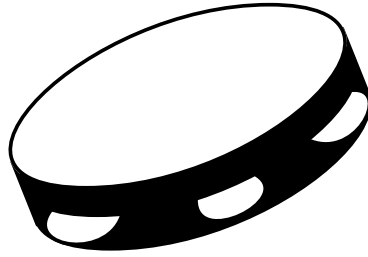


PRIMEIRA EDIÇÃO



---

# Métodos numéricos

---

ESTUDO DE PROBLEMAS INVERSOS

Fernando Pujaico Rivera

*Student number:*  
sXXXXXXXXX

9 de abril do XXXXXXXXXXXX

Copyright © XXXXXXXXXXXX Fernando Pujaico Rivera

IMPRESSO NO BRASIL – ISBN:XXXXXXXXXX

PUBLICADO POR EDITORA XXXXXXXXXXXX

PRIMEIRA IMPRESSÃO , 9 DE ABRIL DO XXXXXXXXXXXX

### Ficha catalográfica

Fernando Pujaico Rivera.

Métodos numéricos : Estudo de problemas inversos/ Fernando Pujaico Rivera. – Brasil, XXXXXXXXXXXX.

51 p. : il. (algumas color.) ; 16x30cm.

Bibliografia – Editora XXXXXXXXXXXX, 9 de abril do XXXXXXXXXXXX.

ISBN:XXXXXXXXXX

1. Samba de gafieira. 2. Samba. 3. Gafieira. 4. Partitura de movimento. 5. Musicalidade. I. Título

Garanta o “download” gratuito da versão digital do livro em <http://XXXXXXXXXX.com>

Limite de responsabilidade e exceção de garantia: O autor ou autores tem feito seu melhor esforço na preparação deste material. Esta edição deve ser proporcionada sem nenhuma modificação. Se distribui gratuitamente com a esperança de que seja útil, porém sem nenhuma garantia expressa ou implícita em relação à exatidão ou completitude do conteúdo.



Esta obra está liberada com uma Licença Creative Commons Atribuição - NãoComercial - SemDerivações 4.0 Internacional. Não é possível usar este arquivo excepto em conformidade com a Licença. Pode obter uma copia da Licença em: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

*Amicum lege feliciter, vivas, gaudeas, floreas in Deo.*

*Fernando*



# Agradecimentos

*Dou muitas graças a Deus*

*Dou muitas graças a Estevam Lawrence por me ajudar a resolver muitas duvidas sobre definições e uso de termos na teoria musical.*

*Dou muitas graças ao Prof. José Henrique de Souza (Henrique Carioca) por suas aulas de samba no pê, e ter-me ensinado exercícios para o desenvolvimento da consciência corporal.*

*Dou muitas graças ao Prof. Jaime Arôxa por me ajudar a resolver algumas duvidas sobre definições e uso de termos na pedagogia para o ensino do samba de gafeira.*

# Patrocínio

Para investir nesta pesquisa e colaborar com o desenvolvimento e crescimento deste projeto:

- você pode comprar uma versão impressa do livro desde o seguinte endereço eletrônico:  
<http://XXXXXXXXXX.com/comprarfisico/>
- Ou pode comprar uma versão digital desde:  
<http://XXXXXXXXXX.com/comprardigital/>

Também pode colaborar com dinheiro em efetivo, desde 20 reais (5 USD), pelos seguintes métodos:

- Método 1
- Método 2

Para verificar a integridade do arquivo da versão digital deste livro, pode seguir as indicações publicadas no sítio oficial do projeto:

<http://XXXXXXXXXX.com>

Se já colaborou com a pesquisa, e se assim o deseja, sintase livre de me mandar um e-mail a [fernando.pujaico.rivera@gmail.com](mailto:fernando.pujaico.rivera@gmail.com), sugerindo abordar um novo assunto ou aprofundar em outro. Sim seu pedido está dentro das minhas capacidades este será agregado sem falta na seguinte edição do livro.

Fernando Pujaico Rivera



# Sumário

## I Derivadas de funções e operadores

<b>1</b>	<b>Funções e operadores notáveis</b>	<b>11</b>
1.1	Operador derivada para $e(\mathbf{x})$ , $\mathbf{f}(\mathbf{x})$ e $\mathbf{G}(\mathbf{x})$	11
1.2	Derivadas notáveis: Gradiente, Matriz Hessiana, Matriz Jacobiana	12
1.3	Serie de Taylor de $d(\mathbf{x})$ , $e(\mathbf{x})$ e $\mathbf{f}(\mathbf{x})$	13
<b>2</b>	<b>Derivada de funções: <math>\mathbb{R}^N \rightarrow \mathbb{R}^M</math></b>	<b>17</b>
2.1	Derivada de $\mathbf{Ax}$	17
2.2	Derivada de $\ \mathbf{Ax}\ ^2$	18
2.3	Derivada de $\ \mathbf{Ax} - \mathbf{b}\ _{\mathbf{C}}^2$	18
2.4	Derivada de $\ \mathbf{f}(\mathbf{x}) - \mathbf{b}\ _{\mathbf{C}}^2$	19
2.5	Derivada de segundo ordem de $\ \mathbf{f}(\mathbf{x}) - \mathbf{b}\ _{\mathbf{C}}^2$	19
2.6	Derivada de $\ \mathbf{f}(\mathbf{x}) - \mathbf{b}\ _{\mathbf{C}}^2 + \alpha \ \mathbf{x} - \mathbf{q}\ _{\mathbf{D}}^2$	19
2.7	Provas dos teoremas	21

## II Problemas de minimização de erro em funções

<b>3</b>	<b>Minimização do erro em funções: <math>\mathbb{R}^N \rightarrow \mathbb{R}^M</math></b>	<b>25</b>
3.1	Minimização de $\ \mathbf{Ax} - \mathbf{b}\ _{\mathbf{C}}^2$	25
3.2	Minimização de $\ \mathbf{f}(\mathbf{x}) - \mathbf{b}\ _{\mathbf{C}}^2$	26
3.3	Minimização de $\ \mathbf{f}(\mathbf{x}) - \mathbf{b}\ _{\mathbf{C}}^2 + \alpha \ \mathbf{x} - \mathbf{q}\ _{\mathbf{D}}^2$	26

3.4	Minimização de $\ f(x) - b\ _C^2 + \alpha \ x\ _D^2$	27
3.5	Minimização de $\ f(x) - b\ _C^2 + \alpha \ x - x_{last}\ _D^2$	27
3.6	Provas dos teoremas	28
<b>4</b>	<b>Minimização do erro em funções: <math>\mathbb{R} \rightarrow \mathbb{R}</math></b>	<b>31</b>
4.1	Minimização de $\ h(x) - a\ ^2$	31
4.2	Minimização de $\ h(x) - a\ ^2 + \alpha \ x - b\ ^2$	33
4.3	Minimização de $\ h(x) - a\ ^2 + \alpha \ x - x_{old}\ ^2$	36
4.4	Provas dos teoremas	37

### III Problemas de achar raízes de funções

<b>5</b>	<b>Raízes de funções: <math>\mathbb{R} \rightarrow \mathbb{R}</math></b>	<b>41</b>
5.1	<b>Raiz de <math>h(x)</math> acotada entre <math>a</math> e <math>b</math></b>	<b>41</b>
5.1.1	Método de la falsa posición	41
5.1.2	Método de bisección	41
5.2	<b>Raiz de <math>h(x)</math> não acotada (Métodos iterativos)</b>	<b>41</b>
5.2.1	Método de las aproximaciones sucesivas	41
5.2.2	Método de la secante	41
5.2.3	Método de Newton	41

### IV Problemas varios

<b>6</b>	<b>Mapeamento</b>	<b>45</b>
6.1	Achar a função de mapeamento $h(x, y) : (x, y) \rightarrow (u, v)$	45

### V Referencias

Bibliography	49
Index	51



# Derivadas de funções e operadores

## 1 Funções e operadores notáveis 11

- 1.1 Operador derivada para  $e(\mathbf{x})$ ,  $\mathbf{f}(\mathbf{x})$  e  $\mathbf{G}(\mathbf{x})$
- 1.2 Derivadas notáveis: Gradiente, Matriz Hessiana, Matriz Jacobiana
- 1.3 Serie de Taylor de  $d(x)$ ,  $e(\mathbf{x})$  e  $\mathbf{f}(\mathbf{x})$

## 2 Derivada de funções: $\mathbb{R}^N \rightarrow \mathbb{R}^M$ 17

- 2.1 Derivada de  $\mathbf{Ax}$
- 2.2 Derivada de  $\|\mathbf{Ax}\|^2$
- 2.3 Derivada de  $\|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{C}}^2$
- 2.4 Derivada de  $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2$
- 2.5 Derivada de segundo ordem de  $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2$
- 2.6 Derivada de  $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha \|\mathbf{x} - \mathbf{q}\|_{\mathbf{D}}^2$
- 2.7 Provas dos teoremas



# 1. Funções e operadores notáveis

## 1.1 Operador derivada para $e(\mathbf{x})$ , $\mathbf{f}(\mathbf{x})$ e $\mathbf{G}(\mathbf{x})$

**Definição 1.1.1** Se  $\mathbf{x} \in \mathbb{R}^N$  é um vetor coluna com elementos  $x_n \in \mathbb{R}$  de modo que  $n \in \mathbb{N}$ ,  $1 \leq n \leq N$ , a função  $e(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$  é um escalar, a função  $\mathbf{f}(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}^M$  é um vetor coluna, e a função  $\mathbf{G}(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}^{M \times L}$  é uma matriz, então definimos que:

$$\frac{\partial e(\mathbf{x})}{\partial \mathbf{x}^T} = \begin{bmatrix} \frac{\partial e(\mathbf{x})}{\partial x_1} & \frac{\partial e(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial e(\mathbf{x})}{\partial x_n} & \cdots & \frac{\partial e(\mathbf{x})}{\partial x_N} \end{bmatrix} = \bigcup_{n=1}^{\rightarrow N} \frac{\partial e(\mathbf{x})}{\partial x_n} \quad (1.1)$$

$$\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^T} = \begin{bmatrix} \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_1} & \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_n} & \cdots & \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_N} \end{bmatrix} = \bigcup_{n=1}^{\rightarrow N} \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_n} \quad (1.2)$$

$$\frac{\partial \mathbf{G}(\mathbf{x})}{\partial \mathbf{x}^T} = \begin{bmatrix} \frac{\partial \mathbf{G}(\mathbf{x})}{\partial x_1} & \frac{\partial \mathbf{G}(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial \mathbf{G}(\mathbf{x})}{\partial x_n} & \cdots & \frac{\partial \mathbf{G}(\mathbf{x})}{\partial x_N} \end{bmatrix} = \bigcup_{n=1}^{\rightarrow N} \frac{\partial \mathbf{G}(\mathbf{x})}{\partial x_n}. \quad (1.3)$$

Assim,  $\frac{\partial e(\mathbf{x})}{\partial \mathbf{x}^T} \in \mathbb{R}^{1 \times N}$ ,  $\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^T} \in \mathbb{R}^{M \times N}$  e  $\frac{\partial \mathbf{G}(\mathbf{x})}{\partial \mathbf{x}^T} \in \mathbb{R}^{M \times (LN)}$ .

**Definição 1.1.2** Se  $\mathbf{x} \in \mathbb{R}^N$  é um vetor coluna com elementos  $x_n \in \mathbb{R}$  de modo que  $n \in \mathbb{N}$ ,  $1 \leq n \leq N$ , a função  $e(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$  é um escalar, a função  $\mathbf{f}(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}^M$  é um vetor coluna,

e a função  $\mathbf{G}(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}^{M \times L}$  é uma matriz, então definimos que:

$$\frac{\partial e(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial e(\mathbf{x})}{\partial x_1} \\ \frac{\partial e(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial e(\mathbf{x})}{\partial x_n} \\ \vdots \\ \frac{\partial e(\mathbf{x})}{\partial x_N} \end{bmatrix} = \text{trans} \left( \frac{\partial e(\mathbf{x})}{\partial \mathbf{x}^T} \right) = \text{vec} \left( \frac{\partial e(\mathbf{x})}{\partial \mathbf{x}^T} \right) = \bigcup_{n=1}^N \frac{\partial e(\mathbf{x})}{\partial x_n} \quad (1.4)$$

$$\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_1} \\ \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_n} \\ \vdots \\ \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_N} \end{bmatrix} = \text{vec} \left( \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^T} \right) = \bigcup_{n=1}^N \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_n} \quad (1.5)$$

$$\frac{\partial \mathbf{G}(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{G}(\mathbf{x})}{\partial x_1} \\ \frac{\partial \mathbf{G}(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial \mathbf{G}(\mathbf{x})}{\partial x_n} \\ \vdots \\ \frac{\partial \mathbf{G}(\mathbf{x})}{\partial x_N} \end{bmatrix} = \bigcup_{n=1}^N \frac{\partial \mathbf{G}(\mathbf{x})}{\partial x_n} \quad (1.6)$$

Assim,  $\frac{\partial e(\mathbf{x})}{\partial \mathbf{x}} \in \mathbb{R}^{N \times 1}$ ,  $\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \in \mathbb{R}^{(MN) \times 1}$  e  $\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} \in \mathbb{R}^{(MN) \times L}$ .

**Corolário 1.1.1 — Igualdade das derivadas cruzadas** Se  $\mathbf{x} \in \mathbb{R}^N$  é um vetor coluna com elementos  $x_n \in \mathbb{R}$  de modo que  $n \in \mathbb{N}$ ,  $1 \leq n \leq N$ , a função  $e(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$  é um escalar, e tendo em consideração as Definições 1.1.1 e 1.1.2 e o “Teorema da igualdade das derivadas cruzadas” [FALTA REFERENCIA!!!]; então é fácil deduzir que:

$$\frac{\partial}{\partial \mathbf{x}} \left( \frac{\partial e(\mathbf{x})}{\partial \mathbf{x}^T} \right) \equiv \frac{\partial}{\partial \mathbf{x}^T} \left( \frac{\partial e(\mathbf{x})}{\partial \mathbf{x}} \right) \quad (1.7)$$

## 1.2 Derivadas notáveis: Gradiente, Matriz Hessiana, Matriz Jacobiana

**Proposição 1.2.1 — Gradiente de  $e(\mathbf{x})$**  Dada uma função  $e : \mathbb{R}^N \rightarrow \mathbb{R}$  com variável  $\mathbf{x} \in \mathbb{R}^N$  como vetor coluna com elementos  $x_n \in \mathbb{R}$  de modo que  $n \in \mathbb{N}$ ,  $1 \leq n \leq N$ , diferenciável em  $\mathbf{x}$ .

$\nabla e(\mathbf{x})$  é chamado gradiente [] **[FALTA REFERENCIA!!!]** de  $e(\mathbf{x})$ , de modo que:

$$\nabla e(\mathbf{x}) \equiv \frac{\partial e(\mathbf{x})}{\partial \mathbf{x}^T} = \begin{bmatrix} \frac{\partial e(\mathbf{x})}{\partial x_1} & \frac{\partial e(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial e(\mathbf{x})}{\partial x_n} & \cdots & \frac{\partial e(\mathbf{x})}{\partial x_N} \end{bmatrix} \quad (1.8)$$

**Proposição 1.2.2 — Matriz Hessiana de  $e(\mathbf{x})$**  Dada uma função  $e : \mathbb{R}^N \rightarrow \mathbb{R}$  com variável  $\mathbf{x} \in \mathbb{R}^N$  como vetor coluna com elementos  $x_n \in \mathbb{R}$  de modo que  $n \in \mathbb{N}$ ,  $1 \leq n \leq N$ , diferenciável em  $\mathbf{x}$ .  $\mathbf{H}(\mathbf{x})$  é chamada matriz Hessiana [] **[FALTA REFERENCIA!!!]** de  $e(\mathbf{x})$ , de modo que:

$$\mathbf{H}(\mathbf{x}) \equiv \frac{\partial}{\partial \mathbf{x}} \left( \frac{\partial e(\mathbf{x})}{\partial \mathbf{x}^T} \right) = \begin{bmatrix} \frac{\partial^2 e(\mathbf{x})}{\partial x_1 \partial x_1} & \frac{\partial^2 e(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 e(\mathbf{x})}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 e(\mathbf{x})}{\partial x_1 \partial x_N} \\ \frac{\partial^2 e(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 e(\mathbf{x})}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 e(\mathbf{x})}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 e(\mathbf{x})}{\partial x_2 \partial x_N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 e(\mathbf{x})}{\partial x_m \partial x_1} & \frac{\partial^2 e(\mathbf{x})}{\partial x_m \partial x_2} & \cdots & \frac{\partial^2 e(\mathbf{x})}{\partial x_m \partial x_n} & \cdots & \frac{\partial^2 e(\mathbf{x})}{\partial x_m \partial x_N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 e(\mathbf{x})}{\partial x_M \partial x_1} & \frac{\partial^2 e(\mathbf{x})}{\partial x_M \partial x_2} & \cdots & \frac{\partial^2 e(\mathbf{x})}{\partial x_M \partial x_n} & \cdots & \frac{\partial^2 e(\mathbf{x})}{\partial x_M \partial x_N} \end{bmatrix} \quad (1.9)$$

**Proposição 1.2.3 — Matriz Jacobiana de  $\mathbf{f}(\mathbf{x})$**  Dado um vetor coluna, como função  $\mathbf{f} : \mathbb{R}^N \rightarrow \mathbb{R}^M$  com variável  $\mathbf{x} \in \mathbb{R}^N$  como vetor coluna com elementos  $x_n \in \mathbb{R}$  de modo que  $n \in \mathbb{N}$ ,  $1 \leq n \leq N$ , diferenciável em  $\mathbf{x}$ .  $\mathbf{J}(\mathbf{x})$  é chamada matriz Jacobiana [] **[FALTA REFERENCIA!!!]** de  $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}) \ f_2(\mathbf{x}) \ \cdots \ f_m(\mathbf{x}) \ \cdots \ f_M(\mathbf{x})]^T$ , de modo que:

$$\mathbf{J}(\mathbf{x}) \equiv \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^T} = \begin{bmatrix} \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_1} & \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_n} & \cdots & \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_N} \end{bmatrix} \quad (1.10)$$

$$\mathbf{J}(\mathbf{x}) \equiv \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_N} \\ \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_2(\mathbf{x})}{\partial x_n} & \cdots & \frac{\partial f_2(\mathbf{x})}{\partial x_N} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \frac{\partial f_m(\mathbf{x})}{\partial x_1} & \frac{\partial f_m(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_N} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \frac{\partial f_M(\mathbf{x})}{\partial x_1} & \frac{\partial f_M(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_M(\mathbf{x})}{\partial x_n} & \cdots & \frac{\partial f_M(\mathbf{x})}{\partial x_N} \end{bmatrix} \quad (1.11)$$

## 1.3 Serie de Taylor de $d(x)$ , $e(\mathbf{x})$ e $\mathbf{f}(\mathbf{x})$

**Proposição 1.3.1 — Serie de Taylor de  $d(x)$**  Dada uma função  $d : \mathbb{R} \rightarrow \mathbb{R}$  com variável  $x \in \mathbb{R}$ ; infinitamente diferenciável em  $a \in \mathbb{R}$ ; esta pode ser expressada mediante uma somatória, em serie de Taylor [] **[FALTA REFERENCIA!!!]** ao redor de  $a$ , como mostra a Eq. (1.12),

$$d(x) = d(a) + \frac{\partial d(x)}{\partial x} \Big|_{x=a} (x-a) + \frac{1}{2} \frac{\partial^2 d(x)}{\partial x^2} \Big|_{x=a} (x-a)^2 + \cdots + \frac{1}{k!} \frac{\partial^k d(x)}{\partial x^k} \Big|_{x=a} (x-a)^k + \cdots \quad (1.12)$$

A equação pode ser escrita de forma mais compacta mediante uma somatória como mostra a Eq. (1.13),

$$d(x) = \sum_{k=0}^{+\infty} \frac{\partial^k d(x)}{\partial x^k} \Big|_{x=a} \frac{(x-a)^k}{k!}. \quad (1.13)$$

É possível ver um exemplo de aproximação da função  $\cos(x)$ , mediante o uso da serie de Taylor da Proposição 1.3.1, truncada ate a derivada de ordem 4, 8 e 10, na Fig. 1.1.

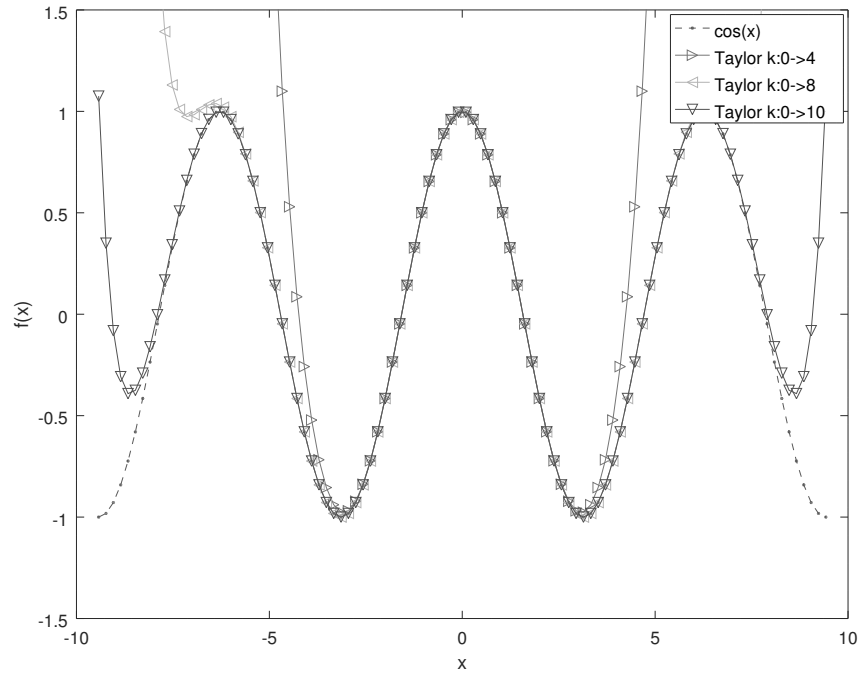


Figura 1.1: Aproximação da função  $\cos(x)$  usando a serie de Taylor truncada.

**Proposição 1.3.2 — Serie de Taylor de  $e(\mathbf{x})$**  Dada uma função  $e : \mathbb{R}^N \rightarrow \mathbb{R}$  com variável  $\mathbf{x} \in \mathbb{R}^N$ , vetor coluna; infinitamente diferenciável em  $\mathbf{a} \in \mathbb{R}^N$ ; esta pode ser expressada mediante uma somatória, em serie de Taylor [] **[FALTA REFERENCIA!!!]** ao redor de  $\mathbf{a}$ , como mostra a Eq. (1.14),

$$e(\mathbf{x}) = \sum_{k_1=0}^{\infty} \cdots \sum_{k_N=0}^{\infty} \left( \frac{\partial^{k_1+\cdots+k_N} e(\mathbf{x})}{\partial x_1^{k_1} \cdots \partial x_N^{k_N}} \right) \Big|_{\mathbf{x}=\mathbf{a}} \frac{(x_1 - a_1)^{k_1} \cdots (x_N - a_N)^{k_N}}{k_1! \cdots k_N!} \quad (1.14)$$

Outra forma alternativa de expressar a função anterior é usando vetores e matrizes, como na Eq. (1.15).

$$e(\mathbf{x}) = e(\mathbf{a}) + \nabla e(\mathbf{a})(\mathbf{x} - \mathbf{a}) + \frac{1}{2!}(\mathbf{x} - \mathbf{a})^T \mathbf{H}(\mathbf{a})(\mathbf{x} - \mathbf{a}) + \cdots \quad (1.15)$$

Onde o vector  $\nabla e(\mathbf{x}) \equiv \frac{\partial e(\mathbf{x})}{\partial \mathbf{x}^T}$  (também chamado gradiente), e a matriz  $\mathbf{H}(\mathbf{x}) \equiv \frac{\partial}{\partial \mathbf{x}} \left( \frac{\partial e(\mathbf{x})}{\partial \mathbf{x}^T} \right)$  (também chamada matriz Hessiana).

**Proposição 1.3.3 — Serie de Taylor de  $\mathbf{f}(\mathbf{x})$**  Dada uma função de contra-domínio vectorial  $\mathbf{f}: \mathbb{R}^N \rightarrow \mathbb{R}^M$ , sendo  $\mathbf{f}$  um vector coluna, com variável  $\mathbf{x} \in \mathbb{R}^N$ , vector coluna; infinitamente diferenciável em  $\mathbf{a} \in \mathbb{R}^N$ ; esta pode ser expressada mediante uma somatória, em serie de Taylor [] **[FALTA REFERENCIA!!!]** ao redor de  $\mathbf{a}$ , como mostra a Eq. (1.16),

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a}) + \mathbf{J}(\mathbf{a})(\mathbf{x} - \mathbf{a}) + \cdots \quad (1.16)$$

Onde a matriz  $\mathbf{J}(\mathbf{x}) \equiv \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^T}$ , também conhecido como matriz Jacobiana de  $\mathbf{f}(\mathbf{x})$ .





## 2. Derivada de funções: $\mathbb{R}^N \rightarrow \mathbb{R}^M$

### 2.1 Derivada de $\mathbf{Ax}$

**Teorema 2.1.1** Se  $\mathbf{x} \in \mathbb{R}^N$  é um vetor coluna com elementos  $x_n$  de modo que  $n \in \mathbb{N}$ ,  $1 \leq n \leq N$ , e  $\mathbf{A} \in \mathbb{R}^{M \times N}$  é uma matriz com elementos  $a_{mn}$  de modo que  $m \in \mathbb{N}$ ,  $1 \leq m \leq M$ , então se cumpre que:

$$\frac{\partial \mathbf{Ax}}{\partial x_n} = a_{:n} \quad (2.1)$$

A demonstração pode ser vista na Prova 2.1.

**Corolário 2.1.2 — Derivada de  $\mathbf{Ax}$  em relação ao vetor  $\mathbf{x}^T$**  Aplicando a Definição 1.1.1 junto ao Teorema 2.1.1, é fácil deduzir que:

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}^T} = \mathbf{A} = [a_{:1} \quad a_{:2} \quad \cdots \quad a_{:N}] \quad (2.2)$$

**Corolário 2.1.3 — Derivada de  $\mathbf{Ax}$  em relação ao vetor  $\mathbf{x}$**  Aplicando a Definição 1.1.2 junto ao Teorema 2.1.1, é fácil deduzir que:

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \text{vec}(\mathbf{A}) = \begin{bmatrix} a_{:1} \\ a_{:2} \\ \vdots \\ a_{:N} \end{bmatrix} \quad (2.3)$$

**Corolário 2.1.4 — Derivada de  $\mathbf{a}^T \mathbf{x}$  em relação ao vetor  $\mathbf{x}^T$**  Aplicando a Definição 1.1.1 junto ao Teorema 2.1.1 e sabendo que  $\mathbf{a}^T$  é um vetor linha, é fácil deduzir que:

$$\frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}^T} = \mathbf{a}^T \quad (2.4)$$

**Corolário 2.1.5 — Derivada de  $\mathbf{a}^T \mathbf{x}$  em relação ao vector  $\mathbf{x}$**  Aplicando a Definição 1.1.2 junto ao Teorema 2.1.1 e sabendo que  $\mathbf{a}^T$  é um vetor linha, é fácil deduzir que:

$$\frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \quad (2.5)$$

## 2.2 Derivada de $\|\mathbf{Ax}\|^2$

**Teorema 2.2.1** Se  $\mathbf{x} \in \mathbb{R}^N$  é um vetor coluna com elementos  $x_n$  de modo que  $n \in \mathbb{N}$ ,  $1 \leq n \leq N$ , e  $\mathbf{A} \in \mathbb{R}^{M \times N}$  é uma matriz com elementos  $a_{mn}$  de modo que  $m \in \mathbb{N}$ ,  $1 \leq m \leq M$ , então se cumpre que:

$$\begin{aligned} \frac{\partial \|\mathbf{Ax}\|^2}{\partial x_n} &= \frac{\partial (\mathbf{Ax})^T (\mathbf{Ax})}{\partial x_n} = 2 (\mathbf{Ax})^T a_{:,n} \\ &= 2 (a_{:,n})^T \mathbf{Ax} \end{aligned} \quad (2.6)$$

A demonstração pode ser vista na Prova 2.2.

**Corolário 2.2.2 — Derivada de  $\|\mathbf{Ax}\|^2$  em relação ao vector  $\mathbf{x}^T$**  Aplicando a Definição 1.1.1 junto ao Teorema 2.2.1, é fácil deduzir que:

$$\frac{\partial \|\mathbf{Ax}\|^2}{\partial \mathbf{x}^T} = 2 (\mathbf{A}^T \mathbf{Ax})^T \quad (2.7)$$

**Corolário 2.2.3 — Derivada de  $\|\mathbf{Ax}\|^2$  em relação ao vector  $\mathbf{x}$**  Aplicando a Definição 1.1.2 junto ao Teorema 2.2.1, é fácil deduzir que:

$$\frac{\partial \|\mathbf{Ax}\|^2}{\partial \mathbf{x}} = 2 \mathbf{A}^T \mathbf{Ax} \quad (2.8)$$

## 2.3 Derivada de $\|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{C}}^2$

**Teorema 2.3.1** Se  $\mathbf{x} \in \mathbb{R}^N$  é um vetor coluna com elementos  $x_n$  de modo que  $n \in \mathbb{N}$ ,  $1 \leq n \leq N$ ,  $\mathbf{b} \in \mathbb{R}^M$  é um vetor coluna com elementos  $b_m$  de modo que  $m \in \mathbb{N}$ ,  $1 \leq m \leq M$ ,  $\mathbf{A} \in \mathbb{R}^{M \times N}$  é uma matriz com elementos  $a_{mn}$ , e  $\mathbf{C} \in \mathbb{R}^{M \times M}$  é uma matriz diagonal, então se cumpre que:

$$\begin{aligned} \frac{\partial \|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{C}}^2}{\partial x_n} &= \frac{\partial (\mathbf{Ax} - \mathbf{b})^T \mathbf{C} (\mathbf{Ax} - \mathbf{b})}{\partial x_n} = 2 (\mathbf{Ax} - \mathbf{b})^T \mathbf{C} a_{:,n} \\ &= 2 (a_{:,n})^T \mathbf{C} (\mathbf{Ax} - \mathbf{b}) \end{aligned} \quad (2.9)$$

A demonstração pode ser vista na Prova 2.3.

**Corolário 2.3.2 — Derivada de  $\|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{C}}^2$  em relação ao vector  $\mathbf{x}^T$**  Aplicando a Definição 1.1.1 junto ao Teorema 2.3.1, é fácil deduzir que:

$$\frac{\partial \|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{C}}^2}{\partial \mathbf{x}^T} = 2 (\mathbf{Ax} - \mathbf{b})^T \mathbf{C} \mathbf{A} \quad (2.10)$$

**Corolário 2.3.3 — Derivada de  $\|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{C}}^2$  em relação ao vector  $\mathbf{x}$**  Aplicando a Definição 1.1.2 junto ao Teorema 2.3.1, é fácil deduzir que:

$$\frac{\partial \|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{C}}^2}{\partial \mathbf{x}} = 2 \mathbf{A}^T \mathbf{C} (\mathbf{Ax} - \mathbf{b}) \quad (2.11)$$

## 2.4 Derivada de $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2$

**Teorema 2.4.1 — Valor exato** Se  $\mathbf{x} \in \mathbb{R}^N$  é um vetor coluna,  $\mathbf{b} \in \mathbb{R}^M$  é um vetor coluna,  $\mathbf{f} : \mathbb{R}^N \rightarrow \mathbb{R}^M$  é uma função de valor vectorial, e  $\mathbf{C} \in \mathbb{R}^{M \times M}$  é uma matriz diagonal, então se cumpre que:

$$\frac{\partial \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2}{\partial \mathbf{x}} = 2\mathbf{J}(\mathbf{x})^T \mathbf{C} [\mathbf{f}(\mathbf{x}) - \mathbf{b}], \quad (2.12)$$

onde  $\mathbf{J}(\mathbf{x})$  é a matriz Jacobiana [] de  $\mathbf{f}(\mathbf{x})$ .

A demonstração pode ser vista na Prova 2.4.

**Teorema 2.4.2 — Valor aproximado** Se  $\mathbf{x} \in \mathbb{R}^N$  é um vetor coluna,  $\mathbf{b} \in \mathbb{R}^M$  é um vetor coluna,  $\mathbf{f} : \mathbb{R}^N \rightarrow \mathbb{R}^M$  é uma função de valor vectorial, e  $\mathbf{C} \in \mathbb{R}^{M \times M}$  é uma matriz diagonal, então se cumpre que:

$$\frac{\partial \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2}{\partial \mathbf{x}} \approx 2\mathbf{J}(\mathbf{p})^T \mathbf{C} [\mathbf{J}(\mathbf{p}) (\mathbf{x} - \mathbf{p}) - (\mathbf{b} - \mathbf{f}(\mathbf{p}))] \quad (2.13)$$

Onde é considerada a aproximação  $\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{p}) + \mathbf{J}(\mathbf{p}) (\mathbf{x} - \mathbf{p})$ , usando a serie de Taylor [] para funções multivariáveis. Sendo  $\mathbf{p}$  um ponto fixo no domínio de  $\mathbf{f}(\mathbf{x})$ , ao redor do qual é feita aproximação da função  $\mathbf{f}(\mathbf{x})$ , e  $\mathbf{J}(\mathbf{p})$  é a matriz Jacobiana [] de  $\mathbf{f}(\mathbf{x})$  avaliado no ponto  $\mathbf{p}$ .

A demonstração pode ser vista na Prova 2.5.

## 2.5 Derivada de segundo ordem de $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2$

**Teorema 2.5.1 — Valor exato** Se  $\mathbf{x} \in \mathbb{R}^N$  é um vetor coluna,  $\mathbf{b} \in \mathbb{R}^M$  é um vetor coluna,  $\mathbf{f} : \mathbb{R}^N \rightarrow \mathbb{R}^M$  é uma função de valor vectorial,  $\mathbf{C} \in \mathbb{R}^{M \times M}$  é uma matriz diagonal, e definida a função  $e(\mathbf{x})$ ,

$$e(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2. \quad (2.14)$$

Então a matriz Hessiana []  $\mathbf{H}(\mathbf{x})$  de  $e(\mathbf{x})$  é igual a:

$$\mathbf{H}(\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}^T} \left( \frac{\partial e(\mathbf{x})}{\partial \mathbf{x}} \right) = 2\mathbf{B}(\mathbf{x}) + 2\mathbf{J}(\mathbf{x})^T \mathbf{C} \mathbf{J}(\mathbf{x}), \quad (2.15)$$

onde

$$\mathbf{B}(\mathbf{x}) = \bigcup_{n=1}^{\rightarrow N} \left\{ \frac{\partial \mathbf{J}(\mathbf{x})^T}{\partial x_n} \mathbf{C} (\mathbf{f}(\mathbf{x}) - \mathbf{b}) \right\}, \quad (2.16)$$

A demonstração pode ser vista na Prova 2.6.

## 2.6 Derivada de $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha \|\mathbf{x} - \mathbf{q}\|_{\mathbf{D}}^2$

**Teorema 2.6.1 — Valor aproximado** Se  $\mathbf{x} \in \mathbb{R}^N$  é um vetor coluna,  $\mathbf{b} \in \mathbb{R}^M$  é um vetor coluna,  $\mathbf{q} \in \mathbb{R}^N$  é um vetor coluna,  $\mathbf{f} : \mathbb{R}^N \rightarrow \mathbb{R}^M$  é uma função de valor vectorial,  $\mathbf{C} \in \mathbb{R}^{M \times M}$  é uma matriz diagonal, e  $\mathbf{D} \in \mathbb{R}^{N \times N}$  é uma matriz diagonal, então se cumpre que:

$$\frac{\partial (||\mathbf{f}(\mathbf{x}) - \mathbf{b}||_{\mathbf{C}}^2 + \alpha ||\mathbf{x} - \mathbf{q}||_{\mathbf{D}}^2)}{\partial \mathbf{x}} \approx 2\mathbf{J}(\mathbf{p})^T \mathbf{C} \{\mathbf{J}(\mathbf{p}) [\mathbf{x} - \mathbf{p}] - [\mathbf{b} - \mathbf{f}(\mathbf{p})]\} + 2\alpha \mathbf{D} [\mathbf{x} - \mathbf{q}] \quad (2.17)$$

Onde é considerada a aproximação  $\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{p}) + \mathbf{J}(\mathbf{p}) (\mathbf{x} - \mathbf{p})$ , usando a serie de Taylor [] para funções multivariáveis. Sendo  $\mathbf{p}$  um ponto fixo no domínio de  $\mathbf{f}(\mathbf{x})$ , ao redor do qual é feita aproximação da função  $\mathbf{f}(\mathbf{x})$ , e  $\mathbf{J}(\mathbf{p})$  é a matriz Jacobiana [] de  $\mathbf{f}(\mathbf{x})$  avaliado no ponto  $\mathbf{p}$ .

A demonstração pode ser vista na Prova 2.7.

## 2.7 Provas dos teoremas

**Prova 2.1 — Prova do Teorema 2.1.1** Dados, uma matriz  $\mathbf{A} = [a_{:1} \ a_{:2} \ \dots \ a_{:n} \ \dots \ a_{:N}]$  e um vetor coluna  $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n \ \dots \ x_N]^T$ , podemos expressar que:

$$\frac{\partial \mathbf{Ax}}{\partial x_n} = \frac{\partial \mathbf{A}}{\partial x_n} \mathbf{x} + \mathbf{A} \frac{\partial \mathbf{x}}{\partial x_n} = \mathbf{A} \frac{\partial \mathbf{x}}{\partial x_n} \quad (2.18)$$

Sabendo que  $\frac{\partial \mathbf{x}}{\partial x_n}$  é igual a um vetor com um 1 na posição  $n$  e 0 em qualquer outra posição, obtemos que

$$\frac{\partial \mathbf{Ax}}{\partial x_n} = [a_{:1} \ a_{:2} \ \dots \ a_{:n} \ \dots \ a_{:N}] \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = a_{:n} \quad (2.19)$$

**Prova 2.2 — Prova do Teorema 2.2.1** Dados, uma matriz  $\mathbf{A} = [a_{:1} \ a_{:2} \ \dots \ a_{:n} \ \dots \ a_{:N}]$  e um vetor coluna  $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n \ \dots \ x_N]^T$ , podemos expressar que:

$$\frac{\partial \|\mathbf{Ax}\|^2}{\partial x_n} = \frac{\partial (\mathbf{Ax})^T (\mathbf{Ax})}{\partial x_n} = \left( \frac{\partial \mathbf{Ax}}{\partial x_n} \right)^T (\mathbf{Ax}) + (\mathbf{Ax})^T \frac{\partial \mathbf{Ax}}{\partial x_n} \quad (2.20)$$

Pelo visto no Teorema 2.1.1 podemos substituir valores na Eq. (2.20) e obter:

$$\frac{\partial \|\mathbf{Ax}\|^2}{\partial x_n} = (a_{:n})^T \mathbf{Ax} + (\mathbf{Ax})^T a_{:n} \quad (2.21)$$

Como cada um dos somandos da equação anterior é um escalar, podemos aplicar o operador transposta (T) sobre qualquer somando sem alterar o resultado; de modo que temos duas possíveis forma de expressar a solução:

$$\begin{aligned} \frac{\partial \|\mathbf{Ax}\|^2}{\partial x_n} &= 2 (\mathbf{Ax})^T a_{:n} \\ &= 2 (a_{:n})^T \mathbf{Ax} \end{aligned} \quad (2.22)$$

**Prova 2.3 — Prova do Teorema 2.3.1** Dados, uma matriz  $\mathbf{A} = [a_{:1} \ a_{:2} \ \dots \ a_{:n} \ \dots \ a_{:N}]$ , uma matriz diagonal  $\mathbf{C} \in \mathbb{R}^{M \times M}$ , um vetor coluna  $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n \ \dots \ x_N]^T$ , e um vetor coluna  $\mathbf{b} \in \mathbb{R}^M$ , podemos expressar que:

$$\frac{\partial \|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{C}}^2}{\partial x_n} = \frac{\partial (\mathbf{Ax} - \mathbf{b})^T \mathbf{C} (\mathbf{Ax} - \mathbf{b})}{\partial x_n} = \left( \frac{\partial \mathbf{Ax}}{\partial x_n} \right)^T \mathbf{C} (\mathbf{Ax} - \mathbf{b}) + (\mathbf{Ax} - \mathbf{b})^T \mathbf{C} \left( \frac{\partial \mathbf{Ax}}{\partial x_n} \right) \quad (2.23)$$

Pelo visto no Teorema 2.1.1, podemos substituir valores na Eq. (2.23) e obter:

$$\frac{\partial \|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{C}}^2}{\partial x_n} = (a_{:n})^T \mathbf{C} (\mathbf{Ax} - \mathbf{b}) + (\mathbf{Ax} - \mathbf{b})^T \mathbf{C} a_{:n} \quad (2.24)$$

Como cada um dos somandos da equação anterior é um escalar, podemos aplicar o operador transposta (T) sobre qualquer somando sem alterar o resultado; de modo que temos duas possíveis forma de expressar a solução:

$$\begin{aligned} \frac{\partial \|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{C}}^2}{\partial x_n} &= 2 (\mathbf{Ax} - \mathbf{b})^T \mathbf{C} a_{:n} \\ &= 2 (a_{:n})^T \mathbf{C} (\mathbf{Ax} - \mathbf{b}) \end{aligned} \quad (2.25)$$

**Prova 2.4 — Prova do Teorema 2.4.1** Dados, uma função  $\mathbf{f}: \mathbb{R}^N \rightarrow \mathbb{R}^M$ , uma matriz diagonal  $\mathbf{C} \in \mathbb{R}^{M \times M}$ , um vetor coluna  $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n \ \dots \ x_N]^T$ , e um vetor coluna  $\mathbf{b} \in \mathbb{R}^M$ ; podemos expressar que:

$$\begin{aligned} \frac{\partial \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2}{\partial x_n} &= \frac{\partial (\mathbf{f}(\mathbf{x}) - \mathbf{b})^T \mathbf{C} (\mathbf{f}(\mathbf{x}) - \mathbf{b})}{\partial x_n} \\ &= \left( \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_n} \right)^T \mathbf{C} (\mathbf{f}(\mathbf{x}) - \mathbf{b}) + (\mathbf{f}(\mathbf{x}) - \mathbf{b})^T \mathbf{C} \left( \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_n} \right). \\ &= 2 \left( \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_n} \right)^T \mathbf{C} (\mathbf{f}(\mathbf{x}) - \mathbf{b}) \end{aligned} \quad (2.26)$$

Assim, usando a Definição 1.1.2 junto com a Eq. (2.26) nos obtemos:

$$\frac{\partial \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2}{\partial \mathbf{x}} = 2\mathbf{J}(\mathbf{x})^T \mathbf{C} (\mathbf{f}(\mathbf{x}) - \mathbf{b}) \quad (2.27)$$

**Prova 2.5 — Prova do Teorema 2.4.2** Dados, uma função  $\mathbf{f}: \mathbb{R}^N \rightarrow \mathbb{R}^M$ , uma matriz diagonal  $\mathbf{C} \in \mathbb{R}^{M \times M}$ , um vetor coluna  $\mathbf{x} \in \mathbb{R}^N$ , um vetor coluna  $\mathbf{b} \in \mathbb{R}^M$ , e considerando a aproximação  $\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{p}) + \mathbf{J}(\mathbf{p})(\mathbf{x} - \mathbf{p})$ , usando a serie de Taylor para funções multivariáveis; podemos expressar que:

$$\frac{\partial \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2}{\partial \mathbf{x}} \approx \frac{\partial \|\mathbf{J}(\mathbf{p})\mathbf{x} - [\mathbf{J}(\mathbf{p})\mathbf{p} + \mathbf{b} - \mathbf{f}(\mathbf{p})]\|_{\mathbf{C}}^2}{\partial \mathbf{x}} \quad (2.28)$$

Pelo visto no Corolário 2.3.3, podemos substituir os valores,  $\mathbf{J}(\mathbf{p})$  e  $[\mathbf{J}(\mathbf{p})\mathbf{p} + \mathbf{b} - \mathbf{f}(\mathbf{p})]$ , da Eq. (2.28), nas variáveis  $\mathbf{A}$  e  $\mathbf{b}$  do Corolário 2.3.3, respectivamente. Assim obtemos:

$$\frac{\partial \|\mathbf{J}(\mathbf{p})\mathbf{x} - [\mathbf{J}(\mathbf{p})\mathbf{p} + \mathbf{b} - \mathbf{f}(\mathbf{p})]\|_{\mathbf{C}}^2}{\partial \mathbf{x}} = 2\mathbf{J}(\mathbf{p})^T \mathbf{C} (\mathbf{J}(\mathbf{p})\mathbf{x} - \mathbf{J}(\mathbf{p})\mathbf{p} - \mathbf{b} + \mathbf{f}(\mathbf{p})) \quad (2.29)$$

**Prova 2.6 — Prova do Teorema 2.5.1** Dados, uma função  $\mathbf{f}: \mathbb{R}^N \rightarrow \mathbb{R}^M$ , uma matriz diagonal  $\mathbf{C} \in \mathbb{R}^{M \times M}$ , um vetor coluna  $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n \ \dots \ x_N]^T$ , e um vetor coluna  $\mathbf{b} \in \mathbb{R}^M$ ; podemos expressar que:

$$\begin{aligned} \frac{\partial}{\partial x_n} \left( \frac{\partial \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2}{\partial \mathbf{x}} \right) &= \frac{\partial 2\mathbf{J}(\mathbf{x})^T \mathbf{C} (\mathbf{f}(\mathbf{x}) - \mathbf{b})}{\partial x_n} \\ &= 2 \frac{\partial \mathbf{J}(\mathbf{x})^T}{\partial x_n} \mathbf{C} (\mathbf{f}(\mathbf{x}) - \mathbf{b}) + 2\mathbf{J}(\mathbf{x})^T \mathbf{C} \frac{\partial (\mathbf{f}(\mathbf{x}) - \mathbf{b})}{\partial x_n}. \\ &= 2 \frac{\partial \mathbf{J}(\mathbf{x})^T}{\partial x_n} \mathbf{C} (\mathbf{f}(\mathbf{x}) - \mathbf{b}) + 2\mathbf{J}(\mathbf{x})^T \mathbf{C} \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_n} \end{aligned} \quad (2.30)$$

Assim, usando a Definição 1.1.1 junto com a Eq. (2.30) nos obtemos:

$$\frac{\partial}{\partial \mathbf{x}^T} \left( \frac{\partial \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2}{\partial \mathbf{x}} \right) = 2 \bigcup_{n=1}^N \left\{ \frac{\partial \mathbf{J}(\mathbf{x})^T}{\partial x_n} \mathbf{C} (\mathbf{f}(\mathbf{x}) - \mathbf{b}) \right\} + 2\mathbf{J}(\mathbf{x})^T \mathbf{C} \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^T} \quad (2.31)$$

**Prova 2.7 — Prova do Teorema 2.6.1** Dados, uma função  $\mathbf{f}: \mathbb{R}^N \rightarrow \mathbb{R}^M$ , uma matriz diagonal  $\mathbf{C} \in \mathbb{R}^{M \times M}$ , uma matriz diagonal  $\mathbf{D} \in \mathbb{R}^{N \times N}$ , um vetor coluna  $\mathbf{x} \in \mathbb{R}^N$ , um vetor coluna  $\mathbf{b} \in \mathbb{R}^M$ , e um vetor coluna  $\mathbf{q} \in \mathbb{R}^N$ , podemos expressar que:

$$\frac{\partial \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha \|\mathbf{x} - \mathbf{q}\|_{\mathbf{D}}^2}{\partial \mathbf{x}} = \frac{\partial \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2}{\partial \mathbf{x}} + \alpha \frac{\partial \|\mathbf{x} - \mathbf{q}\|_{\mathbf{D}}^2}{\partial \mathbf{x}}. \quad (2.32)$$

Usando o Corolário 2.3.3 na Eq. (2.32) obtemos:

$$\frac{\partial \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha \|\mathbf{x} - \mathbf{q}\|_{\mathbf{D}}^2}{\partial \mathbf{x}} = \frac{\partial \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2}{\partial \mathbf{x}} + \alpha 2\mathbf{D}(\mathbf{x} - \mathbf{q}). \quad (2.33)$$

Usando o Teorema 2.4.2 na Eq. (2.33) obtemos:

$$\frac{\partial \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha \|\mathbf{x} - \mathbf{q}\|_{\mathbf{D}}^2}{\partial \mathbf{x}} \approx 2\mathbf{J}(\mathbf{p})^T \mathbf{C} [\mathbf{J}(\mathbf{p})(\mathbf{x} - \mathbf{p}) - (\mathbf{b} - \mathbf{f}(\mathbf{p}))] + 2\alpha \mathbf{D}(\mathbf{x} - \mathbf{q}). \quad (2.34)$$

# Problemas de minimização de erro em funções

## 3 Minimização do erro em funções: $\mathbb{R}^N \rightarrow \mathbb{R}^M$ ..... 25

- 3.1 Minimização de  $\|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{C}}^2$
- 3.2 Minimização de  $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2$
- 3.3 Minimização de  $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha \|\mathbf{x} - \mathbf{q}\|_{\mathbf{D}}^2$
- 3.4 Minimização de  $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha \|\mathbf{x}\|_{\mathbf{D}}^2$
- 3.5 Minimização de  $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha \|\mathbf{x} - \mathbf{x}_{last}\|_{\mathbf{D}}^2$
- 3.6 Provas dos teoremas

## 4 Minimização do erro em funções: $\mathbb{R} \rightarrow \mathbb{R}$ ..... 31

- 4.1 Minimização de  $|h(x) - a|^2$
- 4.2 Minimização de  $|h(x) - a|^2 + \alpha |x - b|^2$
- 4.3 Minimização de  $|h(x) - a|^2 + \alpha |x - x_{old}|^2$
- 4.4 Provas dos teoremas





### 3. Minimização do erro em funções:

$$\mathbb{R}^N \rightarrow \mathbb{R}^M$$

**R** Palavras chave: Pseudo-inversa de Moore-Penrose, regularização de Tikhonov, problema inverso, minimização do erro quadrático.

#### 3.1 Minimização de $\|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{C}}^2$

**Teorema 3.1.1** Dados, um vetor coluna  $\mathbf{x} \in \mathbb{R}^N$ , um vetor coluna  $\mathbf{b} \in \mathbb{R}^M$ , uma matriz  $\mathbf{A} \in \mathbb{R}^{M \times N}$ , uma matriz diagonal  $\mathbf{C} \in \mathbb{R}^{M \times M}$ , e definida a Eq. (3.1),

$$e(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{C}}^2 \quad (3.1)$$

$$= (\mathbf{Ax} - \mathbf{b})^T \mathbf{C} (\mathbf{Ax} - \mathbf{b}). \quad (3.2)$$

Se desejamos ter o valor  $\hat{\mathbf{x}}$  que minimiza o escalar  $e(\hat{\mathbf{x}})$ , devemos usar<sup>a</sup> a Eq. (3.3),

$$\hat{\mathbf{x}} = [\mathbf{A}^T \mathbf{C} \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{C} \mathbf{b}. \quad (3.3)$$

Assim, o mínimo existe só sim  $\mathbf{A}^T \mathbf{C} \mathbf{A}$  tem inversa.

<sup>a</sup>A demonstração pode ser vista na Prova 3.1.

### 3.2 Minimização de $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2$

**Teorema 3.2.1 — Solução iterativa** Dados, um vetor coluna  $\mathbf{x} \in \mathbb{R}^N$ , um vetor coluna  $\mathbf{b} \in \mathbb{R}^M$ , uma função  $\mathbf{f}: \mathbb{R}^N \rightarrow \mathbb{R}^M$ , uma matriz diagonal  $\mathbf{C} \in \mathbb{R}^{M \times M}$ , e definida a Eq. (3.4),

$$e(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 \quad (3.4)$$

$$= (\mathbf{f}(\mathbf{x}) - \mathbf{b})^T \mathbf{C} (\mathbf{f}(\mathbf{x}) - \mathbf{b}). \quad (3.5)$$

Se desejamos ter o valor  $\hat{\mathbf{x}}$  que minimiza o escalar  $e(\hat{\mathbf{x}})$ , este valor pode ser achado usando iterativamente a Eq. (3.6),

$$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + [\mathbf{J}(\mathbf{x}_k)^T \mathbf{C} \mathbf{J}(\mathbf{x}_k)]^{-1} \mathbf{J}(\mathbf{x}_k)^T \mathbf{C} [\mathbf{b} - \mathbf{f}(\mathbf{x}_k)] \quad (3.6)$$

Onde  $\mathbf{J}(\mathbf{x})$  é a matriz Jacobiana [] de  $\mathbf{f}(\mathbf{x})$ . A busca iterativa é considerada falida quando  $\mathbf{J}(\mathbf{x}_k)^T \mathbf{C} \mathbf{J}(\mathbf{x}_k)$  não tem inversa.

Assim,  $\hat{\mathbf{x}}$  pode ser achado iniciando a Eq. (3.6) desde um  $\mathbf{x}_0$  qualquer, ate que  $\mathbf{x}_k$  seja muito próximo a  $\mathbf{x}_{k+1}$ , onde se declara que  $\hat{\mathbf{x}} \approx \mathbf{x}_{k+1}$ ; porem deve ser corroborado que esse ponto tratasse de um máximo ou mínimo usando algum método, por exemplo estudando o comportamento de  $e(\mathbf{x})$  ou analisando a matriz hessiana de  $e(\mathbf{x})$  avaliada em  $\hat{\mathbf{x}}$ .

A demonstração pode ser vista na Prova 3.2.

### 3.3 Minimização de $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha \|\mathbf{x} - \mathbf{q}\|_{\mathbf{D}}^2$

**Teorema 3.3.1 — Solução iterativa** Sabendo que,  $\mathbf{x}$  e  $\mathbf{q}$  são vetores coluna com  $N$  elementos, sendo  $\mathbf{q}$  uma constante,  $\mathbf{f}(\mathbf{x})$  e  $\mathbf{b}$  são vetores coluna de  $M$  elementos, sendo  $\mathbf{b}$  uma constante,  $\mathbf{C}$  uma matriz diagonal de  $M \times M$  e  $\mathbf{D}$  uma matriz diagonal de  $N \times N$ . Sim se deseja achar o valor  $\hat{\mathbf{x}}$  que minimiza o valor de  $e(\mathbf{x})$ , visto na Eq. (3.7),

$$e(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha \|\mathbf{x} - \mathbf{q}\|_{\mathbf{D}}^2 \quad (3.7)$$

$$= (\mathbf{f}(\mathbf{x}) - \mathbf{b})^T \mathbf{C} (\mathbf{f}(\mathbf{x}) - \mathbf{b}) + \alpha (\mathbf{x} - \mathbf{q})^T \mathbf{D} (\mathbf{x} - \mathbf{q}), \quad (3.8)$$

devemos usar a seguinte equação iterativa,

$$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + [\mathbf{J}(\mathbf{x}_k)^T \mathbf{C} \mathbf{J}(\mathbf{x}_k) + \alpha \mathbf{D}]^{-1} (\mathbf{J}(\mathbf{x}_k)^T \mathbf{C} [\mathbf{b} - \mathbf{f}(\mathbf{x}_k)] - \alpha \mathbf{D} [\mathbf{x}_k - \mathbf{q}]) \quad (3.9)$$

Onde  $\mathbf{J}(\mathbf{x})$  é a matriz Jacobiana [] de  $\mathbf{f}(\mathbf{x})$ . A busca iterativa é considerada falida quando  $\mathbf{J}(\mathbf{x}_k)^T \mathbf{C} \mathbf{J}(\mathbf{x}_k) + \alpha \mathbf{D}$  não tem inversa.

Assim,  $\hat{\mathbf{x}}$  pode ser achado iniciando a Eq. (3.9) desde um  $\mathbf{x}_0$  qualquer, ate que  $\mathbf{x}_k$  seja muito próximo a  $\mathbf{x}_{k+1}$ , onde se declara que  $\hat{\mathbf{x}} \approx \mathbf{x}_{k+1}$ ; porem deve ser corroborado que esse ponto tratasse de um máximo ou mínimo usando algum método, por exemplo estudando o comportamento de  $e(\mathbf{x})$  ou analisando a matriz hessiana de  $e(\mathbf{x})$  avaliada em  $\hat{\mathbf{x}}$ .

A demonstração pode ser vista na Prova 3.3.

### 3.4 Minimização de $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha\|\mathbf{x}\|_{\mathbf{D}}^2$

**Teorema 3.4.1 — Solução iterativa** Sabendo que,  $\mathbf{x}$  é um vetor com  $N$  elementos,  $\mathbf{f}(\mathbf{x})$  e  $\mathbf{b}$  são vetores coluna de  $M$  elementos, sendo  $\mathbf{b}$  uma constante,  $\mathbf{C}$  uma matriz diagonal de  $M \times M$  e  $\mathbf{D}$  uma matriz diagonal de  $N \times N$ . Sim se deseja achar o valor  $\hat{\mathbf{x}}$  que minimiza o valor de  $e(\mathbf{x})$ , visto na Eq. (3.10),

$$e(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha\|\mathbf{x}\|_{\mathbf{D}}^2 \quad (3.10)$$

$$= (\mathbf{f}(\mathbf{x}) - \mathbf{b})^T \mathbf{C} (\mathbf{f}(\mathbf{x}) - \mathbf{b}) + \alpha \mathbf{x}^T \mathbf{D} \mathbf{x}, \quad (3.11)$$

devemos usar a seguinte equação iterativa,

$$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + [\mathbf{J}(\mathbf{x}_k)^T \mathbf{C} \mathbf{J}(\mathbf{x}_k) + \alpha \mathbf{D}]^{-1} [\mathbf{J}(\mathbf{x}_k)^T \mathbf{C} \{\mathbf{b} - \mathbf{f}(\mathbf{x}_k)\} - \alpha \mathbf{D} \mathbf{x}_k] \quad (3.12)$$

Onde  $\mathbf{J}(\mathbf{x})$  é a matriz Jacobiana [] de  $\mathbf{f}(\mathbf{x})$ . A busca iterativa é considerada falida quando  $\mathbf{J}(\mathbf{x}_k)^T \mathbf{C} \mathbf{J}(\mathbf{x}_k) + \alpha \mathbf{D}$  não tem inversa.

Assim,  $\hat{\mathbf{x}}$  pode ser achado iniciando a Eq. (3.12) desde um  $\mathbf{x}_0$  qualquer, ate que  $\mathbf{x}_k$  seja muito próximo a  $\mathbf{x}_{k+1}$ , onde se declara que  $\hat{\mathbf{x}} \approx \mathbf{x}_{k+1}$ ; porem deve ser corroborado que esse ponto tratasse de um máximo ou mínimo usando algum método, por exemplo estudando o comportamento de  $e(\mathbf{x})$  ou analisando a matriz hessiana de  $e(\mathbf{x})$  avaliada em  $\hat{\mathbf{x}}$ .

A demonstração pode ser vista na Prova 3.4.

### 3.5 Minimização de $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha\|\mathbf{x} - \mathbf{x}_{last}\|_{\mathbf{D}}^2$

**Teorema 3.5.1 — Solução iterativa** Sabendo que,  $\mathbf{x}$  e  $\mathbf{x}_{last}$  são vetores coluna com  $N$  elementos, sendo  $\mathbf{x}_{last}$  uma constante que representa o ultimo valor de  $\mathbf{x}$  obtido iterativamente,  $\mathbf{f}(\mathbf{x})$  e  $\mathbf{b}$  são vetores coluna de  $M$  elementos, sendo  $\mathbf{b}$  uma constante,  $\mathbf{C}$  uma matriz diagonal de  $M \times M$  e  $\mathbf{D}$  uma matriz diagonal de  $N \times N$ . Sim se deseja achar o valor  $\hat{\mathbf{x}}$  que minimiza o valor de  $e(\mathbf{x})$ , visto na Eq. (3.13),

$$e(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha\|\mathbf{x} - \mathbf{x}_{last}\|_{\mathbf{D}}^2, \quad (3.13)$$

devemos usar a seguinte equação iterativa,

$$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + [\mathbf{J}(\mathbf{x}_k)^T \mathbf{C} \mathbf{J}(\mathbf{x}_k) + \alpha \mathbf{D}]^{-1} [\mathbf{J}(\mathbf{x}_k)^T \mathbf{C} \{\mathbf{b} - \mathbf{f}(\mathbf{x}_k)\}] \quad (3.14)$$

Onde  $\mathbf{J}(\mathbf{x})$  é a matriz Jacobiana [] de  $\mathbf{f}(\mathbf{x})$ . A busca iterativa é considerada falida quando  $\mathbf{J}(\mathbf{x}_k)^T \mathbf{C} \mathbf{J}(\mathbf{x}_k) + \alpha \mathbf{D}$  não tem inversa.

Assim,  $\hat{\mathbf{x}}$  pode ser achado iniciando a Eq. (3.14) desde um  $\mathbf{x}_0$  qualquer, ate que  $\mathbf{x}_k$  seja muito próximo a  $\mathbf{x}_{k+1}$ , onde se declara que  $\hat{\mathbf{x}} \approx \mathbf{x}_{k+1}$ ; porem deve ser corroborado que esse ponto tratasse de um máximo ou mínimo usando algum método, por exemplo estudando o comportamento de  $e(\mathbf{x})$  ou analisando a matriz hessiana de  $e(\mathbf{x})$  avaliada em  $\hat{\mathbf{x}}$ .

A demonstração pode ser vista na Prova 3.5.

### 3.6 Provas dos teoremas

**Prova 3.1 — Prova do Teorema 3.1.1** Dados, um vetor coluna  $\mathbf{x} \in \mathbb{R}^N$ , um vetor coluna  $\mathbf{b} \in \mathbb{R}^M$ , uma matriz  $\mathbf{A} \in \mathbb{R}^{M \times N}$ , uma matriz diagonal  $\mathbf{C} \in \mathbb{R}^{M \times M}$ , e definida a Eq. (3.15),

$$e(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{C}}^2. \quad (3.15)$$

Para achar o valor  $\hat{\mathbf{x}}$  que gere o menor valor de  $e(\hat{\mathbf{x}})$ , é aplicado o critério que um ponto de inflexão pode ser achado quando  $\frac{\partial e(\hat{\mathbf{x}})}{\partial \mathbf{x}} = [0 \ 0 \ \dots \ 0]^T$ . Assim, usando o Corolário 2.3.3 podemos rescrever esta igualdade como a Eq. (3.16),

$$2\mathbf{A}^T \mathbf{C} (\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}) = [0 \ 0 \ \dots \ 0]^T, \quad (3.16)$$

de modo que pode ser obtido:

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{C} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{C} \mathbf{b}. \quad (3.17)$$

Dado que a solução é única e a função  $e(\mathbf{x})$  é sempre positiva, então o valor de  $\hat{\mathbf{x}}$  é um mínimo.

**Prova 3.2 — Prova do Teorema 3.2.1** Dados, um vetor coluna  $\mathbf{x} \in \mathbb{R}^N$ , um vetor coluna  $\mathbf{b} \in \mathbb{R}^M$ , uma função  $\mathbf{f}: \mathbb{R}^N \rightarrow \mathbb{R}^M$ , uma matriz diagonal  $\mathbf{C} \in \mathbb{R}^{M \times M}$ , e definida a Eq. (3.18),

$$e(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2. \quad (3.18)$$

Para achar o valor  $\hat{\mathbf{x}}$  que gere o menor valor de  $e(\hat{\mathbf{x}})$ , é aplicado o critério que um ponto de inflexão pode ser achado quando  $\frac{\partial e(\hat{\mathbf{x}})}{\partial \mathbf{x}} = [0 \ 0 \ \dots \ 0]^T$ . Assim, usando o Teorema 2.4.2 podemos rescrever esta igualdade como a Eq. (3.19),

$$2\mathbf{J}(\mathbf{p})^T \mathbf{C} [\mathbf{J}(\mathbf{p}) (\hat{\mathbf{x}} - \mathbf{p}) - (\mathbf{b} - \mathbf{f}(\mathbf{p}))] \approx \frac{\partial e(\hat{\mathbf{x}})}{\partial \mathbf{x}} = [0 \ 0 \ \dots \ 0]^T, \quad (3.19)$$

de modo que pode ser aproximado  $\hat{\mathbf{x}}$  como:

$$\hat{\mathbf{x}} \approx \mathbf{p} + [\mathbf{J}(\mathbf{p})^T \mathbf{C} \mathbf{J}(\mathbf{p})]^{-1} \mathbf{J}(\mathbf{p})^T \mathbf{C} (\mathbf{b} - \mathbf{f}(\mathbf{p})). \quad (3.20)$$

Assim, quanto mais próximo seja  $\hat{\mathbf{x}}$  ao valor escolhido  $\mathbf{p}$ ; a Eq. (3.20) fica mais próximo a uma igualdade e  $[\mathbf{J}(\mathbf{p})^T \mathbf{C} \mathbf{J}(\mathbf{p})]^{-1} \mathbf{J}(\mathbf{p})^T \mathbf{C} (\mathbf{b} - \mathbf{f}(\mathbf{p})) \approx 0$ . Isto implica duas possibilidades, a primeira é que  $\mathbf{b} \approx \mathbf{f}(\mathbf{p})$ , que é um mínimo absoluto de  $e(\mathbf{x})$ , e a segunda que  $\mathbf{J}(\mathbf{p}) \approx 0$

Assim, um bom critério para procurar um mínimo ou máximo é seguir a seguinte equação iterativa,

$$\mathbf{p}_k \leftarrow \mathbf{p}_{k-1} + [\mathbf{J}(\mathbf{p}_{k-1})^T \mathbf{C} \mathbf{J}(\mathbf{p}_{k-1})]^{-1} \mathbf{J}(\mathbf{p}_{k-1})^T \mathbf{C} (\mathbf{b} - \mathbf{f}(\mathbf{p}_{k-1})), \quad (3.21)$$

iniciando desde um  $\mathbf{p}_0$  qualquer, ate que  $\mathbf{p}_k$  seja muito próximo a  $\mathbf{p}_{k-1}$ , onde se declara que  $\hat{\mathbf{x}} \approx \mathbf{p}_k$ ; porem deve ser corroborado que esse ponto tratasse de um máximo ou mínimo usando algum método, por exemplo estudando o comportamento de  $e(\mathbf{x})$  ou analisando a matriz hessiana de  $e(\mathbf{x})$  avaliada em  $\hat{\mathbf{x}}$ .

**Prova 3.3 — Prova do Teorema 3.3.1** Dados, um escalar  $\alpha \in \mathbb{R}$ , um vetor coluna  $\mathbf{x} \in \mathbb{R}^N$ , um vetor coluna  $\mathbf{q} \in \mathbb{R}^N$ , um vetor coluna  $\mathbf{b} \in \mathbb{R}^M$ , uma função  $\mathbf{f}: \mathbb{R}^N \rightarrow \mathbb{R}^M$ , uma matriz diagonal  $\mathbf{C} \in \mathbb{R}^{M \times M}$ , uma matriz diagonal  $\mathbf{D} \in \mathbb{R}^{N \times N}$ , e definida a Eq. (3.22),

$$e(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha \|\mathbf{x} - \mathbf{q}\|_{\mathbf{D}}^2. \quad (3.22)$$

Para achar o valor  $\hat{\mathbf{x}}$  que gere o menor valor de  $e(\hat{\mathbf{x}})$ , é aplicado o critério que um mínimo ou máximo pode ser achado quando  $\frac{\partial e(\hat{\mathbf{x}})}{\partial \mathbf{x}} = [0 \ 0 \ \dots \ 0]^T$ . Assim, usando o Teorema 2.6.1 podemos rescrever esta igualdade como a Eq. (3.23),

$$2\mathbf{J}(\mathbf{p})^T \mathbf{C} \{ \mathbf{J}(\mathbf{p}) [\hat{\mathbf{x}} - \mathbf{p}] - [\mathbf{b} - \mathbf{f}(\mathbf{p})] \} + 2\alpha \mathbf{D} [\mathbf{x} - \mathbf{q}] \approx \frac{\partial e(\hat{\mathbf{x}})}{\partial \mathbf{x}} = [0 \ 0 \ \dots \ 0]^T, \quad (3.23)$$

de modo que pode ser aproximado  $\hat{\mathbf{x}}$  como:

$$\hat{\mathbf{x}} \approx \mathbf{p} + [\mathbf{J}(\mathbf{p})^T \mathbf{C} \mathbf{J}(\mathbf{p}) + \alpha \mathbf{D}]^{-1} \{ \mathbf{J}(\mathbf{p})^T \mathbf{C} [\mathbf{b} - \mathbf{f}(\mathbf{p})] - \alpha \mathbf{D} [\mathbf{p} - \mathbf{q}] \}. \quad (3.24)$$

Assim, quanto mais próximo seja a  $\hat{\mathbf{x}}$  o valor escolhido  $\mathbf{p}$ , a Eq. (3.24) fica mais próximo a uma igualdade. Por outro lado, a equação nos indica que dado um ponto  $\mathbf{p}$  qualquer,  $\mathbf{p} + [\mathbf{J}(\mathbf{p})^T \mathbf{C} \mathbf{J}(\mathbf{p}) + \alpha \mathbf{D}]^{-1} \{ \mathbf{J}(\mathbf{p})^T \mathbf{C} [\mathbf{b} - \mathbf{f}(\mathbf{p})] - \alpha \mathbf{D} [\mathbf{p} - \mathbf{q}] \}$  é um ponto próximo de  $\mathbf{p}$  na direção de um mínimo ou máximo de  $e(\mathbf{x})$ . Assim, um bom critério para procurar um mínimo ou máximo é seguir a seguinte equação iterativa,

$$\mathbf{p}_k \leftarrow \mathbf{p}_{k-1} + [\mathbf{J}(\mathbf{p}_{k-1})^T \mathbf{C} \mathbf{J}(\mathbf{p}_{k-1}) + \alpha \mathbf{D}]^{-1} \{ \mathbf{J}(\mathbf{p}_{k-1})^T \mathbf{C} [\mathbf{b} - \mathbf{f}(\mathbf{p}_{k-1})] - \alpha \mathbf{D} [\mathbf{p}_{k-1} - \mathbf{q}] \}, \quad (3.25)$$

iniciando desde um  $\mathbf{p}_0$  qualquer, ate que  $\mathbf{p}_k$  seja muito próximo a  $\mathbf{p}_{k-1}$ , onde se declara que  $\hat{\mathbf{x}} \approx \mathbf{p}_k$ ; porem deve ser corroborado que esse ponto tratasse de um máximo ou mínimo usando algum método, por exemplo estudando o comportamento de  $e(\mathbf{x})$  ou analisando a matriz hessiana de  $e(\mathbf{x})$  avaliada em  $\hat{\mathbf{x}}$ .

**Prova 3.4 — Prova do Teorema 3.4.1** Dados, um escalar  $\alpha \in \mathbb{R}$ , um vetor coluna  $\mathbf{x} \in \mathbb{R}^N$ , um vetor coluna  $\mathbf{b} \in \mathbb{R}^M$ , uma função  $\mathbf{f}: \mathbb{R}^N \rightarrow \mathbb{R}^M$ , uma matriz diagonal  $\mathbf{C} \in \mathbb{R}^{M \times M}$ , uma matriz diagonal  $\mathbf{D} \in \mathbb{R}^{N \times N}$ , e definida a Eq. (3.26),

$$e(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha \|\mathbf{x}\|_{\mathbf{D}}^2. \quad (3.26)$$

Para achar o valor  $\hat{\mathbf{x}}$  que gere o menor valor de  $e(\hat{\mathbf{x}})$ , é aplicado o critério que um mínimo ou máximo pode ser achado quando  $\frac{\partial e(\hat{\mathbf{x}})}{\partial \mathbf{x}} = [0 \ 0 \ \dots \ 0]^T$ . Assim, usando a Prova 3.3 para o caso de  $\mathbf{q} = 0$ , sabemos que um bom critério para procurar um mínimo ou máximo é seguir a seguinte equação iterativa,

$$\mathbf{p}_k \leftarrow \mathbf{p}_{k-1} + [\mathbf{J}(\mathbf{p}_{k-1})^T \mathbf{C} \mathbf{J}(\mathbf{p}_{k-1}) + \alpha \mathbf{D}]^{-1} \{ \mathbf{J}(\mathbf{p}_{k-1})^T \mathbf{C} [\mathbf{b} - \mathbf{f}(\mathbf{p}_{k-1})] - \alpha \mathbf{D} \mathbf{p}_{k-1} \}, \quad (3.27)$$

iniciando desde um  $\mathbf{p}_0$  qualquer, ate que  $\mathbf{p}_k$  seja muito próximo a  $\mathbf{p}_{k-1}$ , onde se declara que  $\hat{\mathbf{x}} \approx \mathbf{p}_k$ ; porem deve ser corroborado que esse ponto tratasse de um máximo ou mínimo usando algum método, por exemplo estudando o comportamento de  $e(\mathbf{x})$  ou analisando a matriz hessiana de  $e(\mathbf{x})$  avaliada em  $\hat{\mathbf{x}}$ .

**Prova 3.5 — Prova do Teorema 3.5.1** Dados, um escalar  $\alpha \in \mathbb{R}$ , um vetor coluna  $\mathbf{x} \in \mathbb{R}^N$ , um vetor coluna  $\mathbf{x}_{last} \in \mathbb{R}^N$  (constante), um vetor coluna  $\mathbf{b} \in \mathbb{R}^M$ , uma função  $\mathbf{f}: \mathbb{R}^N \rightarrow \mathbb{R}^M$ , uma matriz diagonal  $\mathbf{C} \in \mathbb{R}^{M \times M}$ , uma matriz diagonal  $\mathbf{D} \in \mathbb{R}^{N \times N}$ , e definida a Eq. (3.28),

$$e(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\|_{\mathbf{C}}^2 + \alpha \|\mathbf{x} - \mathbf{x}_{last}\|_{\mathbf{D}}^2. \quad (3.28)$$

Para achar o valor  $\hat{\mathbf{x}}$  que gere o menor valor de  $e(\hat{\mathbf{x}})$ , é aplicado o critério que um mínimo ou máximo pode ser achado quando  $\frac{\partial e(\hat{\mathbf{x}})}{\partial \mathbf{x}} = [0 \ 0 \ \dots \ 0]^T$ . Assim, usando o Teorema 2.6.1 podemos rescrever esta igualdade como a Eq. (3.29),

$$2\mathbf{J}(\mathbf{p})^T \mathbf{C} \{ \mathbf{J}(\mathbf{p}) [\hat{\mathbf{x}} - \mathbf{p}] - [\mathbf{b} - \mathbf{f}(\mathbf{p})] \} + 2\alpha \mathbf{D} [\mathbf{x} - \mathbf{x}_{last}] \approx \frac{\partial e(\hat{\mathbf{x}})}{\partial \mathbf{x}} = [0 \ 0 \ \dots \ 0]^T, \quad (3.29)$$

de modo que pode ser aproximado  $\hat{\mathbf{x}}$  como:

$$\hat{\mathbf{x}} \approx \mathbf{p} + [\mathbf{J}(\mathbf{p})^T \mathbf{C} \mathbf{J}(\mathbf{p}) + \alpha \mathbf{D}]^{-1} \{ \mathbf{J}(\mathbf{p})^T \mathbf{C} [\mathbf{b} - \mathbf{f}(\mathbf{p})] - \alpha \mathbf{D} [\mathbf{p} - \mathbf{x}_{last}] \}. \quad (3.30)$$

Assim, quanto mais próximo seja a  $\hat{\mathbf{x}}$  o valor escolhido  $\mathbf{p}$ , a Eq. (3.30) fica mais próximo a uma igualdade. Por outro lado, a equação nos indica que dado um ponto  $\mathbf{p}$  qualquer,  $\mathbf{p} + [\mathbf{J}(\mathbf{p})^T \mathbf{C} \mathbf{J}(\mathbf{p}) + \alpha \mathbf{D}]^{-1} \{ \mathbf{J}(\mathbf{p})^T \mathbf{C} [\mathbf{b} - \mathbf{f}(\mathbf{p})] - \alpha \mathbf{D} [\mathbf{p} - \mathbf{x}_{last}] \}$  é um ponto próximo de  $\mathbf{p}$  na direção de um mínimo ou máximo de  $e(\mathbf{x})$ . Assim, igualando  $\mathbf{p}_k = \mathbf{x}_{last}$ , um bom critério para procurar um mínimo ou máximo é seguir a seguinte equação iterativa,

$$\mathbf{p}_k \leftarrow \mathbf{p}_{k-1} + [\mathbf{J}(\mathbf{p}_{k-1})^T \mathbf{C} \mathbf{J}(\mathbf{p}_{k-1}) + \alpha \mathbf{D}]^{-1} \{ \mathbf{J}(\mathbf{p}_{k-1})^T \mathbf{C} [\mathbf{b} - \mathbf{f}(\mathbf{p}_{k-1})] \}, \quad (3.31)$$

iniciando desde um  $\mathbf{p}_0$  qualquer, ate que  $\mathbf{p}_k$  seja muito próximo a  $\mathbf{p}_{k-1}$ , onde se declara que  $\hat{\mathbf{x}} \approx \mathbf{p}_k$ ; porem deve ser corroborado que esse ponto tratasse de um máximo ou mínimo usando algum método, por exemplo estudando o comportamento de  $e(\mathbf{x})$  ou analisando a matriz hessiana de  $e(\mathbf{x})$  avaliada em  $\hat{\mathbf{x}}$ .

## 4. Minimização do erro em funções:

$\mathbb{R} \rightarrow \mathbb{R}$

### 4.1 Minimização de $\|h(x) - a\|^2$

**Teorema 4.1.1 — Solução iterativa** Dados, um escalar  $x \in \mathbb{R}$ , um escalar  $a \in \mathbb{R}$ , uma função  $h : \mathbb{R} \rightarrow \mathbb{R}$ , e definida a Eq. (4.1),

$$e(x) = \|h(x) - a\|^2. \quad (4.1)$$

Se desejamos ter o valor  $\hat{x}$  que minimiza o escalar  $e(\hat{x})$ , este valor pode ser achado usando iterativamente a Eq. (4.2), onde  $h'(x) \equiv \frac{\partial h(x)}{\partial x}$ .

$$x_{k+1} \leftarrow x_k - \frac{h(x_k) - a}{h'(x_k)}, \quad (4.2)$$

Assim,  $\hat{x}$  pode ser achado iniciando a Eq. (4.2) desde um  $x_0$  qualquer,

- ate que  $x_k$  seja muito próximo a  $x_{k-1}$  (convergência de  $x_k$ ) ou
- ate que  $h(x_k)$  seja muito próximo a  $a$ , se conhecemos que esta igualdade é possível, onde pode ser declarado que  $\hat{x} \approx x_k$ .

No processo de busca, podemos ter 3 diferentes casos:

- **Quando  $x_k$  converge:** Se a Eq. (4.2), converge a um valor  $\hat{x}$ , então este se trata de um mínimo absoluto de  $e(x)$ , ou seja  $\hat{x} \approx x_k \approx x_{k-1}$ , ver Prova 4.1 e Exemplos 4.1 e 4.2.
- **Quando  $x_k$  diverge:** Acontece quando se tem passado por um valor de  $x_k$  com  $0 < |h'(x_k)| < \lim_{\varepsilon \rightarrow +0} \varepsilon$ . Esta característica da Eq. (4.2) é um problema, quando o mínimo valor de  $e(x)$  coincide com um lugar onde  $h'(x) = 0$ , ver Exemplo 4.1; pois quando se está perto da solução a busca diverge.



- **Quando é não computável:** Sim se obtêm um ponto  $x_k$  onde  $h'(x_k) = 0$ , deve ser verificado se  $e(x_k)$  é um mínimo, um máximo, relativo ou absoluto, ou um ponto de inflexão.

Para reconhecer o tipo de ponto crítico de  $e(x)$ , e souber se a busca deve ser continuada desde outro ponto, podemos seguir vários procedimentos; por exemplo, estudar o comportamento de  $e(x_k)$  e/ou analisar  $\frac{\partial^2 e(x)}{\partial x^2}$  avaliada em  $x_k$ .

Uma forma de iterar, como a vista na Eq (4.2) é conhecida como método de Newton, porem pela forma de como esta expressão foi deduzida também poderíamos ligar esta ao método de regularização de Tikhonov. A demonstração da Eq (4.2) pode ser vista na Prova 4.1.

■ **Exemplo 4.1** A Fig. 4.1a nos mostra o processo de busca de um mínimo de  $e(x) = (h(x) - a)^2$ , quando  $h(x) = x^2$  e  $a = 1$ . A busca inicia em  $x_0 = -1.5$ , todos os valores  $x_k$  podem ser vistos na segunda coluna da Tabela 4.1. Neste caso a Eq. (4.2) **converge** sem problemas em  $\hat{x} = -1$  com  $e(\hat{x}) = 0$ , pois no caminho não foi achado nenhum ponto  $x_k$  com  $h(x_k) \cong 0$ .

Por outro lado, a Fig. 4.1b nos mostra o processo de busca de um mínimo de  $e(x) = (h(x) - a)^2$ , quando  $h(x) = x^2$  e  $a = -1$ . A busca inicia em  $x_0 = -1.5$ , todos os valores  $x_k$  podem ser vistos na terceira coluna da Tabela 4.1. Neste caso a Eq. (4.2) **diverge** em  $\hat{x} = 0$  com  $e(\hat{x}) = 0$  que é o mínimo; ao principio os valores de  $x_k$  tendem a procurar o mínimo; porem, perto de este valor se cumpre que  $h(x_k) \cong 0$ , e o valor de  $x_k$  diverge a um valor muito longe do mínimo; é fácil de observar, que neste caso se produz uma especie de efeito sanfona, onde  $x_k$  se aproxima ao mínimo  $e(x)$ , e quando está perto do mínimo o valor volta divergir.

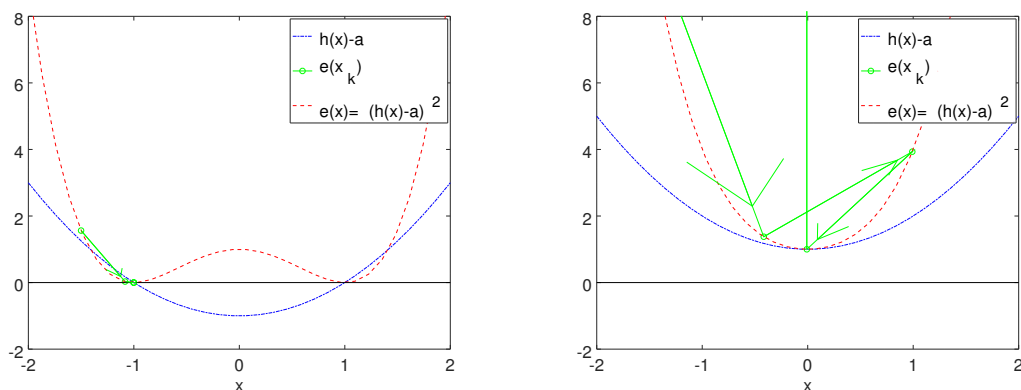
	$e(x) = (x^2 - 1)^2$	$e(x) = (x^2 + 1)^2$
$k$	$x_k$	$x_k$
0	-1.5000	-1.5000
1	-1.0833	-0.41667
2	-1.0032	+0.99167
3	-1.0000	-0.0083683
4	-1.0000	+59.745

Tabela 4.1: Convergência para  $h(x) = x^2$  da equação iterativa do Teorema 4.1.1

■ **Exemplo 4.2** A Fig. 4.2a nos mostra o processo de busca de um mínimo de  $e(x) = (h(x) - a)^2$ , quando  $h(x) = x^3$  e  $a = 1$ ; A busca inicia em  $x_0 = -1.5$ , todos os valores  $x_k$  podem ser vistos na segunda coluna da Tabela 4.2. Neste caso a Eq. (4.2) **diverge** em valores de  $x_k$  próximos a 0, pois provocam valores  $h'(x_k) \cong 0$ , este valor corresponde com um ponto de inflexão de  $e(x)$ .

Por outro lado, a Fig. 4.2b nos mostra o processo de busca de um mínimo de  $e(x) = (h(x) - a)^2$ , quando  $h(x) = x^3$  e  $a = -1$ . A busca inicia em  $x_0 = -1.5$ , todos os valores de  $x_k$  podem ser vistos na terceira coluna da Tabela 4.2. Neste caso a Eq. (4.2) **converge** em  $\hat{x} = -1$  com  $e(\hat{x}) = 0$ , sendo este um mínimo absoluto.





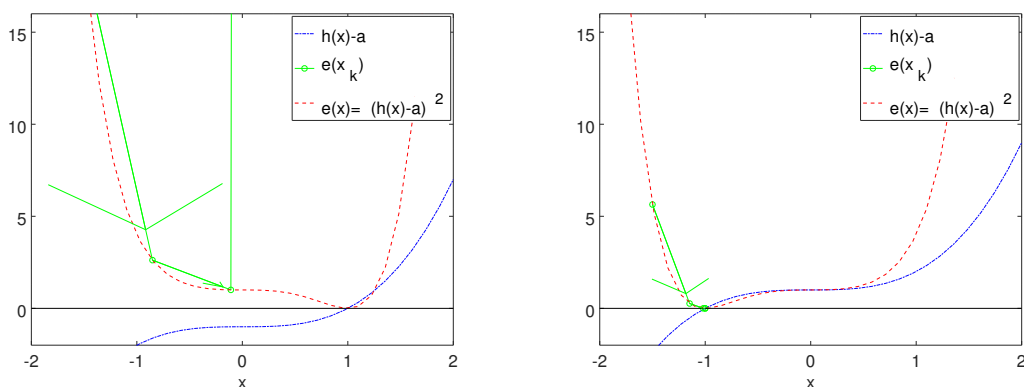
(a) Usando  $h(x) = x^2$  e  $a = 1$ , as iterações convergem

(b) Usando  $h(x) = x^2$  e  $a = -1$ , as iterações divergem

Figura 4.1: Comportamento para  $h(x) = x^2$  da equação iterativa do Teorema 4.1.1

	$e(x) = (x^3 - 1)^2$	$e(x) = (x^3 + 1)^2$
$k$	$x_k$	$x_k$
0	-1.50000	-1.5000
1	-0.85185	-1.1481
2	-0.10854	-1.0183
3	28.21988	-1.0003
4	18.81368	-1.0000

Tabela 4.2: Convergência para  $h(x) = x^3$  da equação iterativa do Teorema 4.1.1



(a) Usando  $h(x) = x^3$  e  $a = 1$ , as iterações divergem

(b) Usando  $h(x) = x^3$  e  $a = -1$ , as iterações convergem

Figura 4.2: Comportamento para  $h(x) = x^3$  da equação iterativa do Teorema 4.1.1

## 4.2 Minimização de $||h(x) - a||^2 + \alpha||x - b||^2$

**Teorema 4.2.1 — Solução iterativa** Dados, um escalar  $\alpha \in \mathbb{R} | \alpha > 0$ , um escalar  $x \in \mathbb{R}$ , um escalar  $a \in \mathbb{R}$ , uma função  $h : \mathbb{R} \rightarrow \mathbb{R}$ , e definida a Eq. (4.3),

$$e(x) = ||h(x) - a||^2 + \alpha ||x - b||^2. \quad (4.3)$$

Se desejamos ter o valor  $\hat{x}$  que minimiza o escalar  $e(\hat{x})$ , este valor pode ser achado usando iterativamente a Eq. (4.4), onde  $h'(x) \equiv \frac{\partial h(x)}{\partial x}$ ,

$$x_{k+1} \leftarrow x_k - \frac{h'(x_k) [h(x_k) - a] + \alpha [x_k - b]}{[h'(x_k)]^2 + \alpha} \quad (4.4)$$

Assim,  $\hat{x}$  pode ser achado iniciando a Eq. (4.4) desde um  $x_0$  qualquer,

- ate que  $x_k$  seja muito próximo a  $x_{k-1}$  (convergência de  $x_k$ ) ou
- ate que  $h(x_k)$  seja muito próximo a  $a$ , se conhecemos que esta igualdade é possível, onde pode ser declarado que  $\hat{x} \approx x_k$ .

No processo de busca, podemos ter 3 diferentes casos:

- **Quando  $x_k$  converge:** Se a Eq. (4.4), converge a um valor  $\hat{x}$ , então este se trata de um mínimo, relativo ou absoluto, ou ponto de inflexão de  $e(x)$ . Ver Prova 4.2 e Exemplo 4.3. É importante ressaltar que quanto maior seja o valor de  $\alpha$  maior será a diferenca do ponto mínimo de  $e(x)$  e  $(h(x) - a)^2$ .
- **Quando  $x_k$  diverge:** Acontece quando  $h'(x) \approx 0$  e o fator  $0 < \alpha < \lim_{\varepsilon \rightarrow +0} \varepsilon$ , ver Exemplo 4.4; pois quando se está perto de um ponto crítico a busca diverge.

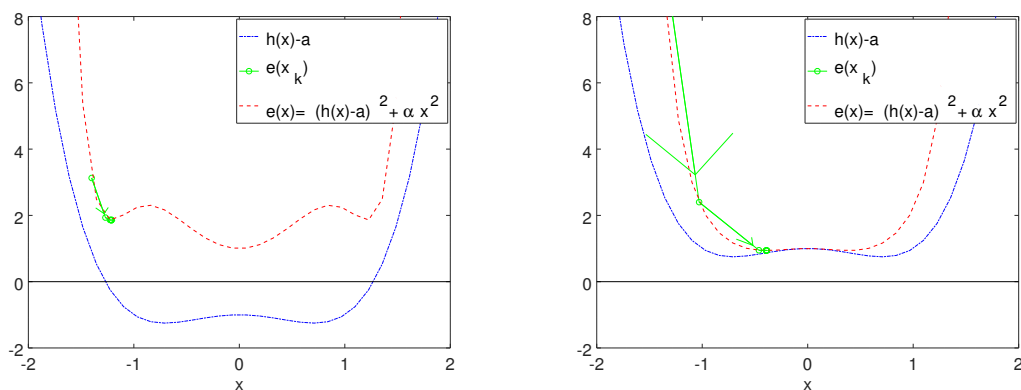
Para reconhecer o tipo de ponto crítico de  $e(x)$ , e souber se a busca deve ser continuada desde outro ponto, podemos seguir vários procedimentos; por exemplo, estudar o comportamento de  $e(x_k)$  e/ou analisar  $\frac{\partial^2 e(x)}{\partial x^2}$  avaliada em  $x_k$ .

Uma forma de iterar, como a vista na Eq (4.4) é conhecida como método de regularização de Tikhonov. A demonstração pode ser vista na Prova 4.2.

■ **Exemplo 4.3** A Fig. 4.3a nos mostra o processo de busca de um mínimo de  $e(x) = (h(x) - a)^2 + \alpha(x - b)^2$ , quando  $h(x) = x^2(x^2 - 1)$ ,  $a = 1$ ,  $b = 0$  e  $\alpha = 1.2$ . A busca inicia em  $x_0 = -1.4$ , todos os valores  $x_k$  podem ser vistos na segunda coluna da Tabela 4.3. Neste caso a Eq. (4.4) **converge** sem problemas em  $\hat{x} = -1.21239$  com  $e(\hat{x}) = 1.85954$ ; porem, este valor de  $x$  está longe do mínimo absoluto de  $e(x)$  que está em  $e(x = 0) = 1.0$ ; por outro lado, este valor de  $x$  está perto do mínimo absoluto de  $(h(x) - a)^2$ , que está em  $x = \pm \sqrt{\frac{1+\sqrt{5}}{2}} \approx \pm 1.2720$ . Isto é consequência do uso do fator  $\alpha > 0$ , pelo qual é de esperar que quanto maior seja o valor de  $\alpha$  maior será a diferenca do mínimo absoluto de  $e(x)$  e de  $(h(x) - a)^2$ .

A Fig. 4.3b nos mostra o processo de busca de um mínimo de  $e(x) = (h(x) - a)^2 + \alpha(x - b)^2$ , quando  $h(x) = x^2(x^2 - 1)$ ,  $a = -1$ ,  $b = 0$  e  $\alpha = 1.2$ . A busca inicia em  $x_0 = -1.4$ , todos os valores  $x_k$  podem ser vistos na terceira coluna da Tabela 4.3. Neste caso a Eq. (4.4) **converge** em  $\hat{x} = -0.39349$  com  $e(\hat{x}) = 0.94120$  que é o mínimo absoluto de  $e(x)$ ; porem, este valor de  $x_k$  está um pouco distante do mínimo de  $(h(x) - a)^2$ , localizado em  $x = \pm \frac{\sqrt{2}}{2} \approx \pm 0.70711$ . Novamente, isto é consequência do uso do fator  $\alpha > 0$ , assim, quanto maior seja o valor de  $\alpha$  maior será a diferenca do mínimo absoluto de  $e(x)$  e de  $(h(x) - a)^2$ .

	$e(x) = (x^2(x^2 - 1) - 1)^2 + 1.2x^2$	$e(x) = (x^2(x^2 - 1) + 1)^2 + 1.2x^2$
$k$	$x_k$	$x_k$
0	-1.4000	-1.40000
1	-1.2694	-1.02908
2	-1.2258	-0.46252
3	-1.2153	-0.38498
4	-1.2130	-0.39259
5	-1.2125	-0.39341
6	-1.2124	-0.39348
7	-1.2124	-0.39349

Tabela 4.3: Convergência para  $h(x) = x^2(x^2 - 1)$  da equação iterativa do Teorema 4.2.1

(a) Usando  $h(x) = x^2(x^2 - 1)$ ,  $a = 1$ ,  $b = 0$  e  $\alpha = 1.2$ , as iterações convergem

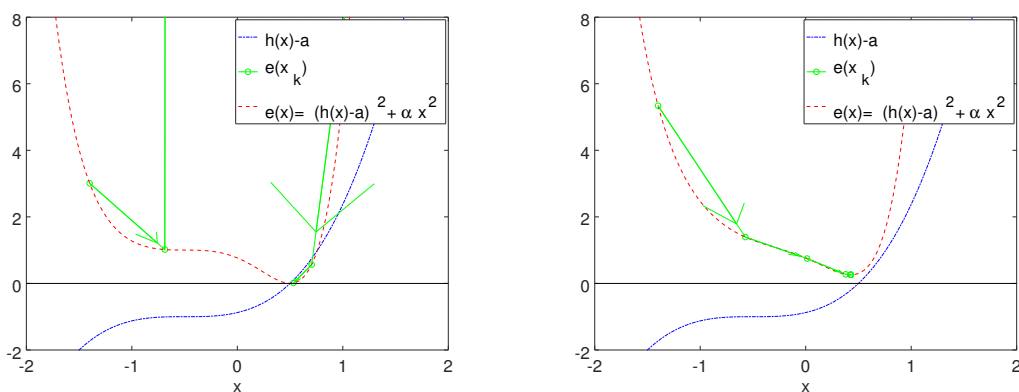
(b) Usando  $h(x) = x^2(x^2 - 1)$ ,  $a = -1$ ,  $b = 0$  e  $\alpha = 1.2$ , as iterações convergem

Figura 4.3: Comportamento para  $h(x) = x^2(x^2 - 1)$  da equação iterativa do Teorema 4.2.1

■ **Exemplo 4.4** A Fig. 4.4a nos mostra o processo de busca de um mínimo de  $e(x) = (h(x) - a)^2 + \alpha(x - b)^2$ , quando  $h(x) = (x + 0.5)^3$ ,  $a = 1$ ,  $b = 0$  e  $\alpha = 0.01$ . A busca inicia em  $x_0 = -1.4$ , todos os valores  $x_k$  podem ser vistos na segunda coluna da Tabela 4.4. Neste caso a Eq. (4.4) num princípio **diverge** em pontos próximos a  $x = -0.5$ , apos a divergência, as iterações continuam desde  $x = +4.665$  e estas **convergem** sem problemas em  $\hat{x} = +0.53255$  com  $e(\hat{x}) \approx 0.013$ ; este valor de  $x$  está perto do mínimo absoluto de  $(h(x) - a)^2$ , que está em  $x = +0.5$ . Isto é consequência do uso do fator  $\alpha > 0$  e pequeno, pelo qual é de esperar que quanto maior seja o valor de  $\alpha$  maior será a diferença do mínimo absoluto de  $e(x)$  e de  $(h(x) - a)^2$ .

A Fig. 4.4b nos mostra o processo de busca de um mínimo de  $e(x) = (h(x) - a)^2 + \alpha(x - b)^2$ , quando  $h(x) = (x + 0.5)^3$ ,  $a = 1$ ,  $b = 0$  e  $\alpha = 1.2$ . A busca inicia em  $x_0 = -1.4$ , todos os valores  $x_k$  podem ser vistos na terceira coluna da Tabela 4.4. Neste caso a Eq. (4.4) **converge** em  $\hat{x} = +0.428757$  com  $e(\hat{x}) \approx 0.26$ , que é o mínimo absoluto de  $e(x)$ ; porem, este valor de  $x_k$  está um pouco distante do mínimo de  $(h(x) - a)^2$ , localizado em  $x = +0.5$ . Novamente, isto é consequência do uso do fator  $\alpha > 0$  e grande.

	$e(x) = ((x+0.5)^3 - 1)^2 + 0.01x^2$	$e(x) = ((x+0.5)^3 + 1)^2 + 1.2x^2$
$k$	$x_k$	$x_k$
0	-1.40000	-1.40000
1	-0.68731	-0.572195
2	+4.66500	+0.012916
3	+2.95583	+0.378950
4	+1.83178	+0.422950
5	+1.11578	+0.427972
6	+0.70475	+0.428660
7	+0.53255	+0.428757

Tabela 4.4: Convergência para  $h(x) = (x+0.5)^3$  da equação iterativa do Teorema 4.2.1

(a) Usando  $h(x) = (x+0.5)^3$ ,  $a = 1$  e  $\alpha = 0.01$ , as iterações divergem e logo convergem  
 (b) Usando  $h(x) = (x+0.5)^3$ ,  $a = 1$  e  $\alpha = 1.2$ , as iterações convergem no mínimo absoluto

Figura 4.4: Comportamento para  $h(x) = (x+0.5)^3$  da equação iterativa do Teorema 4.2.1

### 4.3 Minimização de $\|h(x) - a\|^2 + \alpha\|x - x_{old}\|^2$

**Teorema 4.3.1 — Solução iterativa** Dados, um escalar  $\alpha \in \mathbb{R} | \alpha > 0$ , um escalar  $x \in \mathbb{R}$ , um escalar  $a \in \mathbb{R}$ , uma função  $h : \mathbb{R} \rightarrow \mathbb{R}$ , e definida a Eq. (4.5),

$$e(x) = \|h(x) - a\|^2 + \alpha\|x - x_{old}\|^2. \quad (4.5)$$

Se desejamos ter o valor  $\hat{x}$  que minimiza o escalar  $e(\hat{x})$ , este valor pode ser achado usando iterativamente a Eq. (4.6),

$$x_{k+1} \leftarrow x_k + \frac{h'(x_k)[a - h(x_k)]}{[h'(x_k)]^2 + \alpha} \quad (4.6)$$

Onde  $h'(x) = \frac{\partial h(x)}{\partial x}$ . Assim,  $\hat{x}$  pode ser achado iniciando a Eq. (4.6) desde um  $x_0$  qualquer, ate que  $x_k$  seja muito próximo a  $x_{k+1}$ , onde se declara que  $\hat{x} \approx x_{k+1}$ ; porem deve ser corroborado que esse ponto tratasse de um máximo ou mínimo usando algum método, por exemplo estudando o comportamento de  $e(x)$  ou analisando  $\frac{\partial^2 e(x)}{\partial x^2}$  avaliada em  $\hat{x}$ .

**[FALTA PROVA!!!]** A demonstração pode ser vista na Prova ??.

## 4.4 Provas dos teoremas

**Prova 4.1 — Prova do Teorema 4.1.1** Dados, um escalar  $x \in \mathbb{R}$ , um escalar  $a \in \mathbb{R}$ , uma função  $h : \mathbb{R} \rightarrow \mathbb{R}$ , e definida a Eq. (4.7),

$$e(x) = ||h(x) - a||^2. \quad (4.7)$$

Para achar o valor  $\hat{x}$  que gere o menor valor de  $e(x)$ , é aplicado o critério que um ponto crítico (máximo, mínimo ou ponto de inflexão) de  $e(x)$  pode ser achado quando  $\left. \frac{\partial e(x)}{\partial x} \right|_{x=\hat{x}} \equiv e'(\hat{x}) = 0$ . Assim, usando o Teorema 2.4.1 podemos rescrever esta igualdade como a Eq. (4.8),

$$2h'(x) [h(x) - a] = e'(x) = 0, \quad (4.8)$$

Da Eq. (4.8), observamos que existem duas formas de achar um ponto crítico em  $e(x)$ ,

- a primeira é achando um valor  $x^*$  tal que  $h'(x^*) = 0$ , e
- a segunda forma é achando um valor  $\hat{x}$  tal que  $h(\hat{x}) - a = 0$ ;

de isto se deduz que, um valor  $x^*$  que provoque  $h'(x^*) = 0$ , representa um ponto crítico de  $e(x)$  e  $h(x)$ ; porem, um ponto  $\hat{x}$  que provoque  $h(\hat{x}) - a = 0$ , representa um ponto crítico de  $e(x)$ , mas não necessariamente de  $h(x)$ . Assim, dada a natureza positiva de  $e(x)$ , os pontos críticos diferente dos existentes em  $h(x)$ , são obrigatoriamente mínimos absolutos, pois provocam  $e(\hat{x})=0$ .

Por outro lado, usando o Teorema 2.4.2 podemos rescrever a Eq. (4.8) de forma aproximada como a Eq. (4.9),

$$2h'(p) [(h(p) - a) + h'(p)(\hat{x} - p)] \approx e'(\hat{x}) = 0. \quad (4.9)$$

de modo que, se consideramos  $h'(p) \neq 0$ , a equação pode ser rescrita como:

$$\hat{x} \approx p - \frac{(h(p) - a)}{h'(p)}. \quad (4.10)$$

Isto quer dizer que a Eq. 4.10, **serve só para achar pontos críticos de  $e(x)$  que não pertençam a  $h(x)$** . Sendo este obrigatoriamente um ponto onde  $h(\hat{x}) - a \approx 0$ , equivalente a um mínimo absoluto.

Assim, quanto mais próximo a  $\hat{x}$  seja o valor  $p$ , a Eq. (4.10) fica mais próximo a uma igualdade. Por outro lado, a equação nos indica que dado um ponto  $p$  qualquer,  $p - [h'(p)]^{-1} (h(p) - a)$  é um ponto próximo de  $p$  na direção de um mínimo de  $e(x)$ . De modo que, um bom critério para procurar um mínimo é seguir a seguinte equação iterativa,

$$p_k \leftarrow p_{k-1} - \frac{(h(p_{k-1}) - a)}{h'(p_{k-1})}, \quad (4.11)$$

iniciando desde um  $p_0$  qualquer, ate que  $p_k$  seja muito próximo a  $p_{k-1}$ , onde se declara que  $\hat{x} \approx p_k$ .

porem deve-se ter cuidado com valores de  $p_k$  que provoquem  $0 < |h'(p_k)| < \lim_{\varepsilon \rightarrow +0} \varepsilon$ , pois estes valores tendem a divergir. Por outro lado sim se acha um valor de  $p_k$  com  $h'(p_k) = 0$ , deve ser corroborado se este ponto tratasse de um máximo ou mínimo, relativo ou absoluto, ou ponto de inflexão; usando algum método, por exemplo estudando o comportamento de  $e(x)$  ou analisando a matriz hessiana de  $e(x)$  avaliada em  $p_k$ .

**Prova 4.2 — Prova do Teorema 4.2.1** Dados, um escalar  $x \in \mathbb{R}$ , um escalar  $a \in \mathbb{R}$ , um escalar  $\alpha \in \mathbb{R}^+$ , um escalar  $b \in \mathbb{R}$ , uma função  $h : \mathbb{R} \rightarrow \mathbb{R}$ , e definida a Eq. (4.12),

$$e(x) = ||h(x) - a||^2 + \alpha ||x - b||^2. \quad (4.12)$$

Para achar o valor  $\hat{x}$  que gere o menor valor de  $e(x)$ , é aplicado o critério que um ponto crítico (máximo, mínimo ou ponto de inflexão) de  $e(x)$  pode ser achado quando  $\left. \frac{\partial e(x)}{\partial x} \right|_{x=\hat{x}} \equiv e'(\hat{x}) = 0$ . Assim, usando o Teorema 2.4.1 e 2.3.1 podemos rescrever esta igualdade como a Eq. (4.13),

$$2h'(x)[h(x) - a] + 2\alpha(x - b) = e'(x) = 0, \quad (4.13)$$

Da Eq. (4.13), observamos que existem duas formas de achar um ponto crítico em  $e(x)$ ,

- a primeira é sim  $h'(x)$  tem um fator  $(x - b)$  tal que  $h'(b) = 0$ , e
- a segunda forma é achando um valor  $\hat{x}$  tal que  $e'(\hat{x}) = 0$ ;

de isto se deduz que se  $h'(x)$  tem um fator  $(x - b)$  então,  $x = b$  representa um ponto crítico de  $e(x)$  e  $h(x)$ ; porem, existe um ponto  $\hat{x}$  que provoca  $e'(\hat{x}) = 0$ , representa um ponto crítico de  $e(x)$ , exclusivamente. Assim, dada a natureza positiva de  $e(x)$ , este é obrigatoriamente um mínimo absoluto de  $e(x)$ .

Por outro lado, usando o Teorema 2.6.1 podemos rescrever a Eq. (4.13) de forma aproximada como a Eq. (4.14),

$$2h'(p)[(h(p) - a) + h'(p)(\hat{x} - p)] \approx e'(\hat{x}) = 0. \quad (4.14)$$

de modo que, se consideramos  $h'(p) \neq 0$ , a equação pode ser rescrita como:

$$\hat{x} \approx p - \frac{(h(p) - a)}{h'(p)}. \quad (4.15)$$

Isto quer dizer que a Eq. 4.15, **serve só para achar pontos críticos de  $e(x)$  que não pertençam a  $h(x)$** . Sendo este obrigatoriamente um ponto onde  $h(\hat{x}) - a \approx 0$ , equivalente a um mínimo absoluto.

Assim, quanto mais próximo a  $\hat{x}$  seja o valor  $p$ , a Eq. (4.15) fica mais próximo a uma igualdade. Por outro lado, a equação nos indica que dado um ponto  $p$  qualquer,  $p - [h'(p)]^{-1}(h(p) - a)$  é um ponto próximo de  $p$  na direção de um mínimo de  $e(x)$ . De modo que, um bom critério para procurar um mínimo é seguir a seguinte equação iterativa,

$$p_k \leftarrow p_{k-1} - \frac{(h(p_{k-1}) - a)}{h'(p_{k-1})}, \quad (4.16)$$

iniciando desde um  $p_0$  qualquer, ate que  $p_k$  seja muito próximo a  $p_{k-1}$ , onde se declara que  $\hat{x} \approx p_k$ .

porem deve-se ter cuidado com valores de  $p_k$  que provoquem  $0 < |h'(p_k)| < \lim_{\varepsilon \rightarrow +0} \varepsilon$ , pois estes valores tendem a divergir. Por outro lado sim se acha um valor de  $p_k$  com  $h'(p_k) = 0$ , deve ser corroborado se este ponto tratasse de um máximo ou mínimo, relativo ou absoluto, ou ponto de inflexão; usando algum método, por exemplo estudando o comportamento de  $e(x)$  ou analisando a matriz hessiana de  $e(x)$  avaliada em  $p_k$ .



# Problemas de achar raízes de funções

<b>5</b>	<b>Raízes de funções: <math>\mathbb{R} \rightarrow \mathbb{R}</math> .....</b>	<b>41</b>
5.1	Raiz de $h(x)$ acotada entre $a$ e $b$	
5.2	Raiz de $h(x)$ não acotada (Métodos iterativos)	







## 5. Raíces de funções: $\mathbb{R} \rightarrow \mathbb{R}$

### 5.1 Raiz de $h(x)$ acotada entre $a$ e $b$

---

5.1.1 Método de la falsa posición

5.1.2 Método de bisección

### 5.2 Raiz de $h(x)$ não acotada (Métodos iterativos)

---

5.2.1 Método de las aproximaciones sucesivas

5.2.2 Método de la secante

5.2.3 Método de Newton



# IV

## Problemas varios

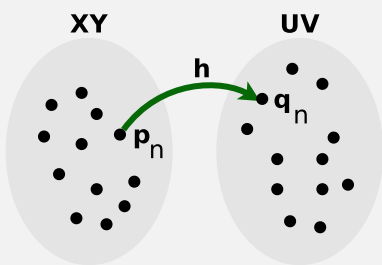
<b>6</b>	<b>Mapeamento .....</b>	<b>45</b>
6.1	Achar a função de mapeamento $\mathbf{h}(x, y) : (x, y) \rightarrow (u, v)$	



## 6. Mapeamento

### 6.1 Achar a função de mapeamento $\mathbf{h}(x, y) : (x, y) \rightarrow (u, v)$

#### Teorema 6.1.1 — Mapeamento usando um polinômio de grau $M$



Se temos,  $N$  pontos (amostras)  $\mathbf{p}_n = [x_n \ y_n]^T \equiv (x_n, y_n)$  no plano  $XY \in \mathbb{R}^2$ , e seus correspondentes  $\mathbf{q}_n = [u_n \ v_n]^T \equiv (u_n, v_n)$  no plano  $UV \in \mathbb{R}^2$ ,  $\forall n \in \{0, 1, 2, \dots, N-1\}$ , vinculados mediante a função  $\mathbf{h}(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , que é um vetor função de transformação de coordenadas,  $\mathbf{q} = \mathbf{h}(\mathbf{p})$ , para um ponto  $\mathbf{p} \in XY$  qualquer e seu correspondente  $\mathbf{q} \in UV$ .

Então, podemos modelar  $[u \ v]^T = \mathbf{h}(\mathbf{p}) \equiv \mathbf{h}(x, y) \equiv [h_1(x, y) \ h_2(x, y)]^T$ , onde as funções  $h_1(x, y)$  e  $h_2(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}$ , são polinômios<sup>a</sup> de grau  $M$ ,

$$\begin{aligned}
 h_1(x, y) &= +c_0 \\
 &\quad +c_1 x + c_2 y \\
 &\quad +c_3 x^2 + c_4 xy + c_5 y^2 \\
 &\quad \dots \\
 &\quad + \sum_{l=0}^M c_{\left\{\frac{M(M+1)}{2}+l\right\}} x^{M-l} y^l; \\
 h_2(x, y) &= \sum_{m=0}^M \sum_{l=0}^m d_{\left\{\frac{m(m+1)}{2}+l\right\}} x^{m-l} y^l.
 \end{aligned} \tag{6.1}$$

Assim, o problema de conhecer  $\mathbf{h}(x, y)$  é equivalente ao problema de achar os vetores  $\mathbf{c} = [c_0 \ c_1 \ c_2 \ \dots \ c_{\left\{\frac{M(M+1)}{2}\right\}}]^T$  e  $\mathbf{d} = [d_0 \ d_1 \ d_2 \ \dots \ d_{\left\{\frac{M(M+1)}{2}\right\}}]^T$ ; com este fim são usadas as  $N$  amostras  $\mathbf{p}_n$  e  $\mathbf{q}_n$ , para procurar quais são os valores de  $\mathbf{c}$  e  $\mathbf{d}$  que provocam o mínimo erro nos

escalares,  $e_1(\mathbf{c})$  e  $e_2(\mathbf{d})$ .

$$e_1(\mathbf{c}) = \|\mathbf{Ac} - \mathbf{u}\|^2 \quad e_2(\mathbf{d}) = \|\mathbf{Ad} - \mathbf{v}\|^2, \quad (6.2)$$

onde

$$\mathbf{u} = \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ \vdots \\ u_{N-1} \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ \vdots \\ v_{N-1} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{a}_{00} & \mathbf{a}_{01} & \dots & \mathbf{a}_{0(M-1)} \\ \mathbf{a}_{10} & \mathbf{a}_{11} & \dots & \mathbf{a}_{1(M-1)} \\ \mathbf{a}_{20} & \mathbf{a}_{21} & \dots & \mathbf{a}_{2(M-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_{(N-1)0} & \mathbf{a}_{(N-1)1} & \dots & \mathbf{a}_{(N-1)(M-1)} \end{bmatrix} \quad (6.3)$$

$$\mathbf{a}_{nm} = [x_n^m \quad x_n^{m-1}y_n \quad x_n^{m-2}y_n^2 \quad \dots \quad x_n^1y_n^{m-1} \quad y_n^m]. \quad (6.4)$$

Um caso especial é quando  $\mathbf{a}_{n0} = 1$ . Assim, utilizando o Teorema 3.1.1, podemos achar o valor  $\hat{\mathbf{c}}$  e  $\hat{\mathbf{d}}$ , que minimizam as funções  $e_1(\hat{\mathbf{c}})$  e  $e_2(\hat{\mathbf{d}})$ :

$$\hat{\mathbf{c}} = [\mathbf{A}^T \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{u}, \quad \hat{\mathbf{d}} = [\mathbf{A}^T \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{v}. \quad (6.5)$$

Devemos ter cuidado, para que exista uma solução única para  $\hat{\mathbf{c}}$  e  $\hat{\mathbf{d}}$ , que o número de linhas de  $\mathbf{A}$  deve ser maior ou igual ao número de suas colunas<sup>b</sup>. Assim, devemos escolher um valor  $M$  que cumpra a seguinte relação com  $N$ ,

$$N \geq \frac{(M+1)(M+2)}{2}. \quad (6.6)$$

<sup>a</sup>Os polinômios  $h_1(x, y)$  e  $h_2(x, y)$  são semelhantes, só a notação na escrita foi diferente.

<sup>b</sup> Que é equivalente ao número de elementos de  $\hat{\mathbf{c}}$  e  $\hat{\mathbf{d}}$



# Referencias

<b>Bibliography .....</b>	<b>49</b>
<b>Index .....</b>	<b>51</b>







## Bibliografía

- [] *Gradient. Encyclopedia of Mathematics.* URL: <http://www.encyclopediaofmath.org/index.php?title=Gradient&oldid=28205> (ver página 13).
- [] *Hessian of a function. L.D. Ivanov (originator), Encyclopedia of Mathematics.* URL: [http://www.encyclopediaofmath.org/index.php?title=Hessian\\_of\\_a\\_function&oldid=19311](http://www.encyclopediaofmath.org/index.php?title=Hessian_of_a_function&oldid=19311) (ver páginas 13, 19).
- [] *Jacobian. Encyclopedia of Mathematics.* URL: <http://www.encyclopediaofmath.org/index.php?title=Jacobian&oldid=28782> (ver páginas 13, 19, 20, 26, 27).
- [] *Taylor series. Encyclopedia of Mathematics.* URL: [http://www.encyclopediaofmath.org/index.php?title=Taylor\\_series&oldid=31211](http://www.encyclopediaofmath.org/index.php?title=Taylor_series&oldid=31211) (ver páginas 13–15, 19, 20).





# Índice Remissivo

## G

Gradiente ..... 13, 14

## M

Mapeamento ..... 45

Matriz Hessiana ..... 13, 14, 19

Matriz Jacobiana ..... 13, 15, 19

Minimização do erro quadrático

    Linear ..... 25

    Não linear ..... 26, 27, 32

Minimização, métodos

    Método de Newton ..... 32

    Regularização de Tikhonov . 26, 27, 32

## P

Problema inverso

    Linear ..... 25, 45

    Não linear ..... 26, 27, 32

Pseudo-inversa de Moore-Penrose ..... 25

## S

Serie de Taylor ..... 13, 19, 20