i = 1, \dots, m\}$ .

Vale ressaltar que a condição de regularidade das restrições é crucial neste Teorema: sem esta condição as afirmações do Teorema não podem ser garantidas.

**Exemplo 3.2.1.** *Seja  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  definida por  $f(x, y) = ax + by$ , com  $a^2 + b^2 \neq 0$ . O gradiente de  $f$  é, em todo ponto  $(x, y)$ , o vetor constante não-nulo  $v = (a, b)$ , ortogonal às linhas de nível  $ax + by = c$ , que são retas, duas a duas paralelas. A função  $f$  não tem pontos críticos. Mas se  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$  for dada por  $h(x, y) = x^2 + y^2 - 1$  então  $\nabla h(x, y) = (2x, 2y)$ ,*

1 é ponto regular de  $h$  e a imagem inversa de 1 pela função  $h$  é a circunferência unitária  $x^2 + y^2 = 1$ . Logo, temos que o conjunto admissível é compacto, e pelo Teorema de Weirstrass,  $f$  possui pelo menos dois pontos críticos nos quais assume seus valores mínimo e máximo. Os pontos críticos de  $f$  são as soluções do sistema

$$\nabla f(x, y) = \lambda \nabla h(x, y)$$

ou seja:

$$2\lambda x = a, 2\lambda y = b, x^2 + y^2 = 1$$

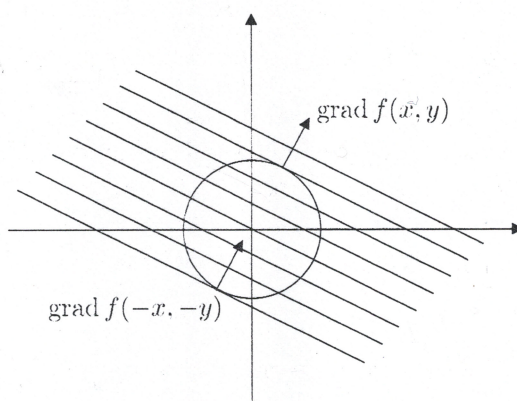


Figura 2 – Exemplo 3.2.1

Isto nos dá

$$(x, y) = \left( \frac{a}{\sqrt{a^2 + b^2}}, \frac{b}{\sqrt{a^2 + b^2}} \right) \text{ ou } (x, y) = \left( \frac{-a}{\sqrt{a^2 + b^2}}, \frac{-b}{\sqrt{a^2 + b^2}} \right)$$

Esses são os pontos em que  $f$  assume seus valores máximos e mínimos em  $S^1$ .

**Teorema 3.2.3 (Condições Necessárias de Segunda Ordem).** *Sejam  $\bar{x}$  um ponto regular, minimizador local de (3.5) e  $T$  como em (3.6). Suponhamos  $f, h \in C^2$ . Então existe  $\bar{\lambda} \in \mathbb{R}^n$  tal que*

$$\nabla f(\bar{x}) + \sum_{j=1}^m \bar{\lambda}_j \nabla h_j(\bar{x}) = 0 \tag{3.8}$$

e

$$y^t \nabla_x^2 L(\bar{x}, \bar{\lambda}) y \geq 0, \forall y \in T, \tag{3.9}$$

onde

$$L(x, \lambda) = f(x) + \lambda^t h(x), x \in \mathbb{R}^n, \lambda \in \mathbb{R}^m$$

é a chamada função lagrangeana.

As condições de segunda ordem expressam sempre informações sobre a curvatura das funções. No caso de restrições lineares, nas condições de segunda ordem aparecerá



somente a função objetivo. Se considerarmos restrições não-lineares, (3.9) significa que as curvaturas, tanto da função objetivo como das restrições, devem ser levadas em conta para caracterizar um minimizador local.

O teorema a seguir apresenta condições suficientes para que um ponto regular seja minimizador local estrito de  $f$  com restrições de igualdade.

**Teorema 3.2.4 (Condições Suficientes de Segunda Ordem).** *Sejam  $\bar{x}$  um ponto regular tal que  $h(\bar{x}) = 0$  e  $T$  como em (3.6). Se  $\bar{\lambda} \in \mathbb{R}^n$  é tal que*

$$\nabla f(\bar{x}) + \sum_{j=1}^m \bar{\lambda}_j \nabla h_j(\bar{x}) = 0 \quad (3.10)$$

e

$$y^t \nabla_x^2 L(\bar{x}, \bar{\lambda}) y > 0, \forall y \in T - \{0\},$$

então  $\bar{x}$  é um minimizador local estrito de (3.5).

**Exemplo 3.2.2.** *Uma refinaria de petróleo deve enviar barris de combustível para alguns tanques de armazenamento. O transporte será feito a partir de duas plataformas A e B. O custo do transporte de  $x$  unidades partindo de A é  $ax^2$ ; o custo do transporte de  $y$  unidades partindo de B é  $by^2$ , onde  $a > 0$  e  $b > 0$  são dados. Como podemos transportar  $q$  barris com o menor custo?*

*A solução para este problema consiste em resolver o seguinte problema de otimização:*

$$(P) \begin{cases} \min ax^2 + by^2 \\ \text{s.a. } x + y = q \end{cases}$$

*Calculando a Lagrangiana do problema temos:*

$$L(x, y, \lambda) = ax^2 + by^2 + \lambda(q - x - y).$$

*Aplicando as condições necessárias de primeira ordem:*

$$\frac{\partial L}{\partial x}(\bar{x}, \bar{y}, \bar{\lambda}) = 2a\bar{x} - \bar{\lambda} = 0$$

$$\frac{\partial L}{\partial y}(\bar{x}, \bar{y}, \bar{\lambda}) = 2b\bar{y} - \bar{\lambda} = 0$$

$$\frac{\partial L}{\partial \lambda}(\bar{x}, \bar{y}, \bar{\lambda}) = q - \bar{x} - \bar{y} = 0$$

*Das duas primeiras igualdades obtemos  $\bar{x} = \frac{b}{a}\bar{y}$ , que ao aplicarmos na terceira igualdade nos dá*

$$\bar{x} = \frac{bq}{a+b}, \quad \bar{y} = \frac{aq}{a+b}, \quad \bar{\lambda} = \frac{2abq}{a+b}$$

*e o custo de  $\frac{abq^2}{a+b}$ .*

A matriz Hessiana  $H(x, y) = \begin{pmatrix} 2a & 0 \\ 0 & 2b \end{pmatrix}$  é positiva definida, pois por hipótese,  $a > 0$  e  $b > 0$ . Logo, pelas condições suficientes de segunda ordem, uma vez que sejam dados  $a$ ,  $b$  e  $q$ , esta solução minimiza o custo do transporte dos barris de combustível.

### 3.3 CONDIÇÕES DE OTIMALIDADE PARA PROBLEMAS COM RESTRIÇÕES DE IGUALDADE E DESIGUALDADE

Nosso objetivo nesta seção é apresentar condições de otimalidade para o problema:

$$(P) \begin{cases} \min f(x) \\ \text{s.a. } h(x) = 0 \\ g(x) \leq 0 \end{cases} \quad (3.11)$$

onde  $f, h, g \in C^1$ ,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , com  $m < n$ , e  $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ .

**Definição 3.3.1.** *Seja  $\bar{x}$  um ponto satisfazendo as restrições*

$$h(\bar{x}) = 0, \quad g(\bar{x}) = 0 \quad (3.12)$$

e seja  $J$  o conjunto de índices  $j$  tais que  $g_j(\bar{x}) = 0$ . Então  $\bar{x}$  é um **ponto regular** se os vetores  $\nabla h_i(\bar{x})$ ,  $\nabla g_j(\bar{x})$ ,  $1 \leq i \leq m$ ,  $j \in J$  são linearmente independentes.

**Teorema 3.3.1 (Condições Necessárias de Primeira Ordem de Kuhn-Tucker).**

*Seja  $\bar{x} \in S$  um ponto regular. Se  $\bar{x}$  é um minimizador local de (3.11), então existem  $\lambda \in \mathbb{R}^m$  e  $\mu \in \mathbb{R}^p$  tais que*

$$\nabla f(\bar{x}) + \lambda^t \nabla h(\bar{x}) + \mu^t \nabla g(\bar{x}) = 0 \quad (3.13)$$

$$\mu^t g(\bar{x}) = 0 \quad (3.14)$$

$$\mu \geq 0 \quad (3.15)$$

**Definição 3.3.2.** *Um ponto viável  $\bar{x} \in \Omega$  é dito **estacionário** quando cumpre as condições necessárias do Teorema (3.3.1).*

**Teorema 3.3.2 (Condições Necessárias de Segunda Ordem).** *Suponhamos  $f, h, g \in C^2$ . Seja  $\bar{x}$  um minimizador local de (3.11). Suponhamos que  $\bar{x}$  é regular. Então, existem*

*$\lambda \in \mathbb{R}^m$ ,  $\mu \in \mathbb{R}^p$  tais que (3.13) e (3.15) se verificam e, além disso, a matriz  $H(\bar{x}, \lambda, \mu)$ , definida por*

$$H(\bar{x}, \lambda, \mu) \equiv \nabla^2 f(\bar{x}) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(\bar{x}) + \sum_{j=1}^p \mu_j \nabla^2 g_j(\bar{x}), \quad j \in J. \quad (3.16)$$

*é semidefinida positiva no espaço tangente  $\bar{T}$ , isto é,*

$$y^t H(\bar{x}, \lambda, \mu) y \geq 0, \quad \forall y \in \bar{T}$$

onde

$$\bar{T} \equiv \{y \in \mathbb{R}^n \mid \nabla h(\bar{x})y = 0, e \nabla g_j(\bar{x})y = 0, \forall j \in J\}$$

**Teorema 3.3.3 (Condições Suficientes de Segunda Ordem).** *Sejam  $f, h, g \in C^2$ . Seja  $\bar{x} \in S$  tal que existam  $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^P$  satisfazendo*

$$\mu \geq 0 \tag{3.17}$$

$$\mu^t \nabla g(\bar{x}) = 0 \tag{3.18}$$

$$\nabla f(\bar{x}) + \lambda^t \nabla h(\bar{x}) + \mu^t \nabla g(\bar{x}) = 0 \tag{3.19}$$

$$\tag{3.20}$$

e a matriz Hessiana

$$H(\bar{x}, \lambda, \mu) = \nabla^2 f(\bar{x}) + \lambda^t \nabla^2 h(\bar{x}) + \mu^t \nabla^2 g(\bar{x}). \tag{3.21}$$

é positiva definida no subespaço

$$T' = \{y \mid \nabla h(\bar{x})y = 0, \nabla g_j(\bar{x})y = 0, \forall j \in J\},$$

onde

$$J = \{j : g_j(\bar{x}) = 0, \mu_j > 0\}.$$

Então,  $\bar{x}$  é um minimizador local estrito de (3.11).

## 4 ALGORITMOS DE PROGRAMAÇÃO NÃO-LINEAR

Quando temos um problema de otimização, dificilmente conseguimos resolver, de forma direta, o sistema (normalmente não-linear) de  $n$  equações e  $n$  incógnitas dado por  $\nabla f(x) = 0$ . Em geral, o que fazemos é tentar encontrar a solução ótima para o problema por meio da implementação de um algoritmo iterativo em um software. Uma das possibilidades para a construção do algoritmo baseia-se na seguinte idéia de processo iterativo: considerando-se um ponto inicial  $x_0$ , obtemos um novo e melhor ponto  $x_1$  e repetimos o processo gerando uma sequência  $(x_k) \subset \mathbb{R}^n$  em que a função objetivo decresce.

Uma forma geral de construir um algoritmo consiste em escolher, a partir de cada ponto obtido, a direção para dar o próximo passo no sentido de decrescimento da função  $f$ .

**Definição 4.0.1.** Considere uma função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , um ponto  $\bar{x} \in \mathbb{R}^n$  e uma direção  $d \in \mathbb{R}^n \setminus \{0\}$ . Dizemos que  $d$  é uma **direção de descida** para  $f$ , a partir de  $\bar{x}$ , quando existe  $\delta > 0$  tal que  $f(\bar{x} + td) < f(\bar{x})$ , para todo  $t \in (0, \delta)$ .

Uma condição suficiente para uma direção ser de descida é apresentada no teorema abaixo.

**Teorema 4.0.1.** Se  $\nabla f(\bar{x})^t d < 0$ , então  $d$  é uma direção de descida para  $f$ , a partir de  $\bar{x}$ .

Quando  $n = 2$  ou  $n = 3$ , podemos interpretar geometricamente o Teorema 4.0.1, dizendo que as direções que formam um ângulo obtuso com  $\nabla f(\bar{x})$  são de descida.

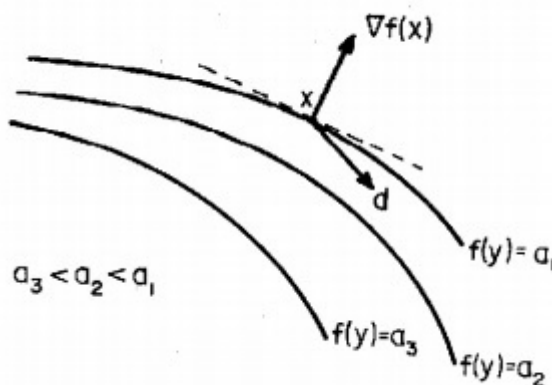


Figura 3 – Ilustração do Teorema 4.0.1

O teorema (4.0.1) nos diz que, dado  $d \in \mathbb{R}^n$  tal que  $\nabla f(\bar{x})^t d < 0$ , certamente podemos encontrar nessa direção pontos onde o valor da função seja estritamente menor que  $f(\bar{x})$ .

A existência dessas direções sugere um modelo geral de algoritmo para minimizar uma função sem restrições.

**Algoritmo 4.0.1 (Algoritmo Básico).** Dado:  $x_0 \in \mathbb{R}^n$ ,  $k = 0$ .

Enquanto  $\nabla f(x_k) \neq 0$

**Passo 1:** Calcule  $d_k \in \mathbb{R}^n$  tal que  $\nabla f(x_k)d_k < 0$ .

**Passo 2:** (Determinação do tamanho do passo)

Calcule  $t_k > 0$  tal que  $f(x_k + t_k d_k) < f(x_k)$ .

**Passo 3:**

Faça  $x_{k+1} = x_k + t_k d_k$ .

$k = k + 1$ .

O algoritmo 4.0.1 ou encontra um ponto estacionário em um número finito de iterações ou gera uma sequência ao longo da qual  $f$  decresce. Porém, é possível que o processo continue indefinidamente sem verificar a condição  $\nabla f(x_k) = 0$  para nenhum valor de  $k$ . Neste caso, mediante este algoritmo, está sendo gerada uma sequência infinita  $(x_k)$  de pontos em  $\mathbb{R}^n$ . Assim, a questão agora é saber se esta sequência tem algum ponto de acumulação e, em caso afirmativo, se este ponto é estacionário. O algoritmo 4.0.1 gera uma sequência de pontos  $(x_k)$  tal que a sequência de números reais associada  $(f(x_k))$  é monótona decrescente.

Tomando a função de uma variável  $f(x) = x^2$ , temos que o único minimizador desta função é  $\bar{x} = 0$ . A sequência definida por

$$x_k = 1 + \frac{1}{k}, \text{ para } k \geq 1$$

pode ser gerada pelo algoritmo porque

$$f(x_{k+1}) = 1 + \frac{1}{(k+1)^2} < \left(1 + \frac{1}{k}\right)^2 = f(x_k).$$

No entanto,

$$\lim_{k \rightarrow \infty} x_k = 1.$$

Logo, a sequência gerada pelo algoritmo pode não convergir para um ponto estacionário do problema.

Portanto, se quisermos garantir convergência, o método deve ser modificado de maneira que a escolha da direção  $d_k$  e do tamanho do passo  $t_k$ , no algoritmo 4.0.1, não seja arbitrária.

Para resolver este subproblema de encontrar o tamanho do passo  $t_k > 0$  tal que  $f(x_k + t_k d_k) < f(x_k)$ , existem vários métodos, denominados *métodos de busca unidirecional*,

os quais garantem que o decrescimento no valor da função é proporcional ao tamanho do passo na direção  $d_k$  a partir de  $x_k$ . A seguir, apresentamos dois métodos de busca unidirecional: a busca exata e a *busca de Armijo*.

#### 4.1 MÉTODOS DE BUSCA UNIDIRECIONAL

Dada uma função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , um ponto  $\bar{x} \in \mathbb{R}^n$  e uma direção de descida  $d \in \mathbb{R}^n$ , queremos encontrar  $\bar{t} > 0$  tal que

$$f(\bar{x} + \bar{t}d) < f(\bar{x}).$$

Na verdade, nosso problema agora se resume a balancear o tamanho do passo  $t$  com o decrescimento promovido em  $f$ . Veremos duas abordagens para este problema: a busca exata e a busca de Armijo.

##### 4.1.1 Busca Unidirecional Exata

Dada uma função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , um ponto  $\bar{x} \in \mathbb{R}^n$  e uma direção  $d \in \mathbb{R}^n$ , o problema da *busca unidirecional exata* ao longo de  $d$  a partir de  $\bar{x}$  é

$$(P) \begin{cases} \min f(\bar{x} + td) \\ \text{s.a. } t > 0. \end{cases} \quad (4.1)$$

Este problema é em geral difícil de resolver. Entretanto, dependendo do tipo de função considerada, existem métodos eficientes para resolvê-lo. Um exemplo, é o *Método da Seção Áurea* que funciona perfeitamente, dentro de uma determinada tolerância, quando a função  $f$  é unimodal.

Um estudo mais detalhado do Método da Seção Áurea pode ser encontrado em [1], [12] e [18]. As idéias discutidas nestas referências estão resumidas no algoritmo a seguir, conforme apresentado em Ribeiro e Karas (2015, p. 76).

**Algoritmo 4.1.1 (Seção Áurea).** *Dados:*  $\epsilon > 0$ ,  $\rho > 0$ ,  $\theta_1 = \frac{3-\sqrt{5}}{2}$ ,  $\theta_2 = 1 - \theta_1$

**Passo 1:** *Obtenção do intervalo*  $[a, b]$

$$a = 0, s = \rho \text{ e } b = 2\rho$$

*REPITA* enquanto  $\varphi(b) < \varphi(s)$

$$a = s, s = b \text{ e } b = 2b.$$

**Passo 2:** *Obtenção de*  $\bar{t} \in [a, b]$

$$u = a + \theta_1(b - a), v = a + \theta_2(b - a)$$

*REPITA* enquanto  $(b - a) > \epsilon$

$$SE \varphi(u) < \varphi(v)$$

$$b = v, v = u, u = a + \theta_1(b - a)$$

SENÃO

$$a = u, u = v, v = a + \theta_2(b - a)$$

$$\text{Defina } \bar{t} = \frac{u + v}{2}.$$

#### 4.1.2 Busca de Armijo

No *Método de Armijo* procuramos uma boa redução da função ao longo da direção  $d$ , sem tentar minimizá-la.

Considerando uma função diferenciável  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , um ponto  $\bar{x} \in \mathbb{R}^n$ , uma direção de descida  $d \in \mathbb{R}^n$  e  $\eta \in (0, 1)$ , basicamente, a regra de Armijo nos garante que existe  $\delta > 0$  tal que

$$f(\bar{x} + td) \leq f(\bar{x}) + \eta t \nabla f(\bar{x})^T d. \quad (4.2)$$

para todo  $t \in [0, \delta)$ .

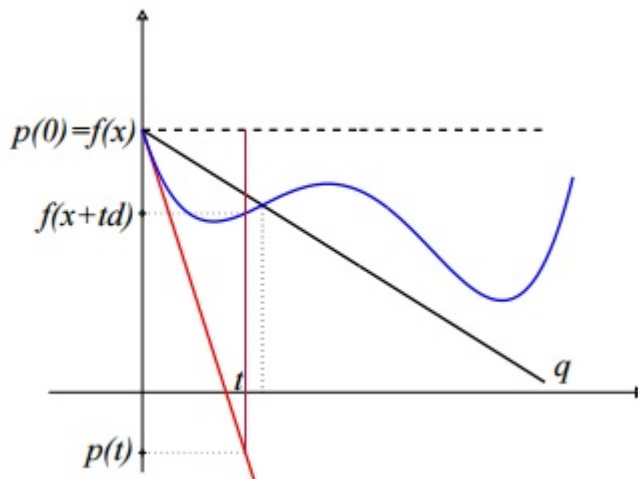


Figura 4 – Interpretação da condição de Armijo

A figura [4] nos dá uma interpretação geométrica para a busca de Armijo. Considerando a função  $\varphi : [0, \infty) \rightarrow \mathbb{R}$  dada por  $\varphi(t) = f(\bar{x} + td)$ , a aproximação de primeira ordem de  $\varphi$  em torno de  $t = 0$  é  $p(t) = \varphi(0) + t\varphi'(0) = f(\bar{x}) + t\nabla f(\bar{x})^T d$ . Nesta figura a reta  $q(t)$  é dada por  $q(t) = f(\bar{x}) + \eta t \nabla f(\bar{x})^T d$ . A condição de Armijo é satisfeita para os pontos tais que  $\varphi$  está abaixo de  $q$ .

Tanto do ponto de vista computacional quanto do ponto de vista teórico, é importante que o tamanho de passo  $t$ , satisfazendo 4.2, não seja muito pequeno. Uma maneira

de conseguirmos isto é iniciar com  $t = 1$  e, se necessário, reduzir  $t$  até que 4.2 seja satisfeita. Podemos ver isso de maneira sintetizada no algoritmo abaixo.

**Algoritmo 4.1.2 (Busca de Armijo).** *Sejam  $\bar{x} \in \mathbb{R}^n$ ,  $d \in \mathbb{R}^n$  (direção de descida),  $\gamma, \eta \in (0, 1)$ .*

**Passo 1:** *Defina  $t = 1$ .*

**Passo 2:** *Enquanto  $f(\bar{x} + td) \leq f(\bar{x}) + \eta t \nabla f(\bar{x})^T d$ , faça  $t = \gamma t$ .*

Embora não encontre um ponto próximo a um minimizador unidirecional, o método de Armijo é muito eficiente para algoritmos bem projetados, pois faz um número muito pequeno de cálculo de função, sendo portanto muito rápido. Para o algoritmo de que trata este trabalho, iremos utilizar a Busca de Armijo, uma vez que este também foi o modelo de busca utilizado em [9].

O sucesso dos algoritmos que resolvem um problema de otimização dependem da escolha adequada da direção  $d_k$  e do tamanho do passo  $t_k$ . A diferença essencial entre os métodos está na escolha da direção  $d_k$ , pois este exerce uma grande influência no processo iterativo  $x_{k+1} = x_k + t_k d_k$ . A seguir apresentaremos alguns métodos baseados nestas escolhas.

## 4.2 MÉTODO DO GRADIENTE

Um dos métodos mais antigos e mais conhecidos para resolver um problema de otimização irrestrita é o método clássico do gradiente, também chamado de *método de Cauchy*, devido ao fato de que foi *August Louis Cauchy* quem demonstrou em 1847, que o gradiente é a direção de máximo crescimento de uma função. Portanto, no algoritmo de Cauchy, a direção é definida em cada iteração como o oposto da função objetivo no ponto corrente, ou seja, este método corresponde a escolher  $d_k$  na direção de  $-\nabla f(x_k)$ . A sequência gerada pelo algoritmo tem a forma

$$x_{k+1} = x_k - t_k \nabla f(x_k),$$

onde  $t_k$  é o comprimento do passo ao longo da direção  $d_k$ , no ponto  $x_k$ .

No algoritmo apresentado a seguir, a determinação do tamanho do passo fica em aberto. Dentre as diversas formas de busca existentes, podemos utilizar a busca exata (algoritmo da seção áurea) ou inexata (busca de Armijo), ambas apresentadas anteriormente.

**Algoritmo 4.2.1 (Método do Gradiente).** *Dado:  $x_0 \in \mathbb{R}^n$  e  $k = 0$ .*

*Enquanto  $\nabla f(x_k) \neq 0$ , REPITA*

**Passo 1:** *Defina  $d_k = -\nabla f(x_k)$ .*



**Passo 2:** Obtenha  $t_k > 0$  tal que  $f(x_k + t_k d_k) < f(x_k)$ .

**Passo 3:** Faça  $x_{k+1} = x_k + t_k d_k$  e  $k = k + 1$ .

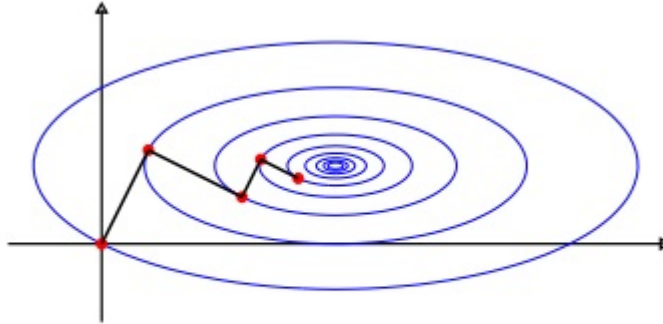


Figura 5 – Passos do Método de Máxima Descida

O método do gradiente com minimização unidirecional chama-se *Método de Máxima Descida*. Uma propriedade importante do método de máxima descida é que as direções utilizadas nas iterações subsequentes, isto é,  $d_k$  e  $d_{k+1}$ , são ortogonais. E por isso a trajetória do método é do tipo "zig-zag" (Figura [5]).

### 4.3 MÉTODO DE NEWTON

O método de Newton é uma das ferramentas mais importantes em otimização. Vimos no capítulo anterior que uma condição necessária para que um ponto  $\bar{x} \in \mathbb{R}^n$  seja solução de um problema irrestrito (P) é

$$\nabla f(\bar{x}) = 0. \quad (4.3)$$

Neste sentido, para encontrarmos os possíveis pontos candidatos à solução do problema (P), devemos resolver o sistema de  $n$  equações e  $n$  incógnitas  $\nabla f(x) = 0$ , geralmente não linear.

Sabemos que o método de Newton é um método muito eficiente para resolver sistema de equações. Assim, se a função  $f$  é de classe  $C^2$ , podemos aplicar o método de Newton para a função  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , onde  $F = \nabla f(x)$ ,  $f \in C^1$ . A sequência gerada pelo método de Newton para resolver o sistema de equações  $F(x) = 0$  é da forma

$$x_{k+1} = x_k + d_k,$$

onde  $d_k$  é definida como

$$d_k = -H(x_k)^{-1} F(x_k).$$

A matriz  $H(x_k)$  é a hessiana da função  $f$ , isto é,  $H(x_k) = \nabla^2 f(x_k)$ . Então, a sequência gerada pelo algoritmo pode ser reescrita em termos da função objetivo como

$$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k).$$

Com base nesta última relação podemos agora formalizar o método de Newton para minimizar a função  $f$ .

**Algoritmo 4.3.1 ( Método de Newton ).** *Dados:*  $x_0 \in \mathbb{R}^n$  e  $k = 0$ .

*Enquanto*  $\nabla f(x_k) \neq 0$ , *REPITA*

**Passo 1:** *Defina*  $x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$

**Passo 2:** *Determine o tamanho do passo*  $t_k > 0$ .

**Passo 3:** *Faça*  $x_{k+1} = x_k + t_k d_k$  e  $k = k + 1$ .

Cabe ressaltar que, diferente do que acontece no algoritmo do gradiente, o passo de Newton pode não estar bem definido, caso a matriz  $\nabla^2 f(x_k)$  seja singular. Além disso, mesmo que o passo  $d_k$  seja calculado, esta direção pode não ser de descida. Entretanto, se  $\nabla^2 f(x_k)$  é definida positiva, então a direção  $d_k$  está bem definida e é de descida.

#### 4.4 MÉTODOS QUASE-NEWTON

Enquanto no método do gradiente escolhemos  $d_k = -\nabla f(x_k)$  e, no método de Newton,  $d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$ , nos métodos *Quase-Newton* definimos a direção  $d_k$  a partir de aproximações para a Hessiana da função objetivo ao longo das iterações.

##### 4.4.1 Algoritmo Básico

Os Métodos quase-Newton são definidos considerando-se as direções de busca dadas por

$$d_k = -H_k \nabla f(x_k), \quad (4.4)$$

onde  $H_k \in \mathbb{R}^{n \times n}$  é simétrica e definida positiva. Assim, temos o seguinte algoritmo básico:

**Algoritmo 4.4.1 (Quase-Newton).** *Dados*  $x_0 \in \mathbb{R}^n$ ,  $H_0 \in \mathbb{R}^{n \times n}$  *definida positiva* e  $k = 0$

*Enquanto*  $\nabla f(x_k) \neq 0$

**Passo 1:** *Calcule*  $d_k = -H_k \nabla f(x_k)$ .

**Passo 2:** *Obtenha*  $t_k$  *através de uma busca linear* e *defina*  $x_{k+1} = x_k + t_k d_k$ .

**Passo 3:** *Determine*  $H_{k+1}$ , *simétrica e definida positiva* e *faça*  $k = k + 1$ .

Observe que se  $H_k = I$ , a direção de busca é a de Cauchy e se  $H_k = (\nabla^2 f(x_k))^{-1}$ , temos a direção de Newton.

Os Métodos que podem ser gerados através deste algoritmo básico dependem da forma de atualização da matriz  $H_k$  ao longo das iterações. Veremos agora, duas maneiras clássicas de atualizar a matriz  $H_k$  de modo que ao longo das iterações as matrizes obtidas se aproximem da inversa de  $\nabla^2 f(\bar{x})$ .

#### 4.4.2 O método DFP

De acordo com Izmailov e Solodov (2007), esta forma de obter a matriz  $H_{k+1}$  foi proposta por Davison, Fletcher e Powell. O método, referenciado como DFP, é caracterizado pela forma recursiva de atualizar  $H_k$  considerando uma correção simétrica de posto 2, isto é, a matriz  $H_{k+1}$  é obtida a partir de uma correção de  $H_k$  que consiste em somar duas matrizes simétricas de posto 1 da forma  $vv^t$ , onde  $v \in \mathbb{R}^n$ .

Assim, a forma para a nova atualização é dada por:

$$H_{k+1} = H_k + \frac{p_k (p_k)^T}{(p_k)^T q_k} - \frac{H_k (q_k)^T H_k}{(q_k)^T H_k q_k}, \quad (4.5)$$

onde  $p_k = x_{k+1} - x_k$  e  $q_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ .

#### 4.4.3 O método BFGS

Outro forma de atualizar as matrizes no algoritmo básico do método Quase-Newton é devido a Broyden, Fletcher, Goldfarb e Shanno (BFGS). Este método é muito popular devido ao seu bom desempenho numérico e consiste em encontrar uma aproximação para a matriz Hessiana, ao invés da sua inversa, utilizando a mesma correção simétrica de posto 2 que o DFP.

A fórmula para a atualização de  $H_k$  no método BFGS é:

$$H_{k+1} = H_k + \left( \frac{1 + q_k^T H_k q_k}{q_k^T p_k} \right) \frac{p_k p_k^T}{p_k^T q_k} - \frac{p_k q_k^T H_k + H_k q_k p_k^T}{q_k^T p_k} \quad (4.6)$$

Um estudo mais detalhado destes métodos pode ser encontrado em [1] e [12].

## 5 ALGORITMO DE PONTOS INTERIORES E DIREÇÕES VIÁVEIS PARA OTIMIZAÇÃO NÃO-LINEAR

Neste capítulo faremos uma abordagem sobre o algoritmo de pontos interiores e direções viáveis **FDIPA**, criado por José Herskovits em [9].

O algoritmo foi desenvolvido procurando-se solucionar o seguinte problema de otimização não-linear:

$$(P) \begin{cases} \min f(x) \\ \text{s. a. } h(x) = 0, \\ g(x) \leq 0. \end{cases} \quad (5.1)$$

onde assume-se que  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  e  $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$  são funções diferenciáveis não necessariamente convexas.

Neste trabalho apresentaremos duas versões do algoritmo: uma em que o problema (P) contém apenas restrições de desigualdade e na outra temos um algoritmo mais generalizado que resolve qualquer um dos possíveis casos de (P).

Este algoritmo é simples para codificar, robusto e eficiente. Não envolve funções penalidade, barreiras ou subproblemas de programação quadrática. Ele apenas requer em cada iteração, a solução de dois sistemas lineares com as mesmas matrizes seguida de uma busca linear inexata.

Muitos problemas de engenharia podem ser resolvidos utilizando este método. Ele pode ser aplicado a problemas de otimização estrutural, mecânica de fluidos, aerodinâmica e eletromagnetismo, entre outros. Podemos encontrar algumas destas aplicações em [2], [4] e [6].

### 5.1 ALGORITMO PARA O PROBLEMA COM RESTRIÇÕES DE DESIGUALDADE

Nesta seção, apresentaremos a versão do algoritmo aplicado ao seguinte problema de otimização não-linear diferenciável:

$$(P) \begin{cases} \min f(x) \\ \text{s.a. } g_i(x) \leq 0, \quad i = 1, 2, \dots, m. \end{cases} \quad (5.2)$$

onde  $f, g_1, g_2, \dots, g_m$  são funções continuamente diferenciáveis.

Introduzindo a variável auxiliar  $\lambda \in \mathbb{R}^m$  (Multiplicador de Lagrange), a função Lagrangeana associada ao problema (P) é

$$L(x, \lambda) = f(x) + \lambda^T g(x)$$

cuja matriz Hessiana é dada por

$$H(x, \lambda) = \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 g_i(x).$$

As condições de otimalidade de primeira ordem de Karush-Kuhn-Tucker (KKT) correspondentes ao problema 5.2 são expressas como:

$$\nabla f(x) + \nabla g(x)\lambda = 0 \quad (5.3)$$

$$G(x)\lambda = 0 \quad (5.4)$$

$$g(x) \leq 0 \quad (5.5)$$

$$\lambda \geq 0 \quad (5.6)$$

onde  $G(x) \in \mathbb{R}^{m \times m}$  é uma matriz diagonal tal que  $G_{ii}(x) = g_i(x)$  e  $g(x) = (g_1(x), g_2(x), \dots, g_m(x))$ .

Consideremos o seguinte sistema de equações lineares extraído das condições de K.K.T. do problem 5.2:

$$\begin{cases} \nabla f(x) + \nabla g(x)\lambda = 0 \\ G(x)\lambda = 0 \end{cases} \quad (5.7)$$

Tomando

$$\Phi(x, \lambda) = \begin{pmatrix} \nabla f(x) + \nabla g(x)\lambda \\ G(x)\lambda \end{pmatrix},$$

obtemos:

$$J\Phi(x, \lambda) = \begin{pmatrix} H(x, \lambda) & \nabla g(x) \\ \Lambda \nabla g^T(x) & G(x) \end{pmatrix}$$

onde  $\Lambda$  é uma matriz diagonal tal que  $\Lambda_{ii} = \lambda_i$ .

Utilizando uma iteração de Newton para resolver o sistema de equações lineares  $\Phi(x, \lambda) = 0$ , encontramos a partir do ponto  $(x, \lambda)$  um novo ponto  $(x_\alpha, \lambda_\alpha)$ , solução do sistema linear:

$$J\Phi(x, \lambda) [(x_\alpha, \lambda_\alpha) - (x, \lambda)] = -\Phi(x, \lambda).$$

Este sistema pode ser reescrito como

$$\begin{pmatrix} B & \nabla g(x) \\ \Lambda \nabla g^T(x) & G(x) \end{pmatrix} \begin{pmatrix} x_\alpha - x \\ \lambda_\alpha - \lambda \end{pmatrix} = - \begin{pmatrix} \nabla f(x) + \nabla g(x)\lambda \\ G(x)\lambda \end{pmatrix} \quad (5.8)$$

Aqui,  $B$  é uma matriz simétrica de ordem  $n$  que pode ser a própria hessiana ou uma aproximação dela obtida por alguma técnica quase-Newton. No entanto,  $B$  deve ser sempre definida positiva para que se garanta a convergência global do algoritmo.

Definindo  $d_\alpha = x_\alpha - x$  e substituindo no sistema linear (5.8) obtemos

$$\begin{cases} Bd_\alpha + \nabla g(x)\lambda_\alpha = -\nabla f(x) \\ \Lambda \nabla g^T(x)d_\alpha + G(x)\lambda_\alpha = 0. \end{cases} \quad (5.9)$$

$$(5.10)$$

A solução deste sistema nos fornece uma direção  $d_\alpha$  no espaço primal e uma nova estimativa  $\lambda_\alpha$  para  $\lambda$ .

Herskovits provou que  $d_\alpha$  é uma direção de descida para  $f$ . No entanto,  $d_\alpha$  pode não ser uma direção viável pois, desenvolvendo (5.10) temos:

$$\lambda_i \nabla g_i^T(x) d_\alpha + g_i(x) \lambda_{\alpha i} = 0, i = 1, 2, \dots, m.$$

Então, se  $g_i(x) \rightarrow 0$ ,

$$\nabla g_i(x) d_{\alpha i} \rightarrow 0,$$

ou seja, quando alguma restrição se aproxima de zero,  $d_\alpha$  tende a uma direção tangente ao conjunto viável.

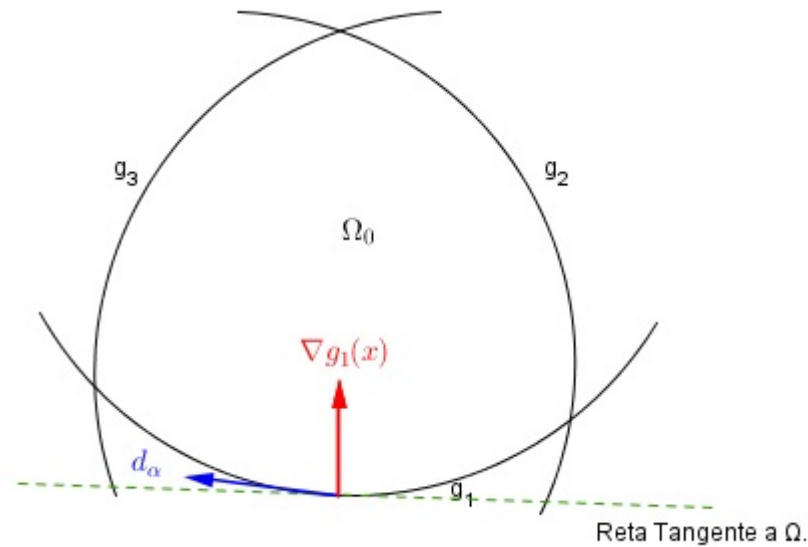


Figura 6 – Exemplo do que acontece com  $d_\alpha$  quando  $g_i(x) \rightarrow 0$ .

Para evitar este problema, define-se o seguinte sistema linear em  $d$  e  $\bar{\lambda}$ ,

$$\begin{cases} Bd + \nabla g(x) \bar{\lambda} = -\nabla f(x) & (5.11) \\ \Lambda \nabla g^T(x) d + G(x) \bar{\lambda} = -\rho \Lambda \omega & (5.12) \end{cases}$$

obtido adicionando-se um vetor negativo ao lado direito da equação (5.10). Os fatores  $\rho, \omega$  pertencem a  $\mathbb{R}_{*,+}^n$  e  $\bar{\lambda}$  é uma nova estimativa de  $\lambda$ . Neste caso, (5.12) é equivalente a

$$\lambda_i \nabla g_i^T(x) d + g_i(x) \bar{\lambda}_i = -\rho \lambda_i \omega_i, i = 1, 2, \dots, m$$

e, conseqüentemente

$$\nabla g_i^T(x) d = -\rho \omega_i < 0$$

para as restrições ativas. Logo,  $d$  é uma direção viável.

Quando o termo  $-\rho\Lambda\omega$  é acrescentado em (5.10) produz-se uma deflexão de  $d_\alpha$  para o interior do conjunto viável. Esta deflexão de  $d_\alpha$  relativa a  $i$ -ésima iteração é proporcional a  $\rho\omega_i$ .

O problema que surge agora é que, ao contrário de  $d_\alpha$ ,  $d$  pode não ser uma direção de descida para  $f$ . Entretanto, como a deflexão de  $d_\alpha$  é proporcional a  $\rho$ , é possível garantir que  $d$  também é uma direção de descida se  $\rho$  for escolhido convenientemente.

Definindo

$$d = d_\alpha + \rho d_\beta$$

e

$$\bar{\lambda} = \lambda_\alpha + \rho\lambda_\beta$$

e substituindo no sistema de equações formado por (5.11) e (5.12), encontramos

$$\begin{cases} Bd_\alpha + B\rho d_\beta + \nabla g(x)\lambda_\alpha + \nabla g(x)\rho\lambda_\beta = -\nabla f(x) \\ \Lambda\nabla g^T(x)d_\alpha + \Lambda\nabla g^T(x)\rho d_\beta + G(x)\lambda_\alpha + G(x)\rho\lambda_\beta = -\rho\Lambda\omega. \end{cases}$$

Uma vez que  $\{d_\alpha, d_\beta\}$  é L.I., podemos dividir este sistema em dois sistemas, um em função de  $d_\alpha$  e  $\lambda_\alpha$ , e o outro em função de  $d_\beta$  e  $\lambda_\beta$ . Assim, obtemos

$$\begin{cases} Bd_\alpha + \nabla g(x)\lambda_\alpha = -\nabla f(x) \\ \Lambda\nabla g^T(x)d_\alpha + G(x)\lambda_\alpha = 0 \end{cases} \quad (5.13)$$

e

$$\begin{cases} Bd_\beta + \nabla g(x)\lambda_\beta = 0 \\ \Lambda\nabla g^T(x)d_\beta + G(x)\rho\lambda_\beta = -\Lambda\omega. \end{cases} \quad (5.14)$$

Resolvendo estes sistemas obtemos a direção  $d$  e  $\bar{\lambda}$ . Desta forma, é fácil estabelecer limites para  $\rho$ .

Uma vez que definimos  $d = d_\alpha + \rho d_\beta$ , temos:

$$d^T \nabla f(x) = d_\alpha^T \nabla f(x) + \rho d_\beta^T \nabla f(x). \quad (5.15)$$

Como já vimos anteriormente,  $d_\alpha$  é uma direção de descida, ou seja,

$$d_\alpha^T \nabla f(x) < 0.$$

Logo, se  $d_\beta^T \nabla f(x) \leq 0$  então  $d^T \nabla f(x) < 0, \forall \rho \in \mathbb{R}_+$ .

Portanto,  $\rho$  deve ser convenientemente escolhido no caso em que  $d_\beta^T \nabla f(x) > 0$ .

Para isto, foi estabelecida a seguinte condição:

$$d^T \nabla f(x) \leq \xi (d_\alpha)^T \nabla f(x), \quad \xi \in (0, 1).$$

Substituindo (5.15) na equação acima obtemos

$$(d_\alpha)^T \nabla f(x) + \rho (d_\beta)^T \nabla f(x) \leq \xi (d_\beta)^T \nabla f(x)$$

ou equivalentemente,

$$\rho \leq (1 - \xi) \frac{(d_\alpha)^T \nabla f(x)}{(d_\beta)^T \nabla f(x)} \quad (5.16)$$

Em [9] provou-se que se  $\rho$  for escolhido de maneira que satisfaça a desigualdade (5.16), então  $d$  será uma direção de descida para  $f$ .

O teste de parada do algoritmo ocorre a cada iteração após o cálculo de  $d_\alpha$  no sistema (5.13).

Vamos agora, apresentar um algoritmo básico para o problema envolvendo desigualdades restritas. Básico no sentido de que vários procedimentos do algoritmo são definidos de forma ampla, sendo possível obter diferentes versões do algoritmo quando esses procedimentos são modificados.

Vale lembrar que no caso em que apenas desigualdades são consideradas, o *conjunto viável* é dado por  $\Omega = \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$ .

### FDIPA

*Parâmetros:*  $\xi \in (0, 1)$ ,  $\eta \in (0, 1)$ ,  $\varphi > 0$  e  $\nu \in (0, 1)$ .

*Dados Iniciais:*  $x \in \Omega_\alpha^0$ ,  $\lambda \in \mathbb{R}_+^m$ ,  $B \in \mathbb{R}^{n \times n}$  simétrica e positiva definida e  $\omega \in \mathbb{R}_+^m$ .

**Passo 1:** Cálculo da direção  $d$  e de  $\bar{\lambda}$ .

**I.** Calcule  $(d_\alpha, \lambda_\alpha)$  resolvendo o sistema linear:

$$\begin{cases} Bd_\alpha + \nabla g(x)\lambda_\alpha = -\nabla f(x) \\ \Lambda \nabla g^T(x)d_\alpha + G(x)\lambda_\alpha = 0 \end{cases} \quad (5.17)$$

Se  $d_\alpha = 0$ , pare.

**II.** Calcule  $(d_\beta, \lambda_\beta)$  resolvendo o sistema linear:

$$\begin{cases} Bd_\beta + \nabla g(x)\lambda_\beta = 0 \\ \Lambda \nabla g^T(x)d_\beta + G(x)\lambda_\beta = -\Lambda\omega \end{cases} \quad (5.18)$$

**III.** Se  $d_\beta^T \nabla f(x) > 0$ , tome

$$\rho = \min \left[ \varphi \|d_\alpha\|_2^2; (\xi - 1) \frac{d_\alpha^T \nabla f(x)}{d_\beta^T \nabla f(x)} \right]. \quad (5.19)$$

Caso contrário, tome

$$\rho = \varphi \|d_\alpha\|_2^2 \quad (5.20)$$



IV. Calcule a direção

$$d = d_\alpha + \rho d_\beta$$

e também

$$\bar{\lambda} = \lambda_\alpha + \rho \lambda_\beta.$$

**Passo 2:** Busca linear.

Calcule  $t$  o primeiro número da sequência  $\{1, \nu, \nu^2, \nu^3, \dots\}$  satisfazendo

$$f(x + td) \leq f(x) + t\eta \nabla f^T(x)d$$

e

$$\begin{aligned} g_i(x + td) &< 0, \text{ se } \bar{\lambda}_i \geq 0, \\ g_i(x + td) &\leq g_i(x), \text{ se } \bar{\lambda}_i < 0. \end{aligned} \tag{5.21}$$

**Passo 3:** Atualizações.

I. Defina

$$x = x + td$$

e defina novos valores para  $\omega$ ,  $\lambda \in \mathbb{R}_+^m$  e  $B$  simétrica e definida positiva.

II. Volte ao passo 1.

O algoritmo inclui no *Passo 2* uma busca linear inexata, baseada no procedimento de Armijo, estudada no capítulo 4. Esta busca garante um decréscimo razoável da função objetivo, mas são as condições (5.21) que fazem com que o novo ponto seja interior.

Diferentes algoritmos podem ser obtidos dependendo das formas de atualização de  $\lambda$ ,  $B$  e  $\omega$ . A seguir, apresentamos algumas considerações importantes sobre  $\lambda$ ,  $B$  e  $\omega$ .

**Proposição 5.1.1.** *Existem números positivos  $\lambda^I$ ,  $\lambda^S$  e  $\bar{g}$  tais que  $0 < \lambda_i \leq \lambda^S$ ,  $i = 1, 2, \dots, m$  e  $\lambda_i \geq \lambda^I$  para todo  $I$  tal que  $g_i(x) \geq -\bar{g}$ .*

**Proposição 5.1.2.** *Existem números positivos  $\sigma_1$  e  $\sigma_2$  tais que  $\sigma_1 \|d\|_2^2 \leq d^T B d \leq \sigma_2 \|d\|_2^2$ ,  $\forall d \in \mathbb{R}^n$ .*

**Proposição 5.1.3.** *Existem números positivos  $\omega_1$  e  $\omega_2$  tais que  $\omega_1 \leq \omega \leq \omega_2$ .*

Herskovits provou que toda sequência  $(x_k)$  gerada pelo algoritmo converge para um ponto K.K.T do problema, para quaisquer regras de atualização de  $\lambda$ ,  $B$  e  $\omega$  que verifiquem as proposições anteriores.

A análise teórica da convergência global do algoritmo, apresentada em [9], inclui primeiramente uma prova de que o algoritmo nunca falha se a solução dos sistemas lineares (5.17) e (5.18) são únicas. Depois um estudo do caso em que o algoritmo para após um

número finito de iterações. E finalmente é apresentada a prova da convergência global do algoritmo, ou seja, de que todo ponto de acumulação de uma sequência  $(x_k)$  gerada pelo algoritmo é um ponto K.K.T. do problema.

## 5.2 INCLUSÃO DE RESTRIÇÕES DE IGUALDADE

Apresentaremos agora, uma versão do algoritmo para o problema geral de programação não-linear (5.1). A técnica proposta estende o domínio de aplicação do algoritmo anterior a problemas que envolvam também restrições de igualdade. Este algoritmo requer um ponto inicial no interior do conjunto

$$\Delta \equiv \{x \in \Omega \mid h(x) \leq 0\}$$

o qual chamaremos  $\Delta^0$ , onde  $\Omega$  foi definido no capítulo 3.

O algoritmo gera uma sequência de pontos interiores  $(x_k) \in \Delta^0$ , que converge para um ponto Karush-Kuhn-Tucker  $\bar{x}$  do problema. As igualdades estão ativas somente no limite. Então, para que elas sejam satisfeitas, pode ser necessário um acréscimo da função objetivo.

O desenvolvimento deste algoritmo segue a mesma linha de idéias do anterior, com um diferencial de que agora, utilizamos uma função auxiliar

$$\phi_c(x) = f(x) + \sum_{i=1}^p c_i |h_i(x)|, \quad (5.22)$$

onde  $c_i$  são constantes positivas. A função  $\phi_c(x)$  é utilizada como uma função penalidade para as restrições de igualdade, de maneira que existe um  $c > 0$  finito tal que o mínimo de  $\phi_c$  sujeito a estas restrições ocorrem exatamente na solução (5.1). Desta forma, o uso de  $\phi_c$  como uma função penalidade é muito vantajoso, uma vez que, não temos que penalizar parâmetros tendendo a infinito. Por outro lado,  $\phi_c$  não possui derivadas nos pontos onde as restrições de igualdades estão ativas. Assim, necessitaremos de técnicas de otimização não diferenciável para minimizar  $\phi_c$ .

O algoritmo apresentado a seguir utiliza  $\phi_c$  na busca linear, mas evita os pontos onde esta função não é diferenciável. Um descrição mais detalhada deste algoritmo pode ser encontrada em [9].

### FDIPA

**Parâmetros:**  $\xi \in (0, 1)$ ,  $\eta \in (0, 1)$ ,  $\varphi > 0$  e  $\eta \in (0, )$ .

**Dados:**  $x \in \Delta^0$ ,  $0 < \lambda \in \mathbb{R}^m$ ,  $B \in \mathbb{R}^{n \times n}$  simétrica e definida positiva,  $0 < \omega^i \in \mathbb{R}^m$ ,  $0 < \omega^e \in \mathbb{R}^p$  e  $0 < c \in \mathbb{R}^p$ .

**Passo 1:** Cálculo de uma direção de descida.

I. Calcule  $(d_\alpha, \lambda_\alpha, \mu_\alpha)$  resolvendo o sistema linear:

$$\begin{cases} Bd_\alpha + \nabla g(x)\lambda_\alpha + \nabla h(x)\mu_\alpha = -\nabla f(x) \\ \Lambda \nabla g^T(x)d_\alpha + G(x)\lambda_\alpha = 0 \\ \nabla h^T(x)d_\alpha = -h(x) \end{cases}$$

Se  $d_\alpha = 0$ , pare.

II. Calcule  $(d_\beta, \lambda_\beta, \mu_\beta)$  resolvendo o sistema linear:

$$\begin{cases} Bd_\beta + \nabla g(x)\lambda_\beta + \nabla h(x)\mu_\beta = 0 \\ \Lambda \nabla g^T(x)d_\beta + G(x)\lambda_\beta = -\Lambda\omega^i \\ \nabla h^T(x)d_\beta = -\omega^e \end{cases}$$

III. Se  $c_i < -1, 2\mu_{\alpha i}$  então calcule  $c_i = -2\mu_{\alpha i}, i = 1, \dots, p$ .

IV. Se  $d_\beta^T \nabla \phi_c(x) > 0$ , defina

$$\rho = \min \left\{ \varphi \|d_\alpha\|_2^2; (\xi - 1) d_\alpha^T \nabla \phi_c(x) / d_\beta^T \nabla \phi_c(x) \right\}. \quad (5.23)$$

Caso contrário, defina

$$\rho = \varphi \|d_\alpha\|_2^2. \quad (5.24)$$

V. Calcule

$$d = d_\alpha + \rho d_\beta,$$

e também

$$\bar{\lambda} = \lambda_\alpha + \rho \lambda_\beta. \quad (5.25)$$

**Passo 2:** *Busca Linear.*

Calcule  $t$ , o primeiro número da sequência  $\{1, \nu, \nu^2, \nu^3, \dots\}$  satisfazendo

$$\begin{aligned} \phi_c(x + td) &\leq \phi_c(x) + t\eta d^T \nabla \phi_c(x), \\ h(x + td) &\leq 0, \end{aligned}$$

e

$$\begin{aligned} g_i(x + td) &< 0, \text{ se } \bar{\lambda}_i \geq 0, \\ g_i(x + td) &\leq g_i(x), \text{ se } \bar{\lambda}_i < 0. \end{aligned}$$

**Passo 3:** *Atualizações*

I. Faça  $x := x + td$ , e defina novos valores para  $\omega^i > 0, \omega^e > 0, \lambda > 0$  e  $B$  simétrica e positiva definida.

II. Volte ao Passo 1.

## 6 ANÁLISE DA INFLUÊNCIA DE PARÂMETROS NO FDIPA

O algoritmo básico apresentado no capítulo anterior utiliza os parâmetros  $\xi$ ,  $\eta$  e  $\nu$  definidos no intervalo  $(0, 1)$  e também o parâmetro  $\varphi > 0$ . Sabemos pela prova de convergência global do FDIPA, que qualquer valor estabelecido para estes parâmetros dentro destes intervalos, torna possível a execução do algoritmo. Diante disto, faz sentido pensar se há algum valor específico para cada um destes parâmetros, que execute o FDIPA de maneira mais eficiente.

Nesta perspectiva, implementamos uma versão do algoritmo básico em MatLab e realizamos uma série de testes numéricos que nos direcionaram a alguns resultados. Este capítulo tem como objetivo descrever os resultados obtidos com a realização destes testes.

Na versão que implementamos, utilizamos a matriz identidade de ordem  $n \times n$ , como sendo a matriz simétrica positiva-definida  $B$ , e a condição de Armijo para a busca linear, uma vez que os resultados apresentados em [9] também foram obtidos utilizando esta condição. Esta versão do FDIPA resolve problemas que contenham apenas restrições de desigualdades.

Para obter os resultados, resolvemos os problemas com restrições de desigualdades listados no apêndice A. Estes problemas foram obtidos em [10]. Não foi feita nenhuma análise sobre os problemas testados. Cabe ressaltar apenas que eram problemas heterogêneos e de fácil implementação computacional. Tomamos os valores iniciais de  $x$  como em [10], quando esses valores satisfaziam as restrições de desigualdade. Caso contrário, tomávamos um outro ponto interior ao conjunto viável. Adotamos uma tolerância de erro não maior que  $10^{-5}$ , uma vez que esta também foi a tolerância utilizada em [9]. Desta forma, podemos fazer uma comparação dos nossos resultados com os resultados apresentados em [9].

Os testes foram realizados fazendo-se variações nos valores dos parâmetros  $\xi$ ,  $\eta$ ,  $\nu$  e  $\varphi$  dentro dos intervalos em que estão definidos. Para cada parâmetro foram fixados 9 valores (ver Apêndice B) e cada um destes valores foi testado em todos os problemas do Apêndice A. Os testes seguiam uma ordem para a variação dos valores dos parâmetros: a cada problema, fixavam-se  $\eta = 0,5$ ,  $\nu = 0,5$  e  $\varphi = 1$  e testavam-se os 9 valores fixados para  $\xi$ . Estes valores eram colocados em uma tabela junto com o número de iterações que o algoritmo fazia para solucionar o problema. Em seguida, tomava-se o valor de  $\xi$  correspondente ao menor número de iterações e fixava-se este valor para  $\xi$ . Realizava-se então o mesmo procedimento com  $\eta$ ,  $\nu$  e  $\varphi$ , respectivamente.

Houve uma preocupação de que a ordem de teste dos parâmetros pudesse influenciar no número de iterações que o algoritmo desenvolvia. Por exemplo, ao realizar os testes com  $\xi$ , o valor de  $\xi$  que apresenta o menor número de iterações é  $\xi = 0,85$ . Quando

fixamos este valor e fazemos o teste com  $\eta$ , encontramos  $\eta = 0,5$ . No entanto, se primeiro tivéssemos testado o  $\eta$ , pode ser que o valor tomado para  $\xi$  fosse diferente de 0,85. Assim, houve a necessidade da realização de alguns testes que não seguissem a ordem fixada para a variação dos parâmetros. Estes testes foram realizados com os problemas de número 4, 6, 11, 31 e 37 do Apêndice A. Os problemas foram escolhidos de forma aleatória. Aqui, ao invés de seguir a ordem fixada para a variação dos parâmetros (a saber, primeiro  $\xi$ , depois  $\eta$ , em seguida  $\nu$  e por último  $\varphi$ ) os testes eram realizados com os parâmetros escolhidos de forma aleatória, por exemplo, escolhia-se  $\varphi$  para ser o primeiro parâmetro a ser testado, depois testava-se  $\xi$ , em seguida  $\eta$  e por último  $\nu$ . Em nenhum caso, houve diferença entre os valores aos quais correspondiam o menor número de iterações com os parâmetros fixados e com os parâmetros sem fixar.

A fim de facilitar a descrição dos resultados, a partir de agora, chamaremos de *melhor valor* (ou *melhores valores*), aquele(s) associado(s) ao menor número de iterações feitas pelo algoritmo e de *pior valor* (ou *piores valores*) aquele(s) relacionado(s) ao maior número de iterações.

O primeiro resultado obtido relaciona-se a matriz  $B$ . Os resultados numéricos apresentados em [9], utilizaram para a atualização de  $B$ , uma técnica quase-Newton baseada na fórmula do BFGS. Na versão do algoritmo que implementamos, utilizamos a matriz identidade de ordem  $n \times n$ . Em todos os testes numéricos, o número de iterações realizadas pelo algoritmo com a matriz identidade foi superior ao número de iterações que utilizava o método BFGS. Em alguns casos, a diferença entre o número de iterações dos dois métodos era superior a 3000 iterações, como podemos observar no quadro abaixo. Vale destacar que, tanto o método BFGS quanto a matriz identidade de ordem  $n$  são uma aproximação para a Hessiana do problema ( $H(x, \lambda) = \nabla^2 f(x) + \sum_{i=1}^m \lambda_i + \nabla^2 g_i(x)$ ), ou seja, nenhuma delas usa diretamente as derivadas de segunda ordem de  $f$ ,  $g$  e  $h$ . No entanto, fica claro pelos resultados numéricos abaixo que a atualização de  $B$  feita pelo BFGS é mais eficiente que  $I_n$ .

Resultados Numéricos					
$N_p$	$N_{it}(B_{BFGS})$	$N_{it}(I_{n \times n})$	$N_p$	$N_{it}(B_{BFGS})$	$N_{it}(I_{n \times n})$
1	36	1015	24	11	26
2	29	934	25	92	643
3	15	3798	27	13	379
4	4	9	29	7	15
5	5	10	30	11	16
6	6	9	31	6	23
7	10	12	33	12	40
10	10	11	35	5	18
11	6	9	36	12	20
12	5	10	37	10	17
13	30	31	38	14	3833
15	5	12	43	7	30
16	10	35	44	21	24
17	21	23	45	34	46
18	13	249	65	8	51
19	19	20	66	5	93
20	21	26	76	7	31
21	15	148	100	13	169
22	8	179	113	16	193
23	7	17			

$N_p$ : Número do Problema

$N_{it}(B_{BFGS})$ : Número de iterações obtidas utilizando o método BFGS

$N_{it}(I_{n \times n})$ : Número de iterações obtidas utilizando-se a matriz identidade

Nesta tabela, os valores apresentados para  $N_{it}(I_{n \times n})$  correspondem ao menor número de iterações obtidos, ou seja, o  $N_{it}(I_{n \times n})$  obtido com os *melhores valores* de cada parâmetro.

Após a realização dos primeiros testes, já pode-se observar que não havia um *melhor valor* para nenhum dos parâmetros. No entanto, estes valores estavam sempre representados em um intervalo bem mais restrito do que o apresentado no algoritmo base. E ainda, não só havia um intervalo onde os melhores valores estavam definidos, mas também intervalos nos quais deveria-se evitar tomar os valores dos parâmetros, pois nestes intervalos o número de iterações que o algoritmo realizava era sempre muito maior.

Algumas vezes, a diferença entre número de iterações dos melhores valores e dos piores valores em um determinado problema era bem pequena. No entanto, houve casos em que esta diferença foi superior a 5 mil iterações. Tomemos como exemplo a variação

do parâmetro  $\xi$  no Problema 2. Conforme podemos observar na tabela abaixo, a diferença entre o *melhor valor* e o *pior valor* para este parâmetro ultrapassa 5000 iterações.

Problema 2		
Parâmetro	$N_{it}$ com o melhor valor	$N_{it}$ com o pior valor
$\xi$	3757	9383

Assim, podemos perceber a importância da escolha correta dos valores dos parâmetros no FDIPA para que tenhamos um algoritmo robusto para solucionar uma classe de problemas.

Apresentaremos agora quais são os intervalos em que obtemos os *melhores valores* e os *piores valores* para cada um dos parâmetros  $\xi$ ,  $\eta$ ,  $\nu$  e  $\varphi$ .

### 6.1 INTERVALOS PARA $\xi$

O parâmetro  $\xi$  é utilizado no FDIPA para o cálculo adequado de  $\rho$ . Este cálculo deve ser feito para garantir que  $d$  seja uma direção de descida para  $f$ . A escolha de valores muito pequenos para  $\xi$  pode provocar uma deflexão desnecessária em  $d$ , o que implicaria que o algoritmo teria que realizar um número maior de iterações para encontrar a solução do problema. De fato, esta idéia está de acordo com o resultado apontado pelos testes.

Para o parâmetro  $\xi$ , o menor número de iterações foi obtido pelos valores de  $\xi \geq 0,65$ , sendo que em 60% dos problemas testados,  $\xi = 0,85$  foi o melhor valor. A escolha de  $\xi < 0,5$  representa o intervalo nos quais estão definidos os piores valores, sendo que 80% dos piores valores para  $\xi$  ocorreram quando tomamos  $\xi = 0,1$ .

Intervalos para $\xi$	
<i>Melhores Valores</i>	<i>Piores Valores</i>
$0,65 \leq \xi < 1$	$0 < \xi \leq 0,5$

### 6.2 INTERVALOS PARA $\nu$

Para este parâmetro os melhores valores ocorreram quando  $\nu \geq 0,5$ . Houve dois valores de  $\nu$  que se destacaram como melhores:  $\nu = 0,9$  e  $\nu = 0,5$ . Os piores valores ocorreram quando  $\nu \leq 0,5$  sendo que em 90% dos casos  $\nu = 0,1$  foi o *pior valor*.

O parâmetro  $\nu$  interfere diretamente na busca linear do algoritmo, a qual é feita utilizando-se o método de Armijo. Como já vimos anteriormente, a busca de Armijo

procura uma boa redução da função ao longo da direção, sem no entanto tentar minimizá-la. Para isto, é necessário calcular adequadamente o tamanho do passo  $t$ . O parâmetro  $\nu$  é utilizado para encontrar um tamanho de passo que satisfaça a desigualdade de Armijo

$$f(x + td) \leq f(x) + t\eta\nabla^T f(x)d$$

Assim, tanto do ponto de vista computacional quanto teórico, é importante que o tamanho de passo  $t$  satisfazendo a equação acima não seja muito pequeno a fim de que possamos garantir um decréscimo razoável da função objetivo. Como  $t$  é calculado em função de  $\nu$ , era de se esperar que os melhores valores para  $\nu$  seriam aqueles que não fossem muito pequenos.

Intervalos para $\nu$	
Melhores Valores	Piores Valores
$0,5 \leq \nu < 1$	$0 < \nu < 0,5$

### 6.3 INTERVALOS PARA $\eta$

O parâmetro  $\eta$ , assim como o  $\nu$  é utilizado na busca de Armijo. Como mencionado anteriormente, deve-se evitar que o tamanho do passo  $t$  seja muito pequeno. No entanto, pode ser que tenhamos passos grandes, mas com pouco decréscimo da função objetivo. Desta forma, o parâmetro  $\eta$  serve para ajustar o decréscimo da função para que este seja proporcional ao tamanho do passo. Neste caso, valores muito grandes de  $\eta$  não seriam apropriados. De fato, os testes realizados corroboram com esta análise teórica.

Os melhores valores de  $\eta$  ocorreram no intervalo  $0,25 \leq \eta \leq 0,65$ , sendo que em 75% dos casos o melhor valor foi representado por  $\eta = 0,5$ . Os piores casos ocorrem para valores de  $\eta$  maiores que  $0,75$ . Em 70% dos casos, o pior valor foi  $\eta = 0,9$ .

Intervalos para $\eta$	
Melhores Valores	Piores Valores
$0,25 \leq \eta \leq 0,65$	$0,75 \leq \eta < 1$

### 6.4 INTERVALOS PARA $\varphi$

Em nossos testes, variamos  $\varphi$  entre valores muito grandes e muito pequenos. Os valores de  $\varphi$  menores que 1 mostraram-se os *piores valores*. Para os valores de  $\varphi$  maiores que 10, embora nem sempre tenham sido valores ruins, todas as vezes que algum valor



de  $\varphi > 10$  era um *melhor valor*, sempre conseguimos um outro valor de  $\varphi$  entre 1 e 10, que desempenhava no algoritmo um número de iterações equivalente ao número de iterações daquele valor. Desta maneira, o intervalo  $[1, 10]$  contém os *melhores valores* a serem tomados para  $\varphi$ . Dentre estes valores, na maioria dos problemas, o *melhor valor* para  $\varphi$  foi  $\varphi = 2$ .

Intervalos para $\varphi$	
<i>Melhores Valores</i>	<i>Piores Valores</i>
$1 \leq \varphi \leq 10$	$0 < \varphi < 1$

## 6.5 ANÁLISE DOS PARÂMETROS $\omega$ E $\lambda$

Também foram testados diferentes valores para  $\omega$  e  $\lambda$ , seguindo as mesmas linhas dos testes realizados para os demais parâmetros. No entanto, ao contrário do que acontecia nos outros casos, a ordem com que se variava estes parâmetros influenciava no número de iterações obtidos. Desta forma, os testes de variação destes parâmetros foi feita de maneira independente da variação dos demais. Fixamos os valores de  $\eta = 0,5$ ,  $\nu = 0,5$ ,  $\xi = 0,85$  e  $\varphi = 2$ , conforme os resultados anteriores e testamos diferentes valores para  $\lambda > 0$  e  $\omega > 0$ . Como  $\omega$  e  $\lambda$  são vetores de  $\mathbb{R}^m$ , consideramos em cada caso todas as componentes do vetor iguais, pois assim teríamos um forma de analisar estes parâmetros que seria mais bem definida e que demandaria menos tempo na atribuição computacional dos valores.

Embora não tenhamos conseguido obter um padrão bem definido para estes parâmetros, pode-se observar que não é conveniente tomar valores muito grandes para eles. Através dos testes realizados, recomendamos tomar os valores destes parâmetros menores que 1. Para os testes realizados com os parâmetros  $\xi$ ,  $\eta$ ,  $\nu$  e  $\varphi$  utilizamos  $\lambda = 0,25$  e  $\omega = 1$ .

## 6.6 CONCLUSÕES

Através dos resultados obtidos podemos ver que uma boa escolha de cada um dos parâmetros apresentados tem influência direta no desempenho computacional do algoritmo. Analisando os parâmetros utilizados por Herkovits para obter os resultados apresentados em [9] vemos que alguns deles foram tomados dentro dos intervalos apresentados em nossos estudos, a saber  $\xi = 0,7$  e  $\eta = 0,7$ . Entretanto, alguns outros não são especificados na referência [9]. Pode-se pensar que um estudo de variação dos parâmetros do FDIPA feita com a versão geral apresentada em [9] pode diminuir ainda mais o número de iterações que o algoritmo necessita para resolver os problemas, e ainda, tal estudo poderá confirmar se

os resultados obtidos independem ou não da forma de atualização da matriz  $B$ , da busca linear e das restrições do problema.

## 7 CONSIDERAÇÕES FINAIS

Este trabalho tem como objetivo principal apresentar um estudo da influência dos parâmetros de um algoritmo de pontos interiores e direções viáveis na solução de problemas de otimização não linear. O estudo apresentado neste trabalho, foi feito considerando-se um algoritmo particular do FDIPA que resolve problemas contendo apenas restrições de desigualdade. Através da realização de uma série de testes numéricos, definimos intervalos que possibilitam uma escolha adequada de parâmetros que levam a um número menor de iterações na resolução dos problemas testes.

O FDIPA vem sendo amplamente utilizado em diversas áreas, seja para aplicações práticas ou para a criação de novos algoritmos que utilizam seus sistemas internos. Este estudo pode contribuir para reduzir o custo computacional do algoritmo tornando-o ainda mais eficiente para se resolver problemas de grande porte

Como sugestão de estudos futuros, proponho os seguintes tópicos:

- Fazer um estudo da variação dos parâmetros utilizando o algoritmo geral apresentado em [9].
- Estudar possíveis regras de atualização da matriz  $B$ .

## REFERÊNCIAS

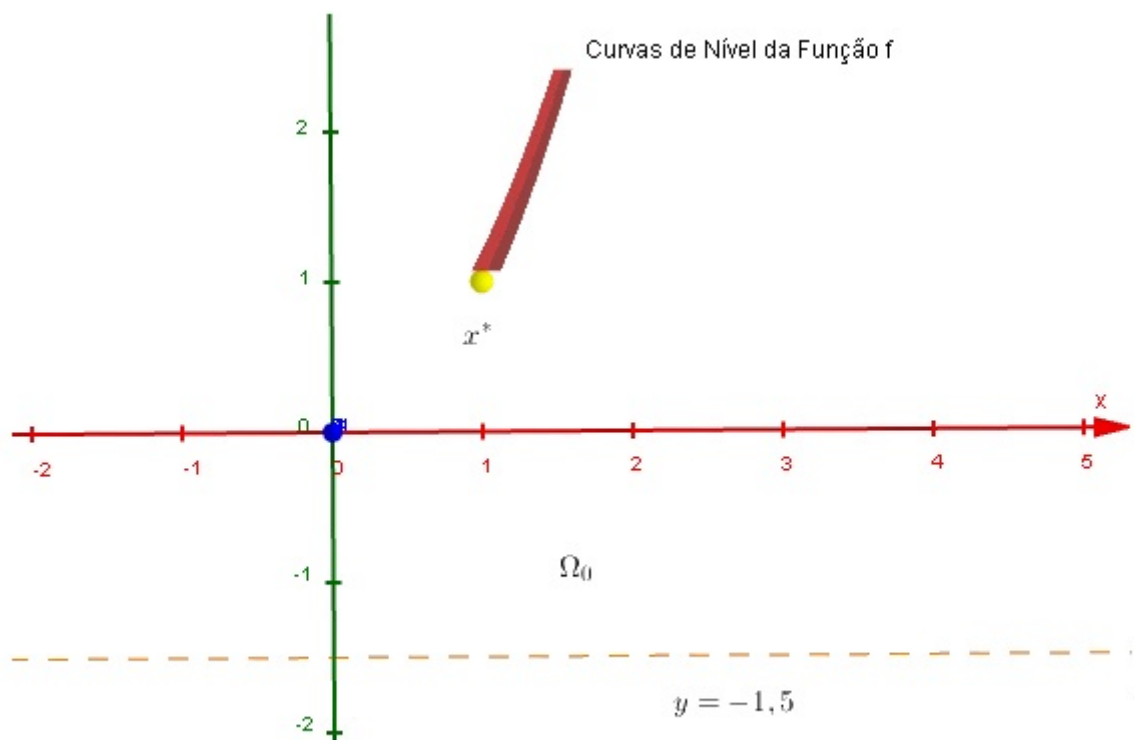
- [1] BAZARAA, M. S., SHERALY, H. DD., SHETTY, C. M., *Nonlinear Programming: Theory and Algorithms*, 3 ed. John Wiley and Sons, New York, 2006.
- [2] BERNADÁ, Gabriel M. Guerra. *Técnicas de Otimização Aplicadas a Problemas de Contato Elástico*. Dissertação de Mestrado, COPPE/PEM - Universidade Federal do Rio de Janeiro, 2006.
- [3] BERTSEKAS, D. P., *Nonlinear Programming*. Athena Scientific, Belmont, USA, 1999.
- [4] BOTTA, Alfredo Canelas. *Um Algoritmo de Newton de Ponto Interior e Aplicações na Fundação Eletromagnética de Metais*. Tese de Doutorado, COPPE/PEM - Universidade Federal do Rio de Janeiro, 2009.
- [5] BOYER, C. B., *História da Matemática*, 3 ed. Edgard Bluncher, São Paulo, 2003.
- [6] FÁLCON, Ginés A. Santos. *Uma Família de Algoritmos de Ponto Interior para Projeto Ótimo em Engenharia*. Tese de Doutorado, COPPE/PEM - Universidade Federal do Rio de Janeiro.
- [7] FREIRE, W. P., *Um algoritmo de Direções viáveis para Otimização Convexa Não Diferenciável*. Tese de Doutorado, COPPE/PEM - Universidade Federal do Rio de Janeiro, 2005.
- [8] FRIEDLANDER, A., *Elementos de Programação Não Linear*. Campinas, SP: Editora da UNICAMP, 1994.
- [9] HERSKOVITS, J., *Feasible Direction Interior Point Technique for Nonlinear Optimization*, Journal of Optimization Theory and Applications, v. 99, pp. 121-146, 1998.
- [10] HOCK, W. e SCHITTKOWSKI, K., *Test Examples for Nonlinear Programming Codes*, Lectures Notes in Economics and Mathematical Systems. Springer Verlag, Berlin, Germany, v. 187, n. 3, 1981.
- [11] IZMAILOV, A. e SOLODOV, M., *Otimização Vol. 1: Condições de Otimzalidade, Elementos de Análise Convexa e de Dualidade*. 3 ed. Rio de Janeiro: IMPA, 2014.
- [12] IZMAILOV, A. e SOLODOV, M., *Otimização Vol. 2: Condições de Otimzalidade, Métodos Computacionais*. 2 ed. Rio de Janeiro: IMPA, 2014.
- [13] LIMA, E. L., *Análise Real Vol. 1: Funções de uma variável*. 10 ed. Rio de Janeiro: IMPA, 2010.
- [14] LIMA, E. L., *Análise Real Vol. 2: Funções de n variáveis*. 5 ed. Rio de Janeiro: IMPA, 2010.
- [15] LIPSCHUTZ, S., *Álgebra Linear: Teoria e problemas*. 3 ed. São Paulo. Pearson Makron Books, 2004.
- [16] LUENBERG, D. G. *Linear and Nonlinear Programming*. 2 Ed., Addison-Wesley Publishing Company, 1984.

- [17] PERESSINI, A. L.; SULLIVAN, F. E., *The Mathematics of nonlinear programming*. New York: Springer-Verlag, 1988.
- [18] RIBEIRO, A. e KARAS, E. W., *Otimização Contínua: Aspectos teóricos e computacionais*. São Paulo: Cengage Learning, 2013.

## APÊNDICE A – Problemas Utilizados Para Realização dos Testes

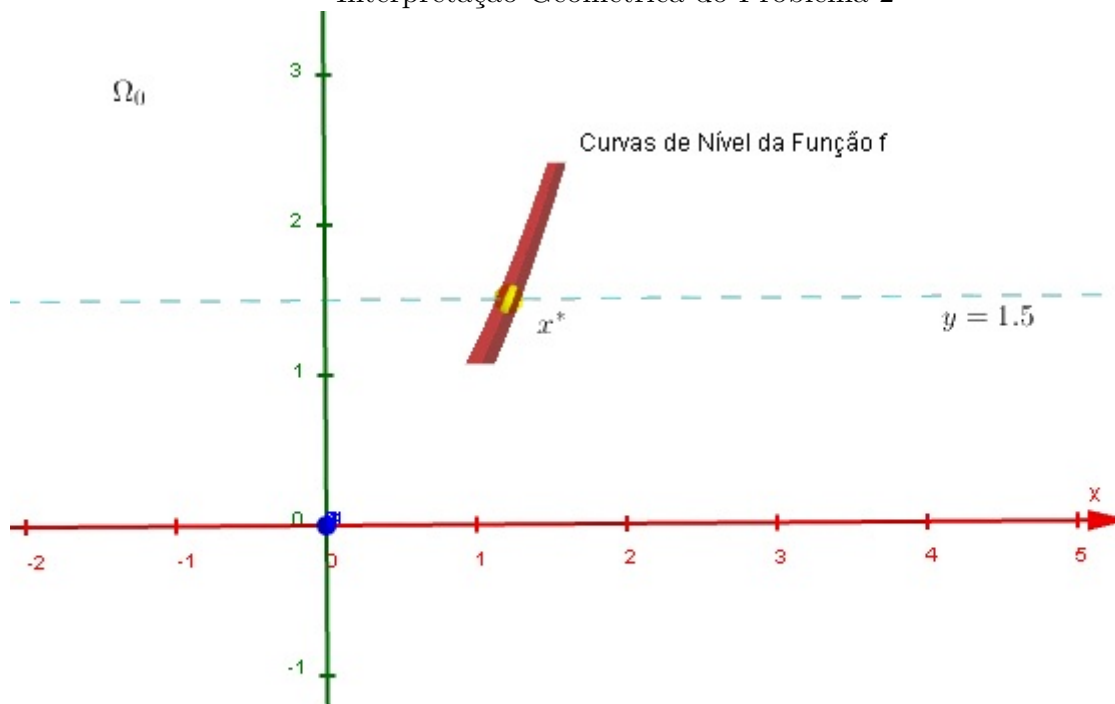
Problema 1	
<i>Função Objetivo</i>	$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$
<i>Restrição</i>	$-1,5 \leq x_2$
<i>Ponto Inicial</i>	$x_0 = (-2, 1)$
<i>Solução</i>	$x^* = (1, 1)$ e $f(x^*) = 0$

### Interpretação Geométrica do Problema 1



<b>Problema 2</b>	
<i>Função Objetivo</i>	$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$
<i>Restrição</i>	$1,5 \leq x_2$
<i>Ponto Inicial</i>	$x_0 = (-2, 1)$
<i>Solução</i>	$x^* = (1.4118546 \cos(\frac{1}{3} \arccos \frac{1}{140.714839}), 1.5)$ $f(x^*) = 0,5042619$

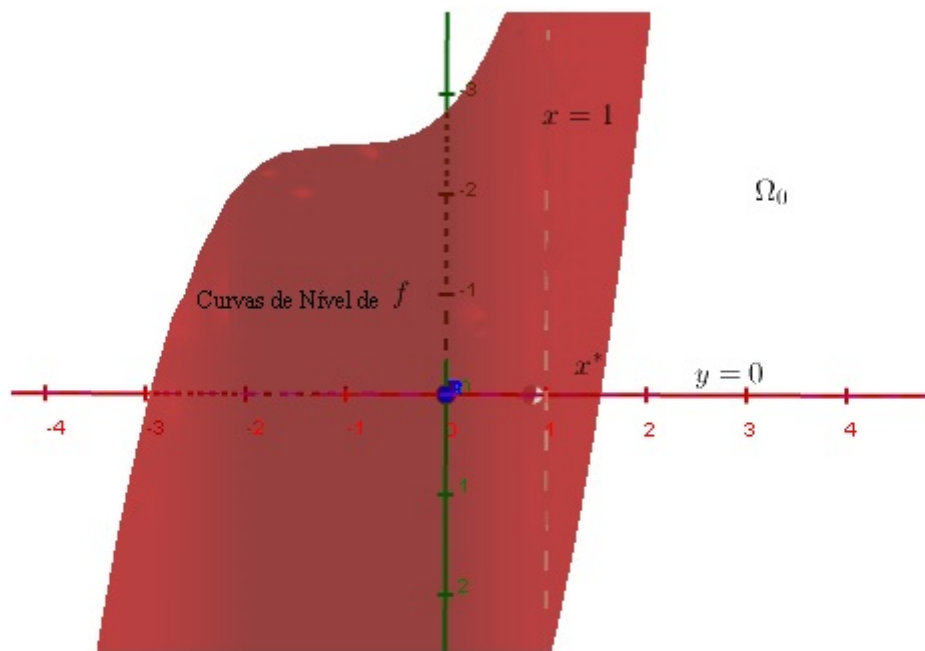
Interpretação Geométrica do Problema 2



<b>Problema 3</b>	
<i>Função Objetivo</i>	$f(x) = x_2 + 10^{-5}(x_2 - x_1^2)$
<i>Restrição</i>	$0 \leq x_2$
<i>Ponto Inicial</i>	$x_0 = (10, 1)$
<i>Solução</i>	$x^* = (0, 0)$ e $f(x^*) = 0$

Problema 4	
Função Objetivo	$f(x) = \frac{1}{3}(x_1 + 1)^3 + x_2$
Restrições	$1 \leq x_1$ $0 \leq x_2$
Ponto Inicial	$x_0 = (1.125, 0.125)$
Solução	$x^* = (1, 0)$ e $f(x^*) = \frac{8}{3}$

Interpretação Geométrica do Problema 4

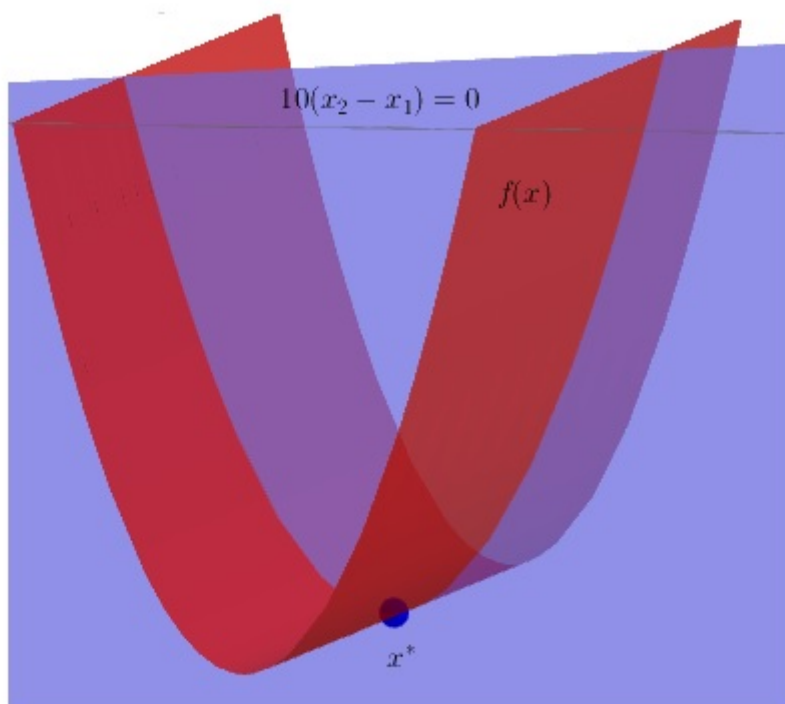


Problema 5	
Função Objetivo	$f(x) = \sin(x_1 + x_2) + (x_1 - x_2)^2 - 1, 5x_1 + 2, 5x_2 + 1$
Restrições	$-1, 5 \leq x_1 \leq 4$ $-3 \leq x_2 \leq 3$
Ponto Inicial	$x_0 = (0, 0)$
Solução	$x^* = \left(-\frac{\pi}{3} + \frac{1}{2}, -\frac{\pi}{3} - \frac{1}{2}\right)$ e $f(x^*) = -\frac{\sqrt{3}}{2} - \frac{\pi}{3}$



Problema 6	
<i>Função Objetivo</i>	$f(x) = (1 - x_1)^2$
<i>Restrições</i>	$10(x_2 - x_1^2) \leq 0$
<i>Ponto Inicial</i>	$x_0 = (-1.2, 1)$
<i>Solução</i>	$x^* = (1, 1)$ e $f(x^*) = 0$

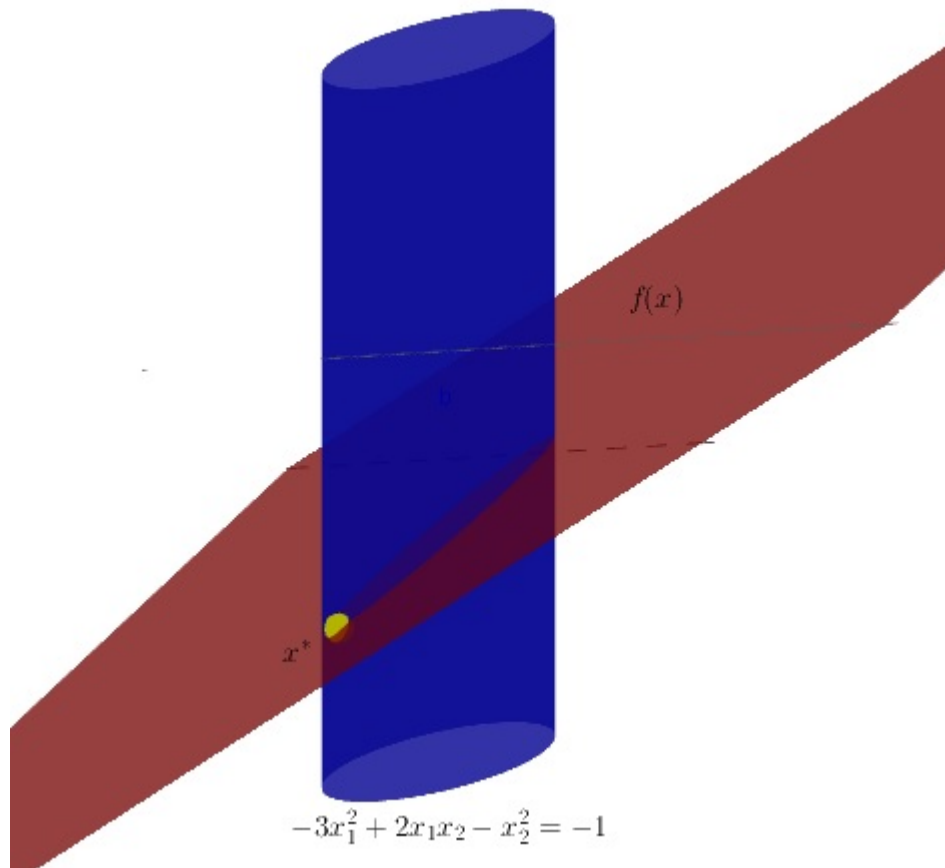
Interpretação Geométrica do Problema 6



Problema 7	
<i>Função Objetivo</i>	$f(x) = \ln(1 + x_1^2) - x_2$
<i>Restrições</i>	$(1 + x_1^2)^2 + x_2^2 - 4 = 0$
<i>Ponto Inicial</i>	$x_0 = (1, 0)$
<i>Solução</i>	$x^* = (0, \sqrt{3})$ e $f(x^*) = -\sqrt{3}$

Problema 10	
<i>Função Objetivo</i>	$f(x) = x_1 - x_2$
<i>Restrições</i>	$-3x_1^2 + 2x_1x_2 - x_2^2 + 1 \geq 0$
<i>Ponto Inicial</i>	$x_0 = (-10, 10)$
<i>Solução</i>	$x^* = (0, 1)$ e $f(x^*) = -1$

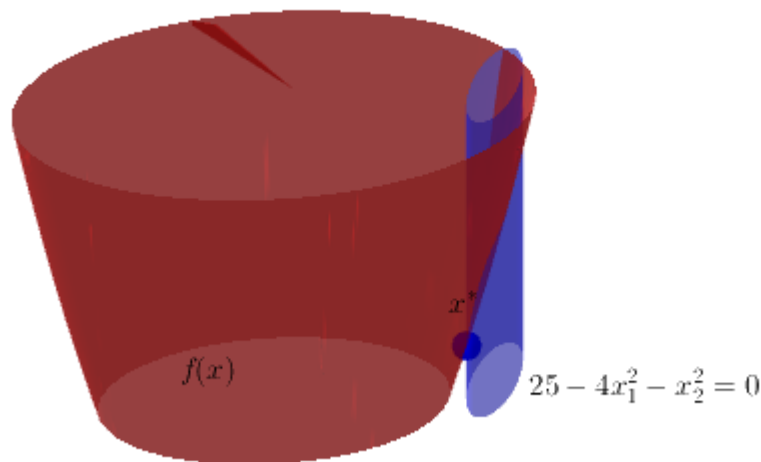
Interpretação Geométrica do Problema 10



Problema 11	
<i>Função Objetivo</i>	$f(x) = (x_1 - 5)^2 + x_2^2 - 25$
<i>Restrições</i>	$-x_1^2 + x_2 \geq 0$
<i>Ponto Inicial</i>	$x_0 = (-1, 10)$
<i>Solução</i>	$x^* = \left( \left( a - \frac{1}{a} \right) / \sqrt{6}, \frac{a^2 - 2 + a^{-2}}{6} \right)$ , onde $a = 7.5\sqrt{6} + \sqrt{338.5}$ $f(x^*) = -8.4984642$

<b>Problema 12</b>	
<i>Função Objetivo</i>	$f(x) = 0,5x_1^2 + x_2^2 - x_1x_2 - 7x_1 - 7x_2$
<i>Restrições</i>	$25 - 4x_1^2 - x_2^2 \geq 0$
<i>Ponto Inicial</i>	$x_0 = (0, 0)$
<i>Solução</i>	$x^* = (2, 3)$ e $f(x^*) = -30$

Interpretação Geométrica do Problema 12



<b>Problema 13</b>	
<i>Função Objetivo</i>	$f(x) = (x_1 - 2)^2 + x_2^2$
<i>Restrições</i>	$(1 - x_1)^3 - x_2 \geq 0$ $0 \leq x_1$ $0 \leq x_2$
<i>Ponto Inicial</i>	$x_0 = (0, 0)$
<i>Solução</i>	$x^* = (1, 0)$ e $f(x^*) = 1$

<b>Problema 15</b>	
<i>Função Objetivo</i>	$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$
<i>Restrições</i>	$x_1x_2 - 1 \geq 0$ $x_1 + x_2^2 \geq 0$ $x_1 \leq 0,5$
<i>Ponto Inicial</i>	$x_0 = (0,4, 4)$
<i>Solução</i>	$x^* = (0,5, 2)$ e $f(x^*) = 306,5$

<b>Problema 16</b>	
<i>Função Objetivo</i>	$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$
<i>Restrições</i>	$x_1 + x_2^2 \geq 0$ $x_1^2 + x_2 \geq 0$ $-0,5 \leq x_1 \leq 0,5$ $x_2 \leq 1$
<i>Ponto Inicial</i>	$x_0 = (0, 1)$
<i>Solução</i>	$x^* = (0.5, 0.25)$ e $f(x^*) = 0.25$

<b>Problema 17</b>	
<i>Função Objetivo</i>	$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$
<i>Restrições</i>	$x_2^2 - x_1 \geq 0$ $x_1^2 - x_2 \geq 0$ $-0,5 \leq x_1 \leq 0,5$ $x_2 \leq 1$
<i>Ponto Inicial</i>	$x_0 = (0, -1)$
<i>Solução</i>	$x^* = (0, 0)$ e $f(x^*) = 1$

<b>Problema 18</b>	
<i>Função Objetivo</i>	$f(x) = 0,01x_1^2 + x_2^2$
<i>Restrições</i>	$x_1x_2 - 25 \geq 0$ $x_1^2 + x_2^2 - 25 \geq 0$ $2 \leq x_1 \leq 50$ $0 \leq x_2 \leq 50$
<i>Ponto Inicial</i>	$x_0 = (6, 6)$
<i>Solução</i>	$x^* = (\sqrt{250}, \sqrt{2.5})$ e $f(x^*) = 5$

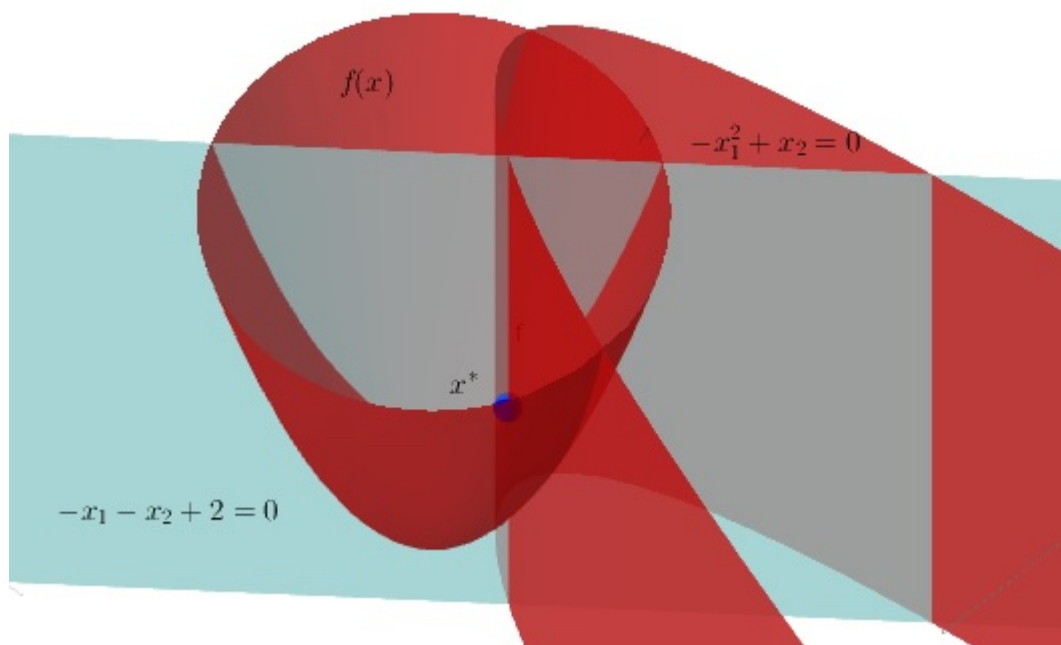
<b>Problema 19</b>	
<i>Função Objetivo</i>	$f(x) = (x_1 - 10)^3 + (x_2 - 20)^3$
<i>Restrições</i>	$(x_1 - 5)^2 + (x_2 - 5)^2 - 100 \geq 0$ $-(x_2 - 5)^2 - (x_1 - 6)^2 + 82.81 \geq 0$ $13 \leq x_1 \leq 100$ $0 \leq x_2 \leq 100$
<i>Ponto Inicial</i>	$x_0 = (15, 6)$
<i>Solução</i>	$x^* = (14.095, 0.84296079)$ e $f(x^*) = -6961.81381$

<b>Problema 20</b>	
<i>Função Objetivo</i>	$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$
<i>Restrições</i>	$x_1 + x_2^2 \geq 0$ $x_1^2 + x_2 \geq 0$ $x_1^2 + x_2^2 - 1 \geq 0$ $-0,5 \leq x_1 \leq 0,5$
<i>Ponto Inicial</i>	$x_0 = (-2, 1)$
<i>Solução</i>	$x^* = (0.5, 0.5\sqrt{3})$ e $f(x^*) = 81.5 - 25\sqrt{3}$

<b>Problema 21</b>	
<i>Função Objetivo</i>	$f(x) = 0,01x_1^2 + x_2^2 - 100$
<i>Restrições</i>	$10x_1 - x_2 - 10 \geq 0$ $2 \leq x_1 \leq 50$ $-50 \leq x_2 \leq 50$
<i>Ponto Inicial</i>	$x_0 = (3, -4)$
<i>Solução</i>	$x^* = (2, 0)$ e $f(x^*) = -99.96$

Problema 22	
<i>Função Objetivo</i>	$f(x) = (x_1 - 2)^2 + (x_2 - 1)^2$
<i>Restrições</i>	$-x_1 - x_2 + 2 \geq 0$ $-x_1^2 + x_2 \geq 0$
<i>Ponto Inicial</i>	$x_0 = (-1, 2)$
<i>Solução</i>	$x^* = (1, 1)$ e $f(x^*) = 1$

Interpretação Geométrica do Problema 22



Problema 23	
<i>Função Objetivo</i>	$f(x) = x_1^2 + x_2^2$
<i>Restrições</i>	$x_1 + x_2 - 1 \geq 0$ $x_1^2 + x_2^2 - 1 \geq 0$ $9x_1^2 + x_2^2 - 9 \geq 0$ $x_1^2 - x_2 \geq 0$ $x_2^2 - x_1 \geq 0$ $-50 \leq x_i \leq 50, i = \{1, 2\}$
<i>Ponto Inicial</i>	$x_0 = (3, 3)$
<i>Solução</i>	$x^* = (1, 1)$ e $f(x^*) = 2$

<b>Problema 24</b>	
<i>Função Objetivo</i>	$f(x) = \frac{1}{27\sqrt{3}} \left( (x_1 - 3)^2 - 9 \right) x_2^3$
<i>Restrições</i>	$\frac{x_1}{\sqrt{3}} - x_2 \geq 0$ $x_1 + \sqrt{3}x_2 \geq 0$ $-x_1 - \sqrt{3}x_2 + 6 \geq 0$ $0 \leq x_1$ $0 \leq x_2$
<i>Ponto Inicial</i>	$x_0 = (1, 0.5)$
<i>Solução</i>	$x^* = (3, \sqrt{3})$ e $f(x^*) = -1$

<b>Problema 26</b>	
<i>Função Objetivo</i>	$f(x) = (x_1 - x_2)^2 + (x_2 - x_3)^4$
<i>Restrições</i>	$(1 + x_2^2)x_1 + x_3^4 - 3 = 0$
<i>Ponto Inicial</i>	$x_0 = (-2.6, 2, 2)$
<i>Solução</i>	$x^* = (1, 1, 1)$ e $f(x^*) = 0$

<b>Problema 27</b>	
<i>Função Objetivo</i>	$f(x) = 0,01(x_1 - 1)^2 + (x_2 - x_1^2)^2$
<i>Restrições</i>	$x_1 + x_3^2 + 1 = 0$
<i>Ponto Inicial</i>	$x_0 = (-5, 2, 2)$
<i>Solução</i>	$x^* = (-1, 1, 0)$ e $f(x^*) = 0.04$

<b>Problema 29</b>	
<i>Função Objetivo</i>	$f(x) = -x_1x_2x_3$
<i>Restrições</i>	$-x_1^2 - 2x_2^2 - 4x_3^3 + 48 \geq 0$
<i>Ponto Inicial</i>	$x_0 = (1, 1, 1)$
<i>Solução</i>	$x^* = (a, b, c), x^* = (a, -b, -c), x^* = (-a, b, -c)$ ou $x^* = (-a, -b, c)$ onde $a = 4, b = 2\sqrt{2}$ e $c = 2$ . $f(x^*) = -16\sqrt{2}$

<b>Problema 30</b>	
<i>Função Objetivo</i>	$f(x) = x_1^2 + x_2^2 + x_3^2$
<i>Restrições</i>	$x_1^2 + x_2^2 - 1 \geq 0$ $1 \leq x_1 \leq 10$ $-10 \leq x_2 \leq 10$ $-10 \leq x_3 \leq 10$
<i>Ponto Inicial</i>	$x_0 = (1, 1, 1)$
<i>Solução</i>	$x^* = (1, 0, 0)$ e $f(x^*) = 1$

<b>Problema 31</b>	
<i>Função Objetivo</i>	$f(x) = 9x_1^2 + x_2^2 + 9x_3^2$
<i>Restrições</i>	$x_1x_2 - 1 \geq 0$ $-10 \leq x_1 \leq 10$ $1 \leq x_2 \leq 10$ $-10 \leq x_3 \leq 1$
<i>Ponto Inicial</i>	$x_0 = (1, 1, 1)$
<i>Solução</i>	$x^* = (1/\sqrt{3}, \sqrt{3}, 0)$ e $f(x^*) = 6$

<b>Problema 33</b>	
<i>Função Objetivo</i>	$f(x) = (x_1 - 1)(x_1 - 2)(x_1 - 3) + x_3$
<i>Restrições</i>	$x_3^2 - x_2^2 - x_1^2 \geq 0$ $x_1^2 + x_2^2 + x_3^2 - 4 \geq 0$ $0 \leq x_1$ $0 \leq x_2$ $0 \leq x_3 \leq 5$
<i>Ponto Inicial</i>	$x_0 = (0, 0, 3)$
<i>Solução</i>	$x^* = (0, \sqrt{2}, \sqrt{2})$ e $f(x^*) = \sqrt{2} - 6$

<b>Problema 35</b>	
<i>Função Objetivo</i>	$f(x) = 9 - 8x_1 - 6x_2 - 4x_3 + 2x_1^2 + 2x_2^2 + x_3^2 + 2x_1x_2 + 2x_1x_3$
<i>Restrições</i>	$3 - x_1 - x_2 - 2x_3 \geq 0$ $0 \leq x_i, i = \{1, 2, 3\}$
<i>Ponto Inicial</i>	$x_0 = (0.5, 0.5, 0.5)$
<i>Solução</i>	$x^* = (4/3, 7/9, 4/9)$ e $f(x^*) = 1/9$



<b>Problema 36</b>	
<i>Função Objetivo</i>	$f(x) = -x_1x_2x_3$
<i>Restrições</i>	$72 - x_1 - 2x_2 - 2x_3 \geq 0$ $0 \leq x_1 \leq 20$ $0 \leq x_2 \leq 11$ $0 \leq x_3 \leq 42$
<i>Ponto Inicial</i>	$x_0 = (10, 10, 10)$
<i>Solução</i>	$x^* = (20, 11, 15)$ e $f(x^*) = -3300$

<b>Problema 37</b>	
<i>Função Objetivo</i>	$f(x) = -x_1x_2x_3$
<i>Restrições</i>	$72 - x_1 - 2x_2 - 2x_3 \geq 0$ $x_1 + 2x_2 + 2x_3 \geq 0$ $0 \leq x_i \leq 42, i = \{1, 2, 3\}$
<i>Ponto Inicial</i>	$x_0 = (10, 10, 10)$
<i>Solução</i>	$x^* = (24, 12, 12)$ e $f(x^*) = -3456$

<b>Problema 38</b>	
<i>Função Objetivo</i>	$f(x) = -100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 90(x_4 - x_3^2)^2 +$ $(1 - x_3)^2 + 10,1((x_2 - 1)^2 + (x_4 - 1)^2) + 19,8(x_2 - 1)(x_4 - 1)$
<i>Restrições</i>	$-10 \leq x_i \leq 10, i = \{1, \dots, 4\}$
<i>Ponto Inicial</i>	$x_0 = (-3, -1, -3, -1)$
<i>Solução</i>	$x^* = (1, 1, 1, 1)$ e $f(x^*) = 0$

<b>Problema 43</b>	
<i>Função Objetivo</i>	$f(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4$
<i>Restrições</i>	$8 - x_1^2 - x_2^2 - x_3^2 - x_4^2 - x_1 + x_2 - x_3 + x_4 \geq 0$ $10 - x_1^2 - 2x_2^2 - x_3^2 - 2x_4^2 + x_1 + x_4 \geq 0$ $5 - 2x_1^2 - x_2^2 - x_3^2 - 2x_1 + x_2 + x_4 \geq 0$
<i>Ponto Inicial</i>	$x_0 = (0, 0, 0, 0)$
<i>Solução</i>	$x^* = (0, 1, 2, -1)$ e $f(x^*) = -44$

<b>Problema 44</b>	
<i>Função Objetivo</i>	$f(x) = x_1 - x_2 - x_3 - x_1x_3 + x_1x_4 + x_2x_3 - x_2x_4$
<i>Restrições</i>	$8 - x_1 - 2x_2 \geq 0$ $12 - 4x_1 - x_2 \geq 0$ $12 - 3x_1 - 4x_2 \geq 0$ $8 - 2x_3 - x_4 \geq 0$ $8 - x_3 - 2x_4 \geq 0$ $5 - x_3 - x_4 \geq 0$ $0 \leq x_i, i = \{1, \dots, 4\}$
<i>Ponto Inicial</i>	$x_0 = (0, 0, 0, 0)$
<i>Solução</i>	$x^* = (0, 3, 0, 4)$ e $f(x^*) = -15$

<b>Problema 45</b>	
<i>Função Objetivo</i>	$f(x) = 2 - \frac{1}{120}x_1x_2x_3x_4x_5$
<i>Restrições</i>	$0 \leq x_i \leq i, i = \{1, \dots, 5\}$
<i>Ponto Inicial</i>	$x_0 = (0.5, 2, 2, 2, 2)$
<i>Solução</i>	$x^* = (1, 2, 3, 4, 5)$ e $f(x^*) = 1$

<b>Problema 65</b>	
<i>Função Objetivo</i>	$f(x) = (x_1 - x_2)^2 + \frac{(x_1+x_2-10)^2}{9} + (x_3 - 5)^2$
<i>Restrições</i>	$48 - x_1^2 - x_2^2 - x_3^2 \geq 0$ $-4, 5 \leq x_i \leq 4, 5, i = \{1, 2\}$ $-5 \leq x_3 \leq 5$
<i>Ponto Inicial</i>	$x_0 = (-5, 5, 0)$
<i>Solução</i>	$x^* = (3.6504618, 3.6504618, 4.6204171)$ e $f(x^*) = 0.9635289$

<b>Problema 66</b>	
<i>Função Objetivo</i>	$f(x) = 0, 2x_3 - 0, 8x_1$
<i>Restrições</i>	$x_2 - \exp(x_1) \geq 0$ $x_3 - \exp(x_2) \geq 0$ $0 \leq x_1 \leq 100$ $0 \leq x_2 \leq 100$ $0 \leq x_3 \leq 10$
<i>Ponto Inicial</i>	$x_0 = (0, 1.05, 2.9)$
<i>Solução</i>	$x^* = (0.1841265, 1.2021679, 3.3273223)$ e $f(x^*) = 0.5181632$

<b>Problema 76</b>	
<i>Função Objetivo</i>	$f(x) = x_1^2 + 0,5x_2^2 + x_3^2 + 0,5x_4^2 - x_1x_3 + x_3x_4 - x_1 - 3x_2 + x_3 - x_4$
<i>Restrições</i>	$5 - x_1 - 2x_2 - x_3 - x_4 \geq 0$ $4 - 3x_1 - x_2 - 2x_3 + x_4 \geq 0$ $x_2 + 4x_3 - 1,5 \geq 0$ $0 \leq x_i, i = \{1, \dots, 4\}$
<i>Ponto Inicial</i>	$x_0 = (0.5, 0.5, 0.5, 0.5)$
<i>Solução</i>	$x^* = (0.2727273, 2.090909, -0.26e^{-10}, 0.5454545)$ e $f(x^*) = -4.6818181$

<b>Problema 100</b>	
<i>Função Objetivo</i>	$f(x) = (x_1 - 10)^2 + 5(x_2 - 12)^2 + x_3^4 + 3(x_4 - 11)^2 + 10x_5^6 + 7x_6^2 + x_7^4 - 4x_6x_7 - 10x_6 - 8x_7$
<i>Restrições</i>	$127 - 2x_1^2 - 3x_2^4 - x_3 - 4x_4^2 - 5x_5 \geq 0$ $282 - 7x_1 - 3x_2 - 10x_3^2 - x_4 + x_5 \geq 0$ $196 - 23x_1 - x_2^2 - 6x_6^2 + 8x_7 \geq 0$ $-4x_1^2 - x_2^2 + 3x_1x_2 - 2x_3^2 - 5x_6 + 11x_7 \geq 0$
<i>Ponto Inicial</i>	$x_0 = (1, 2, 0, 4, 0, 1, 1)$
<i>Solução</i>	$x^* = (2.330499, 1.951372, -0.4775414, 4.365726,$ $-0.6244870, 1.038131, 1.594227)$ $f(x^*) = 680.6300573$

<b>Problema 113</b>	
<i>Função Objetivo</i>	$f(x) = x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 + 4(x_4 - 5)^2 + (x_5 - 3)^2 + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2 + (x_{10} - 7)^2 + 45$
<i>Restrições</i>	$105 - 4x_1 - 5x_2 + 3x_7 - 9x_8 \geq 0$ $-10x_1 + 8x_2 + 17x_7 - 2x_8 \geq 0$ $8x_1 - 2x_2 - 5x_9 + 2x_{10} + 12 \geq 0$ $-3(x_1 - 2)^2 - 4(x_2 - 3)^2 - 2x_3^2 + 7x_4 + 120 \geq 0$ $-5x_1^2 - 8x_2 - (x_3 - 6)^2 + 2x_4 + 40 \geq 0$ $-0,5(x_1 - 8)^2 - 2(x_2 - 4)^2 - 3x_5^2 + x_6 + 30 \geq 0$ $-x_1^2 - 2(x_2 - 2)^2 + 2x_1x_2 - 14x_5 + 6x_6 \geq 0$ $3x_1 - 6x_2 - 12(x_9 - 8)^2 + 7x_{10} \geq 0$
<i>Ponto Inicial</i>	$x_0 = (2, 3, 5, 5, 1, 2, 7, 3, 6, 10)$
<i>Solução</i>	$x^* = (2.171996, 2.363683, 8.773826, 5.095984, 0.9906548,$ $1.430574, 1.321644, 9.828726, 8.280092, 8.375927)$ $f(x^*) = 24.3062091$

**APÊNDICE B – Valores Atribuídos aos Parâmetros Para a Realização dos Testes Numéricos**

Valores Atribuídos a $\xi$								
0,9	0,85	0,75	0,65	0,5	0,35	0,3	0,25	0,1

Valores Atribuídos a $\eta$								
0,9	0,8	0,75	0,65	0,5	0,35	0,3	0,25	0,1

Valores Atribuídos a $\nu$								
0,9	0,8	0,75	0,65	0,5	0,35	0,3	0,25	0,1

Valores Atribuídos a $\varphi$								
0,5	1	2	2,5	5	10	20	50	100