Universidade Federal do Piauí
Centro de Ciências da Natureza
Pós-Graduação em Matemática
Mestrado em Matemática

# Convergência Local do Método de Gauss-Newton para Sistemas de Equações Não-lineares sob Condição Majorante.

**Samara Costa Lima**

**Teresina - 2013**

**Samara Costa Lima**

**Dissertação de Mestrado:**

# Convergência Local do Método de Gauss-Newton para Sistemas de Equações Não-lineares sob Condição Majorante.

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Matemática, da Universidade Federal do Piauí, como requisito parcial para obtenção do grau de Mestre em Matemática.

Orientadora:
Prof$^{a}$. Dr$^{a}$. Sissy da Silva Souza.

**Teresina - 2013**

Lima, Samara Costa.
Convergência Local do Método de Gauss-Newton para Sistemas de Equações Não-lineares sob Condição Majorante.

Samara Costa Lima – Teresina: 2013.

Orientadora: Prof$^{a}$. Dr$^{a}$. Sissy da Silva Souza.
1.Otimização

CDD 516.36

*Dedico esta dissertação a minha mãe e minhas irmãs.*

# Agradecimentos

À Deus, pela proteção durante este percurso da minha carreira estudantil, por ter me dado a oportunidade e a capacidade, pois sem Deus nada disso seria possível.

À minha familia, pela compreensão e paciência, que permitiram a estabilidade necessária ao longo deste periodo.

Agradeço à Profª. Drª. Sissy da Silva Souza, pela orientação, disponibilidade, sugestões feitas durante a escrita e revisão final deste trabalho, bem como ao apoio, paciência, incentivo, confiança, dedicação e amizade que foram indispensáveis para a concretização deste trabalho.

Agradeço à todos os professores da UFPI do departamento de matemática, pelo conhecimento transmitido, dentre eles: Sissy, Paulo Sergio, Marcondes Clark, Xavier, Marcos Vinicios, Jurandir, Paulo Alexandre, Barnabé Pessoa e Alexandro Marinho.

Agradeço aos amigos pelo apoio e parcerias de estudos entre eles: Gilson, Leonardo, Rui Marques, Andrelino, Lucas Vidal, Felipe, Valdinês, Bernardo, Jeffesson, Renata, Diego Prudêncio, Israel Evangelista, Mikael, Emerson, Alexandre, Ailton, Valdir, Sandoel, Ramon, Vitaliano e Yuri.

Agradeço aos professores Paulo Roberto Oliveira (UFRJ) e Paulo Sérgio Marques dos Santos (UFPI) por terem aceito compor a banca de avaliação deste trabalho, pelo apoio e valiosas sugestões.

À CAPES, pelo apoio financeiro.

“*Para ser grande, sê inteiro:*
*Nada teu exagera ou exclui.*
*Sê todo em cada coisa.*
*Põe quanto és no mínimo que fazes.*
*Assim em cada lago a lua toda brilha,*
*Porque alta*”.

Ricardo Reis.

# Resumo

O método de Gauss-Newton e suas variações são alguns dos mais eficientes métodos conhecidos para resolver problemas de mínimos quadrados não-lineares, os quais são aplicados em diversas áreas da ciência e engenharias. Nesta dissertação apresentaremos uma análise de convergência local do método de Gauss-Newton para resolver determinados sistemas de equações não lineares em espaço de Hilbert sob uma condição majorante. Daremos ênfase nos casos em que a função majorante tem derivada convexa e no caso em que ela não tem essa hipótese, em ambos os casos veremos que o método está bem definido e converge para a solução do problema.

# Abstract

The Gauss-Newton method and its variants are some of the most effective known methods for solving nonlinear least squares, which are applied in many areas of science and engineering. In this dissertation we present an analysis of local convergence of Gauss-Newton method for solving certain overdetermined nonlinear systems of equations in Hilbert space under majorant conditions. We will emphasize two cases: where the majorant function has derived convex and where it has no such assumption, in both cases we will see that the method is well defined and converges to the solution of the problem studied.

# Sumário

# Introdução

O método clássico de Newton para determinar zero de operadores continuamente diferenciáveis $F : \Omega \subset X \to Y$, onde X e Y são espaços de Hilbert, ou seja, resolver o seguinte problema:

$$F(x) = 0 \tag{1}$$

é definido por:
*Inicia de um ponto inicial $x_0 \in \Omega$, e a sequência $\{x_k\}$ é definida como:*

$$x_{k+1} = x_k - (F'(x_k))^{-1} F(x_k),$$

Para garantir convergência quadratica do método de Newton para a solução da equação não-linear $F$, necessitamos de hipóteses adequadas sobre $F$ e o ponto inicial $x_0$. Por exemplo, análise de conergência clássica requer que o ponto inicial seja tomado "suficientemente próximo"da solução e a derivada do operador não-linear seja invertível nesta solução, veja por exemplo em [21], [22] e [25].

Este método foi vastamente estudado para vários tipos de problemas de otimização, como por exemplo pode ser visto em [29]. Agora, se $F'(x_*)$ não é invertível, temos uma generalização do método de Newton chamado método de Gauss-Newton, onde $F : \Omega \to Y$ é uma função continuamente diferenciável, tal que $F'(x)$ é injetivo e tem imagem fechada no aberto $\Omega \subseteq X$, com $X$ e $Y$ espaços de Hilbert. Este método é definido por

$$x_{k+1} = x_k - \left[F'(x_k)^* F'(x_k)\right]^{-1} F'(x_k)^* F(x_k), \quad k = 0, 1, ....$$

em que $F'(x_k)^*$ representa o operador adjunto de $F'(x_k)$.

Existe uma vasta literatura dedicada ao estudo dos resultados de convergência para o método de gauss-Newton sob diferentes perspectivas. Em particular, podemos distinguir duas principais correntes de pesquisas: análise local e semi-local. A primeira classe de estudos ([15], [16],[21] e [35]) assumem a existência de minimizador local, e determinam uma região de atração do ponto, o qual significa que se o ponto de partida for escolhido dentro dessa região o processo iterativo é garantido e converge para o minimizador. Por outro lado, o resultado do semi-local, também conhecido como Teorema do Kantoravich não assumem a existência de um minimizador local, ele apenas estabelece condições suficiêntes sobre o ponto de partida para tornar o processo iterativo convergente em direção a um ponto que se provou ser um minimizador local veja por exemplo em [7], [14] e [30].

O algoritmo obtido pelo método de Gauss-Newton é um método para se obter soluções de problemas de mínimos quadrados não-lineares, o qual pode ser visto como uma modificação do método clássico de Newton para minimizar (ou maximizar) funções. Diferentemente do método de Newton, o método de Gauss-Newton só pode ser aplicado para minimizar problemas de mínimos quadrados.

A busca dos mínimos quadrados de (1) é equivalente a buscar solução do seguinte problema não-linear:

$$\min \|F(x)\|^2. \tag{2}$$

Existem diversas aplicações prática para este problema, ver [4]. Tais aplicações, tem o objetivo de encontrar os parâmetros de um modelo matemático, que melhor descreva um conjunto de dados numericos de um experimento químico, físico, estatistico ou econômico, usando uma função da forma (2) para medir a discrepância entre as saidas do modelo e o conjunto de dados. A análise de convergência para os métodos de Gauss-Newton requer hipóteses sobre a $F$ dentre elas $F'$ deve satisfazer a condição Lipschitz (ver [6], [23] e [24]). Nos ultimos anos tem-se trabalhado no método de Gauss-Newton no intuito de relaxar a hipótese de continuidade Lipschitz para $F'$ veja por exemplo em [5], [17], [18], [19], [26] [28],[29], [30] e [31], onde nesta dissertação veremos no caso em que a condição Lipschitz para $F'$ é relaxada usando o principio majorante introduzido por Kantorovich em [26] e usado por Ferreira em [28] e [29], Ferreira, Gonçalves em [31], Ferreira, Svaiter em [30] e Gonçalves [27].

Nesta dissertação estamos interessados no estudo da convergência do método de Gauss-Newton sob condição majorante, como estudado por exemplo em [27] e [32].

Nesta dissertação apresentamos uma análise de convergência local do método de Gauss-Newton para resolver o problema (1) sob condição majorante, com dois enfoques: o caso em que a função majorante tem derivada convexa (hipótese assumida em [32]) e o caso que tal função não tem derivada convexa visto em [27]. Em ambos os casos apresentaremos convergência, singularidade, taxa de convergência e estimativa do melhor raio de convergência possível. Vale ressaltar que a hipótese de convexidade da derivada da função majorante é necessária apenas para obter a taxa de convergência.

Esta dissertação está organizada da seguinte forma. No Capítulo 1, daremos alguns resultados preliminares sobre: fatos básicos de funções convexas, espaços de dimensão infinita, método clássico de Newton, inversa generalizada, função majorante e propriedades e relações entre $F$ e $f$ necessárias para a analise de convergência dos métodos. No Capítulo 2, apresentaremos uma discussão sobre a convergência local do método de Gauss-Newton sob uma condição majorante, cuja derivada da função $f$ é convexa. No Capítulo 3, obteremos resultados análogos aos obtidos no Capítulo 2, retirando a hipótese da derivada da função ser convexa. Em ambos os casos mostraremos que sob certas condições, a sequência gerada pelo método está bem definida e converge para uma solução de (1). Além disso, apresentaremos o raio de convergência ótimo e a unicidade da solução de $F$.

# Capítulo 1

# Noções preliminares

O objetivo deste capítulo é apresentar as definições e resultados básicos de análise convexa, espaços de dimensão infinita, o método clássico de Newton, inversa generalizada e função majorante. Apresentaremos ainda, a relação entre operador não-linear $F$ e a função majorante $f$. Estes resultados serão úteis para os proximos capítulos.

## 1.1 Fatos básicos de análise convexa

Nesta seção, trataremos sobre conceitos relacionados aos conjuntos convexos e às funções convexas. Iniciaremos definindo conjunto convexo.

Um conjunto convexo se caracteriza por conter todos os segmentos cujos extremos pertencem ao conjunto.

**Definição 1.1.1.** *Um conjunto $D \subset \mathbb{R}^n$ é dito convexo se*

$$\alpha x + (1-\alpha)y \in D, \quad \forall\ x, y \in D, \quad \alpha \in [0,1].$$

O ponto $\alpha x + (1-\alpha)y$, onde $\alpha \in [0,1]$, é chamado de combinação convexa de $x$ e $y$ (com parâmetro $\alpha$).

**Definição 1.1.2.** *Seja $D \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : D \to \mathbb{R}$ é chamada convexa se*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) \quad \forall\ x, y \in D, \quad \forall\ \alpha \in [0,1].$$

A função $f$ é dita estritamente convexa quando a última desigualdade é estrita para todo $x \neq y$ e $\alpha \in (0,1)$.

**Exemplo 1.1.1.** *As funções $f, g : \mathbb{R} \to \mathbb{R}$ dadas por $f(x) = x^2$ e $g(x) = e^x$ são convexas.*

*Demonstração.* De fato, pois

$$\begin{aligned} f(((1-t)x + ty) &= ((1-t)x + ty)^2 \\ &= (x + t(y-x))^2 \\ &= x^2 + 2tx(y-x) + t^2(y-x)^2 \\ &\leq x^2 + 2tx(y-x) + t(y-x)^2 \\ &= x^2 + t(y^2 - x^2) \\ &= (1-t)x^2 + ty^2 \\ &= (1-t)f(x) + tf(y), \end{aligned}$$

logo $f$ é convexa.

Agora mostraremos que $g$ é convexa, para isto, considere $z = (1-t)x + ty$, como $e^d \geq 1 + d$, para todo $d \in \mathbb{R}$, temos

$$e^x \geq e^z + e^z(x - z) \tag{1.1}$$

e

$$e^y \geq e^z + e^z(y - z). \tag{1.2}$$

Assim, multiplicando a desigualdade (1.1) por $(1-t)$ e a desigualdade (1.2) por $t$, obtemos

$$e^z \leq (1-t)e^x + te^y,$$

portanto $g$ é convexa. $\square$

A seguir apresentaremos uma caracterização para funções convexas diferenciáveis.

**Proposição 1.1.1** ([3])**.** *Sejam $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável e $D \subset \mathbb{R}^n$ um conjunto convexo. A função $f$ é convexa em $D$ se, e somente se,*

$$f(y) \geq f(x) + f'(x)^T \cdot (y - x) \ \forall x, y \in D, \tag{1.3}$$

*onde $f'(x)^T$ é a transposta do gradiente da função $f$ no ponto $x$.*

*Demonstração.* Primeiramente consideremos $f$ convexa para $x, y \in D$ e $\alpha \in (0, 1]$. Definindo $d = y - x$ temos $x + \alpha d \in D$ ( pois $D$ é convexo) e

$$f(x + \alpha d) = f((1-\alpha)x + \alpha y) \leq (1-\alpha)f(x) + \alpha f(y).$$

Portanto,
$$f(y) - f(x) \geq \frac{f(x+\alpha d) - f(x)}{\alpha}.$$
Logo, passando ao limite na última desigualdade quando $\alpha \to 0^+$ temos

$$f(y) - f(x) \geq \lim_{\alpha \to 0^+} \frac{f(x+\alpha d) - f(x)}{\alpha} = f'(x)^T \cdot d = f'(x)^T \cdot (y-x),$$

então
$$f(y) - f(x) \geq f'(x)^T \cdot (y-x).$$
Assim, concluímos a primeira parte da demonstração.

Agora, tomando $z = (1-\alpha)x + \alpha y$, e usando (1.3), temos

$$f(x) \geq f(z) + f'(z)^T \cdot (x-z)$$

e
$$f(y) \geq f(z) + f'(z)^T \cdot (y-z).$$
Multiplicando a primeira desigualdade por $(1-\alpha)$ e a segunda por $\alpha$ obtemos

$$(1-\alpha)f(x) \geq (1-\alpha)f(z) + (1-\alpha)f'(z)^T \cdot (x-z),$$

$$\alpha f(y) \geq \alpha f(z) + \alpha f'(z)^T \cdot (y-z).$$
Assim, somando as duas últimas desigualdades segue

$$(1-\alpha)f(x) + \alpha f(y) \geq f(z) = f((1-\alpha)x + \alpha y).$$

Então

$$f(1-\alpha)x + \alpha y) \leq (1-\alpha)f(x) + \alpha f(y), \ \forall x, y \in D \quad e \quad \alpha \in [0,1],$$

o que conclui o resultado. □

Observe que as definições e a propriedade de analise convexa que vimos anteriormente, também são válidas para $n = 1$. Nas proposições seguintes estudaremos as funções convexas de uma variável real.

**Proposição 1.1.2** ([10]). *Sejam $f : (a,b) \to \mathbb{R}$ uma função diferenciável e $f'$ uma função estritamente crescente em $(a,b)$, então $f$ é estritamente convexa em $(a,b)$.*

*Demonstração.* Sejam $f'$ crescente e $x < y$ com $x, y \in (a, b)$ e seja $g : [0, 1] \to \mathbb{R}$, definida por

$$g(\alpha) := f(\alpha x + (1-\alpha)y) - \alpha f(x) - (1-\alpha)f(y).$$

Mostraremos que $g(\alpha) < 0 \quad \forall\, \alpha \in (0, 1)$.

$$\begin{aligned} g(\alpha) &= [\alpha + (1-\alpha)]f(\alpha x + (1-\alpha)y) - \alpha f(x) - (1-\alpha)f(y) \\ &= \alpha[f(\alpha x + (1-\alpha)y) - f(x)] + (1-\alpha)[f(\alpha x + (1-\alpha)y) - f(y)]. \end{aligned} \tag{1.4}$$

Pelo Teorema do Valor Médio existem $x_1$ e $y_1$ com $x < x_1 < \alpha x + (1-\alpha)y < y_1 < y$, tais que

$$f(\alpha x + (1-\alpha)y) - f(x) = (1-\alpha)(y-x)f'(x_1), \tag{1.5}$$

e

$$f(y) - f(\alpha x + (1-\alpha)y) = \alpha(y-x)f'(y_1). \tag{1.6}$$

Substituindo as equações (1.5) e (1.6) em (1.4), obtemos

$$\begin{aligned} g(\alpha) &= \alpha[(1-\alpha)(y-x)f'(x_1)] + (1-\alpha)[-\alpha(y-x)f'(y_1)] \\ &= \alpha(1-\alpha)(y-x)[f'(x_1) - f'(y_1)] \\ &< 0. \end{aligned}$$

Daí, temos que $g(\alpha) < 0$ o qual implica em $f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y)$.

Logo, $f$ é estritamente convexa em $(a, b)$. □

**Proposição 1.1.3** ([10])**.** *Seja $f : (a, b) \to \mathbb{R}$ uma função diferenciável em $(a, b)$ e contínua em $[a, b]$, se $f'(x) > 0 \;\; \forall\, x \in (a, b)$ então $f$ é estritamente crescente.*

*Demonstração.* Pelo Teorema do Valor Médio, temos que para $x < y$ ; $x, y \in (a, b)$ existe $z \in (x, y)$ tal que $f(x) - f(y) = f'(z)(x - y)$ como $f'(z) > 0$ e $x - y < 0$ então $f(x) - f(y) < 0 \;\Rightarrow\; f(x) < f(y)$ . Logo $f$ é estritamente crescente. □

**Proposição 1.1.4** ([20])**.** *Sejam $I \subset \mathbb{R}$ um intervalo e $\varphi : I \to \mathbb{R}$ uma função convexa. Então dados $a$, $b$ e $c \in I$ com $a < b < c$, temos*

$$\frac{\varphi(b) - \varphi(a)}{b - a} \le \frac{\varphi(c) - \varphi(a)}{c - a} \le \frac{\varphi(c) - \varphi(b)}{c - b}.$$

*Demonstração.* Sejam $a, b, c \in I$ e $\varphi$ convexa, com $a < b < c$, tomando

$$b = \frac{c-b}{c-a}a + \frac{b-a}{c-a}c$$

e usando a convexidade de $\varphi$ temos

$$\varphi(b) \leq \frac{c-b}{c-a}\varphi(a) + \frac{b-a}{c-a}\varphi(c).$$

Daí,

$$\varphi(b) - \varphi(a) \leq \left(\frac{c-b}{c-a} - 1\right)\varphi(a) + \frac{b-a}{c-a}\varphi(c) = \frac{a-b}{c-a}\varphi(a) + \frac{b-a}{c-a}\varphi(c).$$

Então

$$\frac{\varphi(b)-\varphi(a)}{b-a} \leq \frac{\varphi(c)-\varphi(a)}{c-a}.$$

A segunda desigualdade é feita de modo análogo. □

**Proposição 1.1.5** ([20]). *Sejam $\epsilon > 0$ e $\tau \in [0,1]$. Se $\varphi : [0,\epsilon) \to \mathbb{R}$ é convexa, então $l : (0,\epsilon) \to \mathbb{R}$ definida por*

$$l(x) = \frac{\varphi(x) - \varphi(\tau x)}{x}$$

*é não decrescente. Se $\varphi$ é estritamente convexa e $\tau \neq 1$, então $l$ é crescente.*

*Demonstração.* Se $\tau = 1$, então $l(x) = 0$. Dados $x, y \in (0,\epsilon)$ com $0 < x < y$. Agora, se $\tau = 0$, então $l(x) = \frac{\varphi(x)-\varphi(0)}{x}$ e o resultado segue como consequência da primeira desigualdade da proposição anterior tomando $a = 0$, $b = x$ e $c = y$.

Suponha que $\tau \in (0,1)$, então $\tau < 1$ o que implica $\tau x < x < y$. Daí, aplicando novamente a primeira desigualdade da Proposição 1.1.4, temos

$$\frac{\varphi(x)-\varphi(\tau x)}{x-\tau x} \leq \frac{\varphi(y)-\varphi(\tau x)}{y-\tau x}.$$

Novamente $\tau < 1$ implica $\tau x < \tau y < y$ e aplicando a segunda desigualdade da proposição anterior obtemos

$$\frac{\varphi(y)-\varphi(\tau x)}{y-\tau x} \leq \frac{\varphi(y)-\varphi(\tau y)}{y-\tau y}.$$

Das duas últimas desigualdades, segue

$$\frac{\varphi(x) - \varphi(\tau x)}{x - \tau x} \leq \frac{\varphi(y) - \varphi(\tau x)}{y - \tau x} \leq \frac{\varphi(y) - \varphi(\tau y)}{y - \tau y}.$$

Portanto $l(x) \leq l(y)$ para todo $x < y \in (0, \epsilon)$, que prova a proposição. Agora, se $\varphi$ é estritamente convexa as desigualdades na Proposição 1.1.4 valem para desigualdades estritas e analogamente concluímos que $l$ é crescente. $\square$

## 1.2 Espaços de dimensão infinita

Nesta seção, veremos algumas definições e exemplos de espaços métricos. Veremos ainda, definições e proposições de análise dos espaços de Banach e Hilbert. Em seguida definiremos a derivada de Frechét e suas propriedades. Encerraremos esta seção com dois lemas importantes que servirão como auxilio dos capítulos posteriores.

### 1.2.1 Espaço de Banach

Nesta subseção, apresentaremos definições e proposições sobre espaços de dimensão infinita que serão úteis nos capítulos seguintes. Mais detalhes podem ser visto em [9] e [25].

Uma métrica num conjunto $M$ é uma função $d : M \times M \to \mathbb{R}$, que associa a cada par ordenado de elementos $x, y \in M$ um número real $d(x, y)$, chamado a distância de $x$ a $y$, de modo que sejam satisfeitas as seguintes condições para quaisquer $x, y, z \in M$:

1. $d(x, x) = 0$;
2. Se $x \neq y$ então $d(x, y) > 0$;
3. $d(x, y) = d(y, x)$;
4. $d(x, z) \leq d(x, y) + d(y, z)$.

**Definição 1.2.1** ([9]). *Um espaço métrico é um par $(M, d)$, onde $M$ é um conjunto e $d$ é uma métrica em $M$.*

**Exemplo 1.2.1.** *O conjunto $\mathbb{R}$ dos números reais é um espaço métrico. De fato, pois a distância entre dois pontos $x, y \in \mathbb{R}$ é dada por $d(x, y) = |x - y|$. As condições 1 a 4 resultam das propriedades elementares do valor absoluto de números reais.*

**Exemplo 1.2.2.** *O espaço euclidiano* $\mathbb{R}^n$ *é um espaço métrico,com as distâncias definidas abaixo, onde* $x = (x_1, x_2, ..., x_n)$ *e* $y = (y_1, y_2, ..., y_n)$ *escrevemos:*
$d(x, y) = \sqrt{(x_1 - y_1)^2 + ... + (x_n - y_n)^2} = [\sum_{i=1}^{n}(x_i - y_i)^2]^{1/2}$,
$d'(x, y) = |x_1 - y_2| + ... + |x_n - y_n|$ *e*
$d''(x, y) = \max\{|x_1 - y_1|, ..., |x_n - y_n|\} = \max\{x_i - y_i\}$.
*Este exemplo generaliza o anterior.*

**Definição 1.2.2** ([9]). *Uma sequência* $(x_n)$ *num espaço métrico* $X$ *chama-se sequência de Cauchy quando, para todo* $\epsilon > 0$ *dado, existe* $n_0 \in \mathbb{N}$ *tal que*

$$m, n > n_0 \Longrightarrow d(x_m, x_n) < \epsilon.$$

Um resultado conhecido, que pode ser encontrado em [9], é que toda subsequência de uma sequência de Cauchy é também de Cauchy.

**Definição 1.2.3** ([9]). *Um espaço vetorial normado é uma par* $(E, \|.\|)$*, onde* $E$ *é um espaço vetorial e* $\|.\|$ *é uma norma em* $E$*.*

*Todo espaço vetorial normado* $(E, \|.\|)$ *torna-se um espaço métrico por meio da definição* $d(x, y) = \|x - y\|$*. Esta métrica diz-se proveniente da norma* $\|.\|$*.*

**Definição 1.2.4** ([9]). *Sejam* $(X, \|.\|)$ *um espaço vetorial normado e* $\{x_n\} \subset X$ *uma sequência. Dizemos que* $\{x_n\}$ *converge para* $x_* \in \Omega$ *se dado* $\epsilon > 0$*, existe* $n_0$ *tal que*

$$\|x_n - x_*\| < \epsilon \qquad \forall n \geq n_0$$

Um espaço métrico $M$ é completo quando toda sequência de Cauchy em $M$ é convergente e converge para um ponto de $M$.

**Definição 1.2.5** ([9]). *Um espaço vetorial normado completo é chamado espaço de Banach.*

**Exemplo 1.2.3.** *O espaço das funções contínuas reais definidas no intervalo* $[0, 1]$ *é um espaço de Banach com a norma do supremo:* $\|f\| = sup_{x\in[0,1]}|f(x)|$.

**Lema 1.2.1** ([25]). *Sejam* $X, Y$ *espaços de Banach, e* $\mathcal{L}(X, Y)$ *o espaço de operadores lineares de* $X$ *em* $Y, T \in \mathcal{L}(X, Y)$*. Definimos a norma de operadores* $\|\cdot\|$ *como*

$$\|T\| := \sup_{\|x\|\neq 0} \frac{\|Tx\|}{\|x\|}. \tag{1.7}$$

*Pela definição de supremo. Valem as seguintes desigualdades* *(i)* $\|Tu\| \leq \|T\|\|u\|$ *para todo* $T \in \mathcal{L}(X,Y)$ *e* $u \in X$*,*

*(ii)* $\|ST\| \leq \|S\|\|T\|$ *e* $\|S+T\| \leq \|S\| + \|T\|$ *para todo* $S, T \in \mathcal{L}(X,Y)$*,*

*(iii)* $\|T^k\| \leq \|T\|^k$ *para todo* $T \in \mathcal{L}(X,Y)$*.*

*Demonstração.* Primeiro provaremos o item $(i)$, para isto, se $x = 0$ segue imediato de (1.7). Se $x \neq 0$, considere o vetor $y = \frac{x}{\|x\|}$ e usando (1.7) temos

$$\|T\| \geq \|Ty\| = \frac{1}{\|x\|}\|Tx\|.$$

Portanto $\|Tx\| \leq \|T\|\|x\|$. Agora, De (1.7), item $(i)$ e propriedades do supremo, temos que

$$\|TS\| = \sup_{\|x\| \neq 0} \frac{\|TSx\|}{\|x\|} \leq \sup_{\|x\| \neq 0} \frac{\|T\|\|Sx\|}{\|x\|} = \|T\|\|S\|.$$

Para finalizar o lema, observamos que a prova do item $(iii)$ é consequência imediata do item $(ii)$. □

*Com a norma definida acima no espaço* $\mathcal{L}(X,Y)$ *,* $(\mathcal{L}(X,Y), \|\cdot\|)$ *é um espaço de Banach. Esta propriedade é facilmente verificada devido a norma definida em (1.7).*

### 1.2.2 Espaço de Hilbert

Nesta subseção apresentaremos a definição de espaço de Hilbert e daremos alguns exemplos.

**Definição 1.2.6** ([9])**.** *Dizemos que* $H$ *é um espaço de Hilbert se* $H$ *é um espaço de Banach com a norma derivada do produto interno.*

**Exemplo 1.2.4.** *O espaço* $\mathbb{R}^n$ *com o produto interno* $< x, y >= \sum_{i=1}^{\infty} x_i y_i$ *é um espaço de Hilbert.*

**Exemplo 1.2.5** ([9])**.** *Um dos exemplos importantes de espaço de Hilbert é o espaço* $l^2$*, o espaço das sequências de quadrado somável. Este espaço é constituído por todas as sequências* $x = (x_1, x_2, ..., x_i, ...)$ *de números reais ou complexos tais que* $\sum_{i=1}^{\infty} x_i^2 < +\infty$*.*

A seguir apresentaremos a definição dos tipos de convergência, cuja definição no espaço euclidiano $\mathbb{R}^n$ pode ser vista em [1]. Porém, a definição também é valida em um contexto mais geral conforme veremos a seguir.

**Definição 1.2.7.** *Seja $X$ um espaço de Hilbert e uma sequência $\{x_k\} \subset X$ convergente para $x_*$, então:*
*(i) $\{x_k\}$ converge linearmente para $x_*$, se existe $c \in [0,1)$, tal que*

$$\|x_{k+1} - x_*\| \leq c\|x_k - x_*\| \quad \forall\ k \geq 0,$$

*(ii) $\{x_k\}$ converge quadraticamente para $x_*$, se existe $c > 0$, tal que*

$$\|x_{k+1} - x_*\| \leq c\|x_k - x_*\|^2 \quad \forall\ k \geq 0,$$

*(iii) $\{x_k\}$ converge superlinearmente para $x_*$, se existe uma sucessão $(c_k)$ de termos não-negativos e convergentes para zero, tal que $\|x_{k+1} - x_*\| \leq c_k\|x_k - x_*\| \quad \forall\ k\ \geq 0$, ou seja,*

$$\lim_{k\to+\infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0.$$

**Exemplo 1.2.6.** *A sequência $z_k = \frac{1}{2^{2^k}}$ converge quadraticamente para $0$. De fato, pois*

$$\frac{\|z_{k+1}\|}{\|z_k\|^2} = \frac{(2^{2^k})^2}{2^{2^{k+1}}} = 1.$$

**Exemplo 1.2.7.** *A convergência da sequência $\left\{\frac{1}{k!}\right\}$ é superlinear. De fato, pois como $\displaystyle\lim_{k\to\infty} \frac{1}{k+1} = 0$ temos*

$$\lim_{k\to\infty} \frac{\frac{1}{(k+1)!}}{\frac{1}{k!}} = \lim_{k\to\infty} \frac{1}{k+1} = 0.$$

A convergência quadrática implica na convergência superlinear, no entanto, a recíproca é falsa. Basta observar que a sequência do Exemplo 1.2.7 não converge quadraticamente.

### 1.2.3 Derivada de Fréchet e tópicos de operadores lineares

A seguir definiremos a derivada de Fréchet para operadores em espaços de Banach. Além disso, apresentaremos o Teorema Fundamental do Cálculo e revisaremos alguns tópicos de operadores lineares e norma de operador. Para mais detalhes indicamos as referências [11], [12] e [25]

**Definição 1.2.8.** *Um operador* $T \in \mathcal{L}(X, Y)$ *é a derivada de Fréchet de* $F$ *em* $x_0 \in \Omega$ *se para todo* $u \in X$ *tal que* $u + x_0 \in \Omega$ *vale:*

$$\lim_{\|u\| \to 0} \frac{1}{\|u\|} \|F(x_0 + u) - F(x_0) - Tu\| = 0.$$

Dizemos que $F$ é Fréchet derivável em $\Omega$ se $F$ for Fréchet derivável em todo ponto $x \in \Omega$. Denotaremos a derivada de Fréchet de um operador $F$ em $x$ por $F'(x)$ e assim $F'(x) \in \mathcal{L}(X, Y)$, ou seja, $F'$ é uma aplicação linear de $\Omega$ no espaço dos operadores lineares $\mathcal{L}(X, Y)$. Se $F' : X \to \mathcal{L}(X, Y)$ é contínua em $\Omega$ dizemos que $F$ é de classe $C^1$ em $\Omega$, e usamos a seguinte notação $F \in C^1(\Omega)$. Se o operador $F'$ for derivável em $\Omega$, e o operador $F'' : X \to \mathcal{L}(X, \mathcal{L}(X, Y))$ for contínuo, então dizemos que $F$ é de classe $C^2$ em $\Omega$ e denotamos por $F \in C^2(\Omega)$.

Se $F'$ é uma função diferenciável em $\Omega$, a sua derivada é chamada de segunda derivada e será denotada por $F''$. Assim, $F''(x)$ é um elemento do espaço $\mathcal{L}(X, \mathcal{L}(X, Y))$ dos operadores lineares que levam $X$ em $\mathcal{L}(X, Y)$. Para simplificar a notação identificaremos o espaço $\mathcal{L}(X, \mathcal{L}(X, Y))$ por $\mathcal{L}_2(X, Y)$.

Podemos generalizar o conceito acima, supondo que a n-ésima derivada de uma função $F$ de $X$ em $Y$ é a derivada da derivada de ordem $n - 1$. Assim, a n-ésima derivada, denotada por $F^{(n)}(x)$, será um elemento do espaço $\mathcal{L}_n(X, Y)$ dos operadores n-lineares de $X$ em $Y$. Novamente identificamos o espaço $\mathcal{L}(X, \mathcal{L}(X, ..., Y), ...))$ com o espaço dos operadores n-lineares $\mathcal{L}_n(X, Y)$.

Utilizando as notações acima temos a seguinte observação:

**Observação.** Seja $F : \Omega \to Y$ uma função diferenciável no conjunto aberto $\Omega \subset X$. Então $F'(x) \in \mathcal{L}(X, Y)$ e

$$\|F'(x)\| := \sup \left\{ \|F'(x)u\|; \|u\| \leq 1 \right\}.$$

Agora, se $F'$ é uma função diferenciável no conjunto aberto $\Omega$, então $F''(x \in \mathcal{L}_2(X, Y))$ e

$$\|F''(x)\| := \sup \left\{ \|F''(x)(u, v)\|; \|u\| \leq 1, \|v\| \leq 1 \right\}.$$

Analogamente, se $F$ é uma função n-vezes diferenciável em $\Omega$, introduzimos a norma do operador n-linear $F^{(n)} \in \mathcal{L}_n(X, Y)$ como sendo

$$\|F^{(n)}(x)\| := \sup \left\{ \|F^{(n)}(x)(u_1, ..., u_n)\|; \|u_1\| \leq 1, ..., \|u_n\| \leq 1 \right\}.$$

Para o desenvolvimento dessa dissertação, apresentamos a seguir o Teorema Fundamental do Cálculo e algumas consequências.

**Proposição 1.2.1** ([12])**.** *(Teorema Fundamental do Cálculo) Seja $F : [a,b] \to X$ continuamente diferenciável, então*

$$\int_a^s F(t)'dt = F(s) - F(a) \qquad \forall\ s \in [a,b].$$

*Demonstração.* A prova desta proposição pode ser encontrada em [12]. $\square$

**Proposição 1.2.2** ([12])**.** *Sejam $X$ e $Y$ espaços de Banach, $F : U \to Y$ um operador aberto e convexo e $F : U \to Y$ um operador Fréchet derivável. Para todo $x, y \in U$ temos que*

$$\int_0^1 \frac{d}{dt}[F(x + t(y-x))]dt = F(y) - F(x).$$

*Demonstração.* Segue imediatamente da Proposição 1.2.1. $\square$

As Propriedades 1.2.1 e 1.2.2 também são válidas em espaço de Hilbert.

A seguir apresentaremos dois lemas a respeito dos operadores lineares em espaço de Hilbert.

**Lema 1.2.2** ([25])**.** [Lema de Banach] *Sejam o operador linear $T \in \mathcal{L}(X,Y)$ e $I$ o operador identidade em um espaço de Hilbert, tal que $\|T\| < 1$. Então $I - T$ é invertível e vale*

$$\|(I-T)^{-1}\| \leq \frac{1}{1 - \|T\|}.$$

*Demonstração.* Sejam $\{S_k\}$ e $\{t_k\}$ sequências definidas respectivamente por

$$S_k = I + T + T^2 + ... + T^k; \qquad t_k = 1 + \|T\| + ... + \|T^k\|.$$

Note que

$$\|S_{k+1} - S_k\| = \|T^{k+1}\| \leq \|T\|^{k+1} = t_{k+1} - t_k,$$

onde a desigualdade é proveniente do item $(iii)$ do Lema 1.2.1.

Agora, como $\|T\| < 1$ e $t_{k+1} - t_k = \|T\|^{k+1} \geq 0$, o que implica $t_{k+1} \geq t_k$ temos que $\{t_k\}$ é uma sequência monótona crescente e convergente, com o limite $t^* = \frac{1}{1-\|T\|}$.

Consequentemente $\{S_k\}$ é uma sequência de Cauchy em $\mathcal{L}(X,Y)$ e assim existe $\lim\limits_{k\to\infty} S^k$.

Agora, observe que

$$S_k(I-T) = (I + T + ... + T^k)(I-T) = I - T^{k+1} \tag{1.8}$$

Como $\|I-(I-T^k)\| = \|T^k\| \leq \|T\|^k$ e $\lim_{k\to\infty} \|T\|^k = 0$ (pois $\|T\| < 1$). Então $\lim_{k\to\infty} I - T^k = I$.
Assim, da expressão (1.8) e do último limite obtemos

$$\lim_{k\to\infty} S_k(I-T) = \lim_{k\to\infty}(I - T^{k+1}) = I.$$

Daí, como existe $\lim_{k\to\infty} S_k$,

$$(I-T)\lim_{k\to\infty} S_k = I \quad \Rightarrow \quad \lim_{k\to\infty} S_k = (I-T)^{-1},$$

ou seja, $I-T$ é invertível.

a Agora note que

$$\begin{aligned} \|(I-T)^{-1}\| &= \|\lim_{k\to\infty} S_k\| \\ &= \|\lim_{k\to\infty}(I + T + \dots + T^k)\| \\ &\leq \lim_{k\to\infty}(\|I\| + \|T\| + \dots + \|T^k\|) \\ &\leq \lim_{k\to\infty}(\|I\| + \|T\| + \dots + \|T\|^k) \\ &= \lim_{k\to\infty} t^k = \frac{1}{1-\|T\|}, \end{aligned}$$

que prova a última parte do lema. $\square$

**Lema 1.2.3** ([25])**.** *Sejam $B \in \mathcal{L}(X,X)$ e $I$ o operador identidade em $X$. Se $\|B-I\| < 1$, então $B$ é invertível e vale $\|B^{-1}\| \leq 1/(1-\|B-I\|)$.*

*Demonstração.* Segue do lema anterior, com $A = I$ e $c = \|B-I\|$. $\square$

## 1.3 Método clássico de Newton

Nosso objetivo neste seção é apresentar a definição do método de Newton.

O método de Newton consiste em aproximar o zero da função não-linear por zeros de aproximações lineares sucessivas. Mais precisamente, queremos resolver a equação

$$F(x) = 0, \quad x \in \Omega \tag{1.9}$$

onde $\Omega \subset X$ é um conjunto aberto e $F : \Omega \to Y$ é uma função continuamente diferenciável não-linear. Começamos com um ponto inicial $x_0 \in \Omega$ podemos definir uma sequência $\{x_k\}$ contida em $\Omega$ dada por

$$x_{k+1} = x_k - F'(x_k)^{-1}F(x_k), \quad k = 0, 1, .... \tag{1.10}$$

Para esta sequência fazer sentido é necessário que todos os $\{x_k\}$ estejam contidos no domínio da $F$ e que $F'(x_k) \neq 0$, para todo $k \geq 0$. portanto necessitamos de condições adequadas sobre $F$ e o ponto inicial $x_0$ para garantir que $\{x_k\}$ converge para um zero de $F$.

No próximo capítulo estaremos interessados em analisar a convergência do Método de Gauss-Newton que é uma generalização dos resultados do Método de Newton para resolver o sistema de equações não-lineares $F(x) = 0$, quando $F'(x_*)$ é invertível, o método que analisaremos no Capítulo 2 é similar ao Teorema 2 em [29]. No caso em que $F'(x_*)$ não é invertível, mas é injetiva usamos o método de Gauss-Newton. Nesta dissertação será usada inversa generalizada de Moore-Penrose.

A seguir definimos e daremos resultados a respeito da inversa generalizada de Moore-Penrose que serão necessários mais tarde, para garantir a boa definição do método de Gauss- Newton.

## 1.4 Inversa generalizada

Iniciaremos esta seção com algumas definições básicas de matrizes. Em seguida apresentaremos a definição da inversa de Moore-Penrose e algumas propriedades da inversa de Moore-Penrose.

O conjunto de todas as matrizes $m \times n$ ($m$ linhas e $n$ colunas) com entradas complexas será denotado por $M(\mathbb{C}, m, n)$. O conjunto de todas as matrizes quadradas $n \times n$ com entradas complexas será denotado simplesmente por $M(\mathbb{C}, n)$.

Uma matriz $A \in M(\mathbb{C}, m, n)$ é frequentemente representada na forma de um arranjo como

$$\begin{pmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \dots & A_{mn} \end{pmatrix}.$$

$M(\mathbb{C}, m, n)$ é um espaço vetorial complexo, com a operação de soma e de multiplicação por escalar(complexos) definidos por:

$$(A_1 + A_2)_{ij} := (A_1)_{ij} + (A_2)_{ij},$$

$$(aA)_{ij} := aA_{ij},$$

tal que $\alpha \in \mathbb{C},\ A,\ A_1\ e\ A_2 \in M(\mathbb{C}, m, n)$ e $i \in \{1, ..., m\}, j \in \{1, ..., n\}$.

Dada uma matriz $A \in M(\mathbb{C}, m, n)$ denotamos por $A^T$ a matriz de $M(\mathbb{C}, n, m)$ cujos elementos são dados por $(A^T)_{ij} = A_{ji}$ para todos $i \in \{1, ..., n\}, j \in \{1, ..., m\}$. A matriz $A^T$ é dita a matriz transposta de $A$. Temos que $(A^T)^T = A$. Além disso,

$m, n, p \in \mathbb{N}$ vale, pela regra de produto de matrizes, a relação $(AB)^T = B^T A^T$ para quaisquer $A \in M(\mathbb{C}, m, n)$ e $B \in M(\mathbb{C}, n, p)$.

Uma matriz $A \in M(\mathbb{C}, n)$ define uma aplicação linear de $\mathbb{C}^n$ sobre si mesmo. Se essa aplicação for bijetora, então existe uma aplicação inversa, denotada por $A^{-1} : \mathbb{C}^n \to \mathbb{C}^n$, tal que $A^{-1}(Ax) = x$ para todo $x \in \mathbb{C}^n$. Seja $\mathbb{C}^n$ um espaço vetorial dotado de um produto escalar $\langle \cdot, \cdot \rangle$ e seja $A : \mathbb{C}^n \to \mathbb{C}^n$ um operador linear. Um operador linear $A^*$ que para todos $u, v \in \mathbb{C}^n$ satisfaça:

$$\langle u, Av \rangle = \langle A^* u, v \rangle$$

é chamado o operador adjunto de $A$. O operador adjunto $A^*$ de $A$ é representado (na base canônica) por uma matriz cujos elementos de matriz são $(A^*)_{ij} = \bar{A}_{ji}$, com $i, j \in \{1, ..., n\}$. Suponha que H é um espaço de Hilbert, com o produto interno $\langle \cdot, \cdot \rangle$. Considere uma operador linear contínuo $A : H \to H$ (isso é o mesmo que um operador linear limitado). Do teorema da representação de Riesz, ganhamos que existe um operador linear contínuo único $A^* : H \to H$ com a seguinte propriedade:

$$\langle Ax, y \rangle = \langle x, A^* y \rangle \qquad \forall \; x, y \in H.$$

Esse operador $A^*$ é o adjunto de $A$. Isto pode ser visto como uma generalização da matriz adjunta, cuja demonstração pode ser encontrada em [2].

**Proposição 1.4.1** ([25])**.** *Sejam $A$ e $B$ operadores em espaço de Hilbert, temos as seguintes afirmações:*

1. $(A^*)^* = A$,
2. *Se $A$ é inversível, então assim também o é $A^*$, com* $(A^*)^{-1} = (A^{-1})^*$,
3. $(A + B)^* = A^* + B^*$ *, onde $A$ e $B$ são operadores contínuos de $H$,*
4. $(\lambda A)^* = \lambda^* A^*$*, onde $\lambda^*$ denota o conjugado do número complexo $\lambda$,*
5. $(AB)^* = B^* A^*$*, onde $A$ e $B$ são operadores contínuos de $H$.*

Um determinado operador $A : H \to H$ é chamado Hermitiano, ou auto-adjunto, se

$$A = A^*$$

o qual é equivalente à

$$\langle Ax, y \rangle = \langle x, Ay \rangle \;\; \forall x, y \in H.$$

Agora, estamos aptos a definir a inversa de Moore-Penrose, cuja definição e propriedades podem ser encontrados [2], [8] e [34].

**Definição 1.4.1** ([34]). *(Definição da inversa generalizada de Moore) Seja $A \in \mathcal{L}(X,Y)$ com imagem fechada então $A^\dagger$ é o único operador em $\mathcal{L}(Y,X)$ satisfazendo*

$$A^\dagger A = P_{R(A^*)}, \quad AA^\dagger = P_{R(A)}$$

*onde $P_M$ é a projeção ortogonal em $M$.*

E. H. Moore foi o primeiro que definiu de forma explícita a inversa generalizada de uma matriz arbitrária, sua definição foi dada em 1935. O trabalho de Moore foi pouco notado devido a sua notação complicada. Então o assunto das inversas generalizadas permaneceu pouco desenvolvida. Em 1949 Tseng em uma série de três (mas novamente pouco notada) artigos definiu uma inversa generalizada para operadores lineares no espaço de Hilbert. A Teoria da inversa generalizada foi rapidamente desenvolvida após o trabalho de Penrose em 1955, que aparentemente estava inconsciente do trabalho de Moore. Em [34], Campbell e Meyer Jr apresentam uma prova da equivalência dessas definições. Desde então esta inversa generalizada é chamada de inversa de Moore-Penrose.

**Definição 1.4.2** ([34]). *(Definição de Penrose da inversa generalizada) Se $A \in \mathcal{L}(X,Y)$ tem imagem fechada, a inversa de Moore-Penrose de $A$ é o operador linear $A^\dagger \in \mathcal{L}(Y,X)$ tal que*

1. *$AA^\dagger A = A$,*
2. *$A^\dagger AA^\dagger = A^\dagger$,*
3. *$AA^\dagger \in M(\mathbb{C},m)$ e $A^\dagger A \in M(\mathbb{C},n)$ são auto-adjuntas, ou seja, $(AA^\dagger)^* = AA^\dagger$ e $(A^\dagger A)^* = A^\dagger A$.*

Denotamos $P_{N(A)}$ e $P_{R(A)}$ os operadores ortogonais sobre núcleo e imagem de A respectivamente, daí temos

$$AA^\dagger = I - P_{N(A)}, \quad A^\dagger A = P_{R(A)}.$$

Quando $A$ é injetiva, $N(A) = \{0\}$ e $A^\dagger A = I$, então existe uma única $A^\dagger$ inversa à esquerda de $A$. Além disso, para $A \in B(X,Y)$ as seguintes afirmações são equivalentes

(i) $A$ é injetiva e a imagem é fechada,

(ii) $A^*A$ é invertível em $B(X,X)$.

**Exemplo 1.4.1.** *Se* $A \in M(\mathbb{C},m,1)$, $A = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}$, *um vetor-coluna não nulo, então* $A^\dagger = \frac{1}{\|A\|_{\mathbb{C}}^2 A^*} = \frac{1}{\|A\|_{\mathbb{C}}^2}(\bar{a}_1,...,\bar{a}_m)$, *onde* $\|A\|_{\mathbb{C}} = \sqrt{|a_1|^2+...+|a_m|^2}$. *Observe que se* $z \in \mathbb{C}$, *podemos considerar* $z$ *como uma matriz complexa* $1\times\ 1$, *ou seja, como elementos de* $M(\mathbb{C},1,1)$ *e, com isso, obtemos*

$$(z)^\dagger = \begin{cases} 0 & , \ se \quad z=0 \\ \frac{1}{z} & , \ se \quad z\neq 0 \end{cases}.$$

Temos diretamente da definição da inversa de Moore-Penrose que, se $A \in M(\mathbb{C},n)$ é uma matriz quadrada invertível, então $A^\dagger$ coincide com a $A^{-1}$. Além disso, para $O_{mn}$, a matriz $mxn$ identicamente nula, vale $(O_{mn})^\dagger = O_{nm}$.

Seja $A \in M(\mathbb{C},m,n)$ é possível mostrar que existe uma única inversa de Moore-Penrose. A demonstração da existência pode ser encontrada em [2]. Apresentaremos a demonstração da unicidade.

**Proposição 1.4.2** ([2])**.** *Seja* $A \in M(\mathbb{C},m,n)$, *então* $A$ *possui uma única inversa de Moore-Penrose.*

*Demonstração.* Seja $A^\dagger \in M(\mathbb{C},n,m)$ uma inversa de Moore-Penrose de $A \in M(\mathbb{C},m,n)$ e seja $B \in M(\mathbb{C},n,m)$ uma outra inversa de Moore-Penrose de $A$, ou seja, tal que $ABA = A$, $BAB = B$ com $AB$ e $BA$ auto-adjuntas.

Seja $M_1 := AB - AA^\dagger = A(B-A^\dagger) \in M(\mathbb{C},m)$. Pelas hipóteses, $M_1$ é auto-adjunta, pois é a diferença de duas matrizes auto-adjuntas e

$$(M_1)^2 = (AB-AA^\dagger)A(B-A^\dagger) = (ABA-AA^\dagger A)(B-A^\dagger) = (A-A)(B-A^\dagger) = 0.$$

Como $M_1$ é auto-adjunta, o fato que $(M_1)^2 = 0$ implica $M_1 = 0$, pois para todo $x \in \mathbb{C}^m$ temos

$$\|M_1x\|_{\mathbb{C}}^2 = \langle M_1x, M_1x\rangle_{\mathbb{C}} = \langle x, (M_1)^2x\rangle_{\mathbb{C}} = 0,$$

o que implica $M_1 = 0$. Isso prova que $AB = AA^\dagger$.

Analogamente, prova-se que $AB = AA^\dagger$, neste caso considere a matriz auto-adjunta

$$M_2 := BA - A^\dagger A \in M(\mathbb{C},n)$$

e proceda como no caso anterior. Assim, obtemos:

$$A^\dagger = A^\dagger AA^\dagger = A^\dagger(AA^\dagger) = A^\dagger AB = (AA^\dagger)B = BAB = B,$$

provando a unicidade. □

As seguintes propriedades da inversa de Moore-Penrose seguem da definição e unicidade. Apresentaremos a prova de alguns itens.

**Proposição 1.4.3** ([2]). *Para $A^\dagger \in M(\mathbb{C}, m, n)$ valem:*

1. $(A^\dagger)^\dagger = A$,
2. $(A^\dagger)^T = (A^T)^\dagger$, $\bar{A}^\dagger = (\bar{A})^\dagger$ *e , consequentemente,* $(A^*)^\dagger$,
3. $(zA)^\dagger = z^{-1}A^\dagger$ *para todo* $z \in \mathbb{C}$ *não-nulo,*
4. $A^\dagger = A^\dagger (A^\dagger)^* A^*$,
5. $A = AA^*(A^\dagger)^*$
6. $A^* = A^*AA^\dagger$
7. $A^* = A^\dagger AA^*$.

Faremos a prova dos itens 4, 5, 6.

*Demonstração.* Como $AA^\dagger$ é auto-adjunta, vale $AA^\dagger(AA^\dagger)^* = (A^\dagger)^*A^*$. Multiplicando-se a esquerda por $A^\dagger$ obtemos $A^\dagger(A^\dagger)A^*$, que prova o item 4. Substituindo $A$ por $A^\dagger$ e usando o fato que $A = (A^\dagger)^\dagger$, obtemos do item 4 que $A = AA^*(A^\dagger)^*$ provando o item 5. Substituindo-se $A$ por $A^*$ e usando o item 2 onde $(A^*) = (A^\dagger)^*$, obtemos do item 5 que $A^* = A^*AA^\dagger$, onde prova o item 6. □

**Proposição 1.4.4** ([2]). *Seja $A \in M(\mathbb{C}, m, n)$. Se $(A^*A)^{-1}$ existe, então $A^\dagger = (A^*A)^{-1}A^*$.*

*Demonstração.* Como $(A^*A)^{-1}$ existe, então multiplicando por $(A^*A)^{-1}$ no lado direito do item 6 da proposição anterior temos que, $(A^*A)^{-1}A^* = (A^*A)^{-1}A^*AA^\dagger$ que implica $A^\dagger = (A^*A)^{-1}A^*$. □

**Exemplo 1.4.2.** *Seja* $A = \begin{pmatrix} 1 & 2 \\ 0 & i \\ 0 & 3 \end{pmatrix}$, *calcule* $A^\dagger$.

*Como* $A^* = \begin{pmatrix} 1 & 0 & 0 \\ 2 & -i & 3 \end{pmatrix}$, *temos que*

$$A^*A = \begin{pmatrix} 1 & 0 & 0 \\ 2 & -i & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & i \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 14 \end{pmatrix}.$$

*Então, obtemos* $(A^*A)^{-1} = \begin{pmatrix} \frac{7}{5} & \frac{-2}{10} \\ \frac{-1}{5} & \frac{1}{10} \end{pmatrix}$, *e portanto,*

$$A^\dagger = (A^*A)^{-1}A^* = \begin{pmatrix} \frac{7}{5} & \frac{-2}{10} \\ \frac{-1}{5} & \frac{1}{10} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 2 & -i & 3 \end{pmatrix} = \frac{1}{10} \begin{pmatrix} 1 & 2i & -6 \\ 0 & -i & 3 \end{pmatrix},$$

*Logo* $A^\dagger = \frac{1}{10} \begin{pmatrix} 10 & -2i & -6 \\ 0 & -i & 3 \end{pmatrix}$.

A seguir apresentaremos alguns lemas a respeito da inversa generalizada de Moore-Penrose que serão necessários mais tarde, para garantir a boa definição do método de Gauss-Newton.

**Lema 1.4.1** ([2])**.** *Sejam $A, B \in \mathcal{L}(X, Y)$ tais que $A$ e $B$ têm imagens fechadas. Se $A$ é injetiva, $E = B - A$ e $\|EA^\dagger\| < 1$, então $B$ é injetiva.*

*Demonstração.* Como $B = A + E = (I + EA^\dagger)A$ e da hipótese $\|EA^\dagger\| < 1$, onde $\|EA^\dagger\| = \|I + EA^\dagger - I\| < 1$, segue do Lema 1.2.3 que $I + EA^\dagger$ é invertível. Portanto, $B$ é injetiva. □

O próximo lema é provado em Stewart [13] (ver também, Wedin [33]) para matrizes $A$ e $B$ de ordem $mxn$ com $m \geq n$ e $posto(A) = posto(B) = n$. Porém, o resultado também é valido em um contexto mais geral conforme veremos a seguir.

**Lema 1.4.2.** *Sejam $A, B \in \mathcal{L}(X, Y)$ tais que $A$ e $B$ são injetivos e têm imagens fechadas. Assuma que $E = B - A$ e $\|A^\dagger\|\|E\| < 1$, então*

$$\|B^\dagger\| \leq \frac{\|A^\dagger\|}{1 - \|A^\dagger\|\|E\|}, \qquad \|B^\dagger - A^\dagger\| \leq \frac{\sqrt{2}\|A^\dagger\|^2\|E\|}{1 - \|A^\dagger\|\|E\|}.$$

## 1.5 Função majorante e propriedades

Nesta seção, definiremos a função majorante e estudaremos algumas propriedades relacionadas a ela. Os resultados apresentados aqui são os principais instrumentos no estudo de convergência local do método de Gauss-Newton para problemas de mínimos quadrados não-lineares. Dividiremos esta seção em duas subseção, onde na primeira apresentaremos os resultados usando a hipótese de que $f$ majorante possui derivada convexa (conforme em [32]) e na segunda subseção não é considerado essa hipótese (conforme [27]).

**Definição 1.5.1.** *Sejam $X, Y$ espaços de Hilbert, $\Omega \subseteq X$ um conjunto aberto e $F : \Omega \to Y$ uma função continuamente diferenciável tal que $F'(x)$ tem imagem fechada em $\Omega$. Tome $x_* \in \Omega$, $R > 0$,*

$$\beta := \|F'(x_*)^{\dagger}\|, \quad \kappa := sup\{t \in [0, R) : B(x_*, t) \in \Omega\}.$$

*Suponha que $F(x_*) = 0$ e $F'(x_*)$ seja injetiva. Então, uma função continuamente diferenciável $f : [0, R) \to \mathbb{R}$ é uma função majorante de $F$ em $B(x_*, \kappa)$ se satisfaz*

$$\|F'(x) - F'(x_* + \tau(x - x_*))\| \leq f'(\|x - x_*\|) - f'(\tau\|x - x_*\|), \tag{1.11}$$

*para todo $\tau \in [0, 1]$, $x \in B(x_*, \kappa)$ e*

*(h1)* $f(0) = 0$ *e* $f'(0) = -1$*;*

*(h2)* $f'$ *é convexa e estritamente crescente;*

**Exemplo 1.5.1.** *Algumas funções que satisfazem $(h1)$ e $(h2)$, cuja derivada é convexa.*

*(i)* $f : \mathbb{R} \to \mathbb{R}$, *tal que* $f(t) = e^t - 2t - 1$,

*(ii)* $f : [0, 1) \to \mathbb{R}$, *tal que* $f(t) = -\ln(1 - t) - 2t$,

**Exemplo 1.5.2** (Exemplo de função majorante de $F$)**.** *Sejam*

- $F : \mathbb{R} \to \mathbb{R}^2$, *tal que* $F(x) = (x, x^2)^T$ *e*
- $f : \mathbb{R} \to \mathbb{R}$, *tal que* $f(t) = t^2 - t$.

Note que a função $f$ é um exemplo de uma função majorante da $F$, onde $F$ satisfaz todas as hipóteses dada na definição da função majorante e $f$ satisfaz $(h1)$ e $(h2)$.

**Exemplo 1.5.3.** *Algumas funções que satisfazem $h1$ e $(h2)$, cuja derivada é não convexa.*

*i)* $f : [0, +\infty) \to \mathbb{R}$, *tal que* $f(t) = t^{1+p} - t$

*ii)* $f : [0, +\infty) \to \mathbb{R}$, *tal que* $f(t) = e^{-t} + t^2 - 1$

*onde* $0 < p < 1$.

Conhecida a definição de função majorante vista na Definição (1.5.1), apresentaremos as principais propriedades desse tipo de função, cuja derivada é convexa.

### 1.5.1 Propriedades da função majorante com a hipótese de $f'$ convexa

Nesta subseção, estamos interessados em propriedades da função majorante com a hipótese de $f'$ convexa, que serão necessárias para o Capítulo 2.

Veremos a seguir uma proposição da função majorante no qual não utilizaremos a hipótese da derivada convexa, mas que será utilizada na prova de propriedades com $f'$ convexa.

**Proposição 1.5.1.** *Defina a constante $\upsilon := \{t \in [0, R) : 1 - \beta[f'(t) + 1] > 0\}$. Então, $\upsilon$ é positiva e são válidas as seguintes desigualdades*

$$0 < \beta[f'(t) + 1] < 1 \quad \forall t \in (0, \upsilon) \tag{1.12}$$

*e*

$$tf'(t) - f(t) > 0 \;\; \forall \; t \in (0, \upsilon). \tag{1.13}$$

*Demonstração.* Como $f'$ é contínua em $(0, R)$ e $f'(0) = -1$:

$$\lim_{t\to 0}(1 - \beta(f'(t) + 1)) = 1 - \beta(f'(0) + 1) = 1 - \beta(-1 + 1) = 1.$$

Assim, existe $\delta > 0$, tal que $|1 - \beta(f'(t) + 1) - 1| < 1$ para todo $t \in (0, \delta)$.

Daí,
$$1 - \beta(f'(t) + 1) > 0, \quad t \in (0, \delta).$$

Então pela definição de $\upsilon$ temos $\upsilon \geq \delta > 0$, consequentemente, $\upsilon > 0$.
Para prova de (1.12), usamos a definição de $\upsilon$, daí, temos

$$1 - \beta[f'(\upsilon) + 1] > 0 \quad \Rightarrow \quad \frac{1}{\beta} > f'(\upsilon) + 1 \quad \Rightarrow \quad \frac{1}{\beta} - 1 > f'(\upsilon).$$

Como $f'$ é estritamente crescente, temos $f'(\upsilon) > f'(t) \quad \forall \; t \in (0, \upsilon)$. Assim,

$$\frac{1}{\beta} - 1 > f'(t) \Rightarrow \; 1 - \beta(f'(t) + 1) > 0, \quad t \; \in (0, \upsilon),$$

donde
$$\beta(f'(t) + 1) < 1 \quad \forall t \in (0, \upsilon),$$

e como $f'(0) < f'(t) \;\; \forall \; t \; \in (0, \upsilon)$, obtemos

$$-1 < f'(t) \quad \Rightarrow \quad 0 < \beta(f'(t) + 1) \;\; \forall \; t \; \in (0, \upsilon).$$

Agora, provaremos a segunda. De $(h2)$ temos que $f'$ é estritamente crescente em $[0, R)$, então pela Proposição 1.1.2, segue que $f$ é estritamente convexa em $[0, R)$. Assim, pela Proposição 1.1.1, temos:

$$f(0) > f(t) - tf'(t) \quad \forall t \in (0, R). \tag{1.14}$$

Portanto, de $(h1)$,

$$0 = f(0) > f(t) - tf'(t) \;\; \forall\, t \in (0, R).$$

Consequentemente:

$$tf'(t) - f(t) > 0 \;\forall\, t \in (0, \upsilon).$$

$\square$

**Proposição 1.5.2.** *A função* $[0, R) \ni \mapsto 1/[1 - \beta(f'(t) + 1)]$ *é crescente.*

*Demonstração.* Usando o fato de $f'$ ser estritamente crescente, temos para $x < y$

$$\begin{aligned} f'(x) < f'(y) \quad &\Longrightarrow \quad \beta(f'(x)+1) < \beta(f'(y)+1) \\ &\Longrightarrow \quad 1 - \beta(f'(x)+1) > 1 - \beta(f'(y)+1) \\ &\Longrightarrow \quad \frac{1}{1 - \beta(f'(x)+1)} < \frac{1}{1 - \beta(f'(y)+1)}. \end{aligned}$$

Portanto a função dada é crescente. $\square$

**Proposição 1.5.3.** *As seguintes funções são crescentes:*

1. $(0, R) \ni t \mapsto [tf'(t) - f(t)]\,/t^2$;
2. $(0, R) \ni t \mapsto \dfrac{tf'(t) - f(t)}{t^2[1 - \beta(f'(t)+1)]}$.

*Demonstração.* Para provar o item 1, usaremos $(h1)$, obtemos:

$$\begin{aligned} \int_0^1 \frac{f'(t) - f'(\tau t)}{t} d\tau &= \int_0^1 \frac{tf'(t)}{t^2} d\tau - \int_0^1 \frac{tf'(\tau t)}{t^2} d\tau \\ &= \frac{tf'(t)}{t^2} - \frac{1}{t^2} \int_0^1 [f(\tau t)]' d\tau \\ &= \frac{tf'(t)}{t^2} - \frac{f(t)}{t^2} \\ &= \frac{tf'(t) - f(t)}{t^2}. \end{aligned}$$

Da Proposição 1.1.5, tomando $f' = \varphi$, usando o fato de $f'$ ser convexa, concluímos que a função em consideração é crescente.

Para provar o item 2, basta usar o fato de $[0, R) \ni \mapsto 1/[1 - \beta(f'(t) + 1)]$ ser crescente dada na Proposição 1.5.2 , e o item 1 desta proposição.

$\square$

**Proposição 1.5.4.** *Defina a constante*

$$\rho := \sup \left\{ t \in (0, \upsilon) : \frac{\beta[tf'(t) - f(t)]}{t[1 - \beta(f'(t) + 1)]} < 1 \right\}.$$

*Então, $\rho$ é positivo e são válidas as seguintes desigualdades*

$$0 < \frac{\beta[tf'(t) - f(t)]}{t[1 - \beta(f'(t) + 1)]} < 1 \quad \forall t \in (0, \rho). \tag{1.15}$$

*Demonstração.* Usando $(h2)$, temos

$$\frac{\beta[tf'(t) - f(t)]}{t[1 - \beta(f'(t) + 1)]} = \frac{\beta \left[ f'(t) - \frac{f(t) - f(0)}{t - 0} \right]}{1 - \beta(f'(t) + 1)}.$$

Segue da última equação e da continuidade de $f'$ que

$$\lim_{t \to 0} \frac{\beta[tf'(t) - f(t)]}{t[1 - \beta(f'(t) + 1)]} = 0,$$

então,

$$\frac{\beta[tf'(t) - f(t)]}{t[1 - \beta(f'(t) + 1)]} < 1, \quad t \in (0, \delta). \tag{1.16}$$

Daí, segue que $\delta \leq \rho$, o que prova a primeira afirmação.
Para provar a primeira desigualdade de (1.15), usamos as desigualdades (1.12) e (1.13) da Proposição 1.5.1.
Agora, a segunda desigualdade (1.12) para a primeira desigualdade segue da definição de $\rho$ e (1.16). $\square$

### 1.5.2 Propriedades da função majorante relaxada

Nesta subseção, estamos interessados em propriedades das funções majorantes sem derivada convexa. Tais propriedades serão necessárias para o Capítulo 3.

**Proposição 1.5.5.** *Sejam* $\sigma := \sup\{t \in (0,\kappa) : \beta[(f(t)/t)+1] < 1\}$ *e* $\kappa$ *como na Definição 1.5.1. Então estas constantes são positivas e, além disso:*

$$\beta\left[(f(t)/t)+1\right] < 1 \ \ \forall\ t \in (0,\sigma). \tag{1.17}$$

*Demonstração.* Como $\Omega$ é aberto e $x_* \in \Omega$, existe $\epsilon > 0$ tal que $B(x_*,\epsilon) \subset \Omega$. Deste modo, pela definição de $\kappa$, temos que $\kappa > 0$.

De $(h1)$, temos que

$$\lim_{t\to 0} \beta\left(\frac{f(t)}{t}+1\right) = \lim_{t\to 0} \beta\left(\frac{f(t)-f(0)}{t-0} - f'(0)\right) = 0.$$

Analogamente, existe $\delta > 0$ tal que $|\beta(\frac{f(t)}{t}+1)| < 1$, para todo $t \in (0,\delta)$. Logo,

$$\beta\left(\frac{f(t)}{t}+1\right) < 1, \quad t \in (0,\delta).$$

Assim, pela definição de $\sigma$ temos $\sigma \geq \delta > 0$, donde $\sigma > 0$.
Observe que:

$$\left[\frac{f(t)}{t}\right]' = \frac{tf'(t)-f(t)}{t^2} > 0, \tag{1.18}$$

onde a positividade é obtida da desigualdade (1.13) da Proposição 1.5.1.
De (1.18), usando a Proposição 1.1.3, concluímos que $\frac{f(t)}{t}$ é crescente.

Por conseguinte, $\frac{f(t)}{t} < \frac{f(\sigma)}{\sigma}$ $\forall\ t \in (0,\sigma)$. A definição de $\sigma$ implica em

$$\beta\left(\frac{f(\sigma)}{\sigma}+1\right) \leq 1 \ \Rightarrow\ \frac{f(\sigma)}{\sigma}+1 \leq \frac{1}{\beta}.$$

Logo

$$\frac{1}{\beta}-1 \geq \frac{f(\sigma)}{\sigma} > \frac{f(t)}{t} \ \Rightarrow\ \frac{f(t)}{t}+1 < \frac{1}{\beta}.$$

Portanto,

$$\beta\left(\frac{f(t)}{t}+1\right) < 1 \ \ \forall\ t \in (0,\sigma),$$

concluindo a demonstração. □

Seja $g_f : [0,\upsilon) \to [0,+\infty)$ uma função contínua, definida por

$$g_f(t) = \frac{\beta[tf'(t) - f(t)]}{1 - \beta(f'(t)+1)}. \tag{1.19}$$

Observe que pela Proposição 1.5.1 $g_f$ está bem definida em $[0,\upsilon)$, pois

$$1 - \beta(f'(t)+1) > 0 \quad \forall\, t \in (0,\upsilon).$$

**Proposição 1.5.6.** *Sejam $g_f$ definida em (1.19) e*

$$\rho := sup\ \{\ \delta \in (0,\upsilon) : \frac{\beta[tf'(t) - f(t)]}{t[1 - \beta(f'(t)+1)]} < 1,\ t \in (0,\delta)\ \},$$

*então temos que* $\lim_{t\to 0} \frac{g_f(t)}{t} = 0,\ \rho > 0$ *e* $0 < g_f(t) < t \quad \forall\, t \in (0,\rho)$.

*Demonstração.* Usando a definição de $g_f(t)$ e $(h1)$, obtemos:

$$\frac{g_f(t)}{t} = \frac{\beta[tf'(t) - f(t)]}{t[1 - \beta(f'(t)+1)]} = \frac{\beta(f'(t) - \frac{f(t)-f(0)}{t-0})}{1 - \beta(f'(t)+1)}. \tag{1.20}$$

Usando a continuidade de $f'$, passando ao limite em (1.20) e novamente usando $(h1)$, segue que

$$\lim_{t\to 0} \frac{g_f(t)}{t} = \frac{\beta[f'(0) - f'(0)]}{1 - \beta(f'(0)+1)} = \frac{0}{1} = 0 \quad \Rightarrow \quad \lim_{t\to 0} \frac{g_f(t)}{t} = 0.$$

Como $\lim_{t\to 0} \frac{g_f(t)}{t} = 0$, existe $\delta \in (0,\upsilon)$ tal que $\frac{g_f(t)}{t} < 1$ para todo $t \in (0,\delta)$. Daí, obtemos

$$\frac{\beta[tf'(t) - f(t)]}{t[1 - \beta(f'(t)+1)]} < 1 \quad \forall\, t \in (0,\delta). \tag{1.21}$$

Pela definição de $\rho$, temos que $\rho \geq \delta > 0$. Logo $\rho > 0$.

Finalmente, provaremos que $0 < \frac{g_f(t)}{t} < 1 \quad \forall\, t \in (0,\rho)$. De fato, de (1.20), (1.21) e da Proposição 1.5.1 (desigualdades (1.12) e (1.13)), vale que

$$1 > \frac{g_f(t)}{t} = \frac{\beta[tf'(t) - f(t)]}{t[1 - \beta(f'(t)+1)]} > 0 \quad \forall\, \delta \in (0,\upsilon)\ e\ \forall\, t \in (0,\delta).$$

Pela definição da $\rho$ e a última desigualdade:

$$0 < \frac{g_f(t)}{t} < 1 \quad \forall t \in (0,\rho). \tag{1.22}$$

□

## 1.6 Relação entre $F$ e $f'$

Nesta seção apresentaremos a relação entre a função majorante e o operador não linear, obtido em [32]. Estes resultados serão necessários para os Capítulo 2 e 3. Iniciaremos com uma definição do erro linear de $F \in C^1$ em cada ponto de $\Omega$.

Definimos

$$E_F(x,y) := F(y) - [F(x) + F'(x)(y-x)], \quad y, x \in \Omega. \tag{1.23}$$

Utilizaremos uma limitação de $E_F$ pelo erro da linearização da função majorante $f$:

$$e_f(t,u) := f(u) - [f(t) + f'(t)(u-t)], \quad t, u \in [0, R). \tag{1.24}$$

Sejam $X$ e $Y$ espaços de Hilbert. As bolas aberta e fechada de centro $a \in X$ e raio $\delta > 0$ são denotados, respectivamente, por

$$B(a,\delta) := \{x \in X; \|x-a\| < \delta\}, \quad B[a,\delta] := \{x \in X; \|x-a\| \leq \delta\}$$

O seguinte lema nos dá uma relação entre o erro linear de $F$ e o erro da linearização de f. Destacamos que a prova do lema, não necessita da hipótese de convexidade da derivada da função majorante.

**Lema 1.6.1.** *Se $\|x - x_*\| < r$ então $\|E_F(x, x_*)\| \leq e_f(\|x - x_*\|, 0)$.*

*Demonstração.* Como $B(x_*, \kappa)$ é um conjunto convexo, segue $x_* + \tau(x - x_*) \in B(x_*, \kappa)$, $\tau \in [0,1]$. Logo, usando a definição de $E_F$, e o fato que $F \in C^1$ em $\Omega$, obtemos

$$\|E_F(x,x_*)\| = \|F(x_*) - [F(x) + F'(x)(x_* - x)]\| = \|F'(x)(x - x_*) + F(x_*) - F(x)\|. \tag{1.25}$$

Pela Proposição 1.2.2 temos que

$$\int_0^1 [F(x_* + \tau(x-x_*))]' d\tau = \int_0^1 (x - x_*) F(x_* + \tau(x - x_*)) d\tau = F(x) - F(x_*). \tag{1.26}$$

De (1.25) e (1.26):

$$
\begin{aligned}
\|E_F(x,x_*)\| &= \|F'(x)(x-x_*) - \int_0^1 F(x_* + \tau(x-x_*))d\tau(x-x_*)\| \\
&\leq \int_0^1 \|F'(x) - F'(x_* + \tau(x-x_*))\|\|x-x_*\|d\tau \\
&\leq \int_0^1 [f'(\|x-x_*\|) - f'(\tau\|x-x_*\|)]\|x-x_*\|d\tau \\
&= f'(\|x-x_*\|)\|x-x_*\| - \int_0^1 f'(\tau\|x-x_*\|)\|x-x_*\|d\tau \\
&= f'(\|x-x_*\|)\|x-x_*\| - \int_0^1 [f(\tau\|x-x_*\|)]'d\tau \\
&= f'(\|x-x_*\|)\|x-x_*\| - f(\|x-x_*\|) + f(0) \\
&= f(0) - [f(\|x-x_*\|) + f'(\|x-x_*\|)(\|x-x_*\| - 0)] \\
&= e_f(\|x-x_*\|, 0).
\end{aligned}
$$

□

O lema a seguir garante a não-singularidade de $F'(x)^*F'(x)$ em $B(x_*, r)$ e, assim, uma boa definição da aplicação iteração de Gauss-Newton, que será usado nos capítulos seguintes. Novamente é importante ressaltar que na prova do lema a seguir não faremos uso da hipótese de convexidade da derivada da função majorante.

**Lema 1.6.2.** *Se* $\|x-x_*\| < \min\{\upsilon, \kappa\}$, *então* $F'(x)^*F'(x)$ *é invertível e*

$$\|F'(x)^T\| \leq \frac{\beta}{1-\beta[f'(\|x-x_*\|)+1]}, \quad \|F'(x)^\dagger - F'(x_*)^\dagger\| < \frac{\sqrt{2}\beta^2[f'(\|x-x_*\|)+1]}{1-\beta[f'(\|x-x_*\|)+1]}.$$

*Em particular, se* $\|x-x_*\| < r$, *onde* $r = \min\{\kappa, \rho\}$, *então* $F'(x)^*F'(x)$ *é invertível em* $B(x_*, r)$.

*Demonstração.* Primeiramente, para simplificar a demonstração definiremos as matrizes

$$A = F'(x_*), \quad B = F'(x), \quad E = F'(x) - F'(x_*).$$

Seja $x \in \Omega$ tal que $\|x-x_*\| < \min\{\upsilon, \kappa\}$. Daí, usando a definição de $\beta$, a desigualdade (1.11) com $\tau = 0$ e a desigualdade (1.12), obtemos:

$$
\begin{aligned}
\|F'(x) - F'(x_*)\|\|[f'(x_*)^*F'(x_*)]^{-1}F'(x_*)\| &< [f'(\|x-x_*\|) - f'(0)]\|[f'(x_*)^*F'(x_*)]^{-1}F'(x_*)\| \\
&= \beta[f'(\|x-x_*\|) - f'(0)] \\
&= \beta[f'(\|x-x_*\|) + 1] \\
&< 1. \qquad (1.27)
\end{aligned}
$$

Assim, de (1.27) e da definição da inversa generalizada:

$$\|F'(x) - F'(x_*)\| \|F'(x_*)^\dagger\| < 1 \quad \Rightarrow \quad \|E\|\|A^\dagger\| < 1,$$

e como $F'(x_*)$ é injetiva segue do Lema 1.4.1 que $F'(x)$ é injetiva. Consequentemente, $F'(x)^*F'(x)$ é invertível. Logo, pela definição de $r$, segue que $F'(x)^*F'(x)$ é invertível $\forall\ x \in B(x_*, r)$.

Agora, como $F'(x)$ e $F'(x_*)$ são injetivas, então pelo Lema 1.4.2 e usando a desigualdade (1.11) para $\tau = 0$, temos

$$\begin{aligned}
\|F'(x)^T\| &\leq \frac{\|F'(x_*)^\dagger\|}{1 - \|F'(x_*)^\dagger\|\|F'(x) - F(x_*)\|} \\
&= \frac{\beta}{1 - \beta\|F'(x) - F(x_*)\|} \\
&\leq \frac{\beta}{1 - \beta(f'(\|x - x_*\| - f'(0)))} \\
&= \frac{\beta}{1 - \beta[f'(\|x - x_*\|) + 1]},
\end{aligned}$$

e

$$\begin{aligned}
\|F'(x)^\dagger - F'(x_*)^\dagger\| &\leq \frac{\sqrt{2}\|F'(x_*)^\dagger\|^2\|F'(x) - F(x_*)\|}{1 - \|F'(x_*)^\dagger\|\|F'(x) - F(x_*)\|} \\
&= \frac{\sqrt{2}\beta^2\|F'(x) - F(x_*)\|}{1 - \beta\|F'(x) - F(x_*)\|} \\
&\leq \frac{\sqrt{2}\beta^2(f'(\|x - x_*\| - f'(0)))}{1 - \beta(f'(\|x - x_*\| - f'(0)))} \\
&= \frac{\sqrt{2}\beta^2[f'(\|x - x_*\|) + 1]}{1 - \beta[f'(\|x - x_*\|) + 1]}.
\end{aligned}$$

Concluindo assim a prova do lema.

□

Nos capítulos seguintes veremos a convergência local do método de Gauss-Newton para resolver problemas do tipo min $\|F(x)\|^2$. Com algumas exigências para a função $F$ e admitindo que o ponto $x_*$ é zero de $F$, mostraremos que tomando o ponto inicial $x_0$ numa vizinhança apropriada de $x_*$, a sequência gerada pelo método de Gauss-Newton está bem definida e converge para $x_*$, onde veremos que sobre certas condiçãos sobre $f'$ temos uma convergência quadrática(Capítulo 2) e superlinear (Capítulo 3). Determinaremos ainda o raio de convergência ótimo e unicidade do zero de $F$.

# Capítulo 2

# Convergência local sob uma condição majorante

Neste capítulo, apresentaremos uma análise de convergência local do método de Gauss-Newton para encontrar solução para sistemas de equações não lineares, sob uma condição majorante, onde utilizaremos aqui a convexidade da derivada da função majorante. Veremos que a taxa de convergência é quadrática, conforme o Corolário 8 em [32].

O objetivo principal deste capítulo é mostrar a validade do seguinte teorema:

**Teorema 2.0.1.** *Sejam $X$ e $Y$ espaços de Hilbert, $\Omega \subset X$ um conjunto aberto e $F : \Omega \to Y$ uma função continuamente diferenciável tal que $F'$ tem imagem fechada em $\Omega$. Tome $x_* \in \Omega,\ R > 0$ ,*

$$\beta := \|F'(x_*)^\dagger\|, \quad \kappa := sup\{t \in [0, R) : B(x_*, t) \in \Omega\}.$$

*Suponha que $F(x_*) = 0$, $F'(x_*)$ é injetiva e que existe uma função continuamente diferenciável $f : [0, R) \to \mathbb{R}$ tal que*

$$\|F'(x) - F'(x_* + \tau(x - x_*))\| \leq f'(\|x - x_*\|) - f'(\tau\|x - x_*\|),$$

*para todo $\tau \in [0, 1],\ x \in B(x_*, \kappa)$ e*

*(h1) $f(0) = 0$ e $f'(0) = -1$,*

*(h2) $f'$ é convexa e estritamente crescente.*

*Sejam as constantes positivas $\upsilon := \{t \in [0, R) : 1 - \beta[f'(t) + 1] > 0\}$*

$$\rho := sup\ \left\{\ \delta \in (0, \upsilon) : \frac{\beta[tf'(t) - f(t)]}{t[1 - \beta(f'(t) + 1)]} < 1,\ t \in (0, \delta)\ \right\}, \quad r := \min\ \{\kappa, \rho\}.$$

*Então, o método de Gauss-Newton para resolver o Problema (2.2), com ponto inicial* $x_0 \in B(x_*, r) - \{x_*\}$,

$$x_{k+1} = x_k - F'(x_k)^{\dagger} F(x_k), \quad k = 0, 1, ...,$$

*está bem definido, a sequência* $\{x_k\}$ *está contida em* $B(x_*, r)$*, converge para* $x_*$ *o qual é o único zero de* $F$ *em* $B(x_*, \sigma)$*, onde* $\sigma := \sup\{t \in (0, \kappa) : \beta[f(t)/t + 1] < 1\}$ *e vale*

$$\|x_{k+1} - x_*\| \leq \frac{\beta[f'(\|x_0 - x_*\|)\|x_0 - x_*\| - f(\|x_0 - x_*\|)]}{\|x_0 - x_*\|^2[1 - \beta(f'(\|x_0 - x_*\|) + 1)]}\|x_k - x_*\|^2. \tag{2.1}$$

*Além disso, se* $\frac{\beta(\rho f'(\rho) - f(\rho))}{\rho(1 - \beta(f'(\rho)+1))} = 1$ *e* $\rho < \kappa$*, então* $r = \rho$ *é o melhor raio de convergência possível.*

Iniciaremos com a prova da boa-definição e, em seguida, apresentaremos resultados que garantem a validade do Teorema 2.0.1.

## 2.1 Definição do algoritmo e boa definição

Sejam $X$ e $Y$ espaços de Hilbert real ou complexo, $\Omega \subseteq X$ um conjunto aberto e $F : \Omega \to Y$ uma função continuamente diferenciável. Consideraremos o seguinte problema

$$F(x) = 0. \tag{2.2}$$

O método de Gauss-Newton, usado para encontrar zero do problema (2.2), é descrito como:

Dado $x_0 \in \Omega$, defina

$$x_{k+1} = x_k - [F'(x_k) * F'(x_k)]^{-1} F'(x_k)^* F(x_k), \quad k = 0, 1, .... \tag{2.3}$$

Pelo Lema 1.6.2 temos $F'(x)^* F'(x)$ invertível em $B(x_*, r)$, o que garante a boa definição da seguinte aplicação.

Seja $G_F : B(x_*, r) \to Y$, definida por:

$$G_F(x) = x - F'(x)^{\dagger} F(x). \tag{2.4}$$

Observe que $G_F(x)$, a iteração de Gauss-Newton para qualquer $x \in B(x_*, r)$, pode não pertencer a $B(x_*, r)$ ou até mesmo não pertencer ao domínio de $F$. Nesta situação garantiríamos a boa-definição de apenas uma iteração.

Para o nosso propósito de apresentar a boa-definição do método, precisamos assegurar que a iteração de Gauss-Newton possa ser repetida indefinidamente. Os lemas a seguir serão úteis para atingir este objetivo.

O próximo lema nos dá a bola ótima de convergência e o lema seguinte garante a unicidade do zero de F . Para prova de tais lemas utilizaremos a convexidade de $f'$.

**Lema 2.1.1.** *Se* $\frac{\beta(\rho f'(\rho)-f(\rho))}{\rho(1-\beta(f'(\rho)+1))} = 1$ *e* $\rho < k$, *então* $r = \rho$ *é o melhor raio de convergência possível.*

*Demonstração.* Considerando a função $h : (-k, k) \to \mathbb{R}$, definida por

$$h(t) = \begin{cases} -\frac{t}{\beta} + t - f(-t) & \text{se} \quad t \in (-k, 0] \\ -\frac{t}{\beta} + t + f(t) & \text{se} \quad t \in [0, k). \end{cases} \tag{2.5}$$

temos $h(0) = 0$, $h'(t) = -\frac{1}{\beta} + 1 + f'(|t|)$ e $h'(0) = -\frac{1}{\beta}$. Logo,

$$|h'(t) - h'(rt)| \leq f'(|t|) - f'(\tau|t|), \quad \tau \in [0, 1], t \in (-k, k).$$

Assim $F = h$ satisfaz todas as hipóteses do Teorema 2.0.1. Como $\rho < k$, basta mostrar que o método de Gauss-Newton para resolver o problema (2.2), com $F = h$ e ponto inicial $x_0 = \rho$ não converge.

Agora lembrando que $\rho < k$, consideraremos a última equação definida em (2.5), temos:

$$\begin{aligned} x_1 &= \rho - h'(\rho)^{\dagger} h(\rho) \\ &= \rho - \frac{h(\rho)}{h'(\rho)} \\ &= \rho - \frac{-\frac{\rho}{\beta} + \rho + f(\rho)}{-\frac{1}{\beta} + 1 + f'(\rho)} \\ &= \rho \left( 1 - \frac{-\frac{\rho}{\beta} + \rho + f(\rho)}{-\frac{\rho}{\beta} + \rho + \rho f'(\rho)} \right) \\ &= \rho \left( \frac{-\frac{\rho}{\beta} + \rho + \rho f'(\rho) + \frac{\rho}{\beta} - \rho - f(\rho)}{-\frac{\rho}{\beta} + \rho + \rho f'(\rho)} \right) \\ &= \rho \left( \frac{\rho f'(\rho) - f(\rho)}{-\frac{\rho}{\beta} + \rho + \rho f'(\rho)} \right) \\ &= \rho \left( -\frac{\beta \left[ \rho f'(\rho) - f(\rho) \right]}{\rho \left[ 1 - \beta(1 + \rho f'(\rho)) \right]} \right) \\ &= -\rho, \end{aligned}$$

Considerando a primeira equação definida em (2.5), temos

$$\begin{aligned}
x_2 &= -\rho - h'(-\rho)^{\dagger}h(-\rho) \\
&= -\rho - \frac{h(-\rho)}{h'(-\rho)} \\
&= -\rho - \frac{\frac{\rho}{\beta} - \rho - f(\rho)}{-\frac{1}{\beta} + 1 + f'(\rho)} \\
&= \rho\left(-1 - \frac{\frac{\rho}{\beta} - \rho - f(\rho)}{-\frac{\rho}{\beta} + \rho + \rho f'(\rho)}\right) \\
&= \rho\left(\frac{\frac{\rho}{\beta} - \rho - \rho f'(\rho) - \frac{\rho}{\beta} + \rho + f(\rho)}{-\frac{\rho}{\beta} + \rho + \rho f'(\rho)}\right) \\
&= \rho\left(\frac{\beta\left[\rho f'(\rho) - f(\rho)\right]}{\rho\left[1 - \beta(1 + \rho f'(\rho))\right]}\right) \\
&= \rho.
\end{aligned}$$

Assim, o método de Gauss-Newton para resolver o problema (2.2) com $F = h$ e ponto inicial $x_0 = \rho$, nos da o ciclo

$$x_0 = \rho, \quad x_1 = -\rho, \quad x_2 = \rho, \; ....$$

Portanto, a sequência não converge, concluindo assim a prova do lema. □

**Lema 2.1.2.** *O ponto $x_*$ é o único zero de $F$ em $B(x_*, \sigma)$.*

*Demonstração.* Seja $y \in B(x_*, \sigma)$ outro zero de $F$, ou seja, $F(x_*) = 0$ e $F(y) = 0$, temos

$$y - x_* = F'(x_*)^{\dagger}\left[F'(x_*)(y - x_*) - F(y) + F(x_*)\right].$$

Combinando a última equação com as propriedades de norma, definição de $\beta$, e Proposição 1.2.2, obtemos

$$\begin{aligned}
\|y - x_*\| &\leq \|F'(x_*)^{\dagger}\| \, \|F'(x_*)(y - x_*) - F(y) + F(x_*)\| = \\
&= \beta\left\|F'(x_*)(y - x_*) - \int_0^1 \left[F'(x_* + u(y - x_*))(y - x_*)\right] du\right\| \\
&= \beta\left\|\int_0^1 \left[F'(x_*) - F'(x_* + u(y - x_*))\right](y - x_*)du\right\| \\
&\leq \beta\int_0^1 \|F'(x_*) - F'(x_* + u(y - x_*))\| \, \|y - x_*\| du. \qquad (2.6)
\end{aligned}$$

Usando (1.11) com $x = x^* + u(y - x_*)$ e $\tau = 0$, temos:

$$\begin{aligned} \|F'(x) - F'(x_* + \tau(x - x_*))\| &= \|F'(x_* + u(y - x_*)) - F'(x_*)\| \\ &\leq f'(u\|y - x_*\|) - f'(0). \end{aligned} \tag{2.7}$$

Das desigualdades (2.6) e (2.7), segue que:

$$\begin{aligned} \|y - x_*\| &\leq \beta \int_0^1 [f'(u\|y - x_*\|) - f'(0)] \|y - x_*\| du \\ &\leq \beta \left( \frac{f(\|y - x_*\|)}{\|y - x_*\|} - f'(0) \right) \|y - x_*\| \\ &= \beta \left( \frac{f(\|y - x_*\|)}{\|y - x_*\|} + 1 \right) \|y - x_*\|. \end{aligned} \tag{2.8}$$

Sabendo que $0 < \|y - x_*\| < \sigma$, da Proposição 1.5.5 (com $t = \|y - x_*\|$) e (2.8) concluímos que

$$\|y - x_*\| < \|y - x_*\|,$$

o que é um absurdo. Logo $y = x_*$. $\square$

**Lema 2.1.3.** *Seja $x \in \Omega$. Se $\|x - x_*\| < r$, então*

$$\|G_F(x) - x_*\| \leq \frac{\beta[f'(\|x - x_*\|)\|x - x_*\| - f(\|x - x_*\|)]}{\|x - x_*\|^2[1 - \beta(f'(\|x - x_*\|) + 1)]} \|x - x_*\|^2. \tag{2.9}$$

*Em particular,*

$$\|G_F(x) - x_*\| < \|x - x_*\|.$$

*Demonstração.* Como $\|x - x_*\| < r$, o Lema 1.6.2 implica que $F'(x)^*F'(x)$ é invertível. Daí, pela Proposição 1.4.4, $G_F(x)$ está bem definida. Assim, usando a definição de $G_F$ dada em (2.4) e $F(x_*) = 0$, obtemos

$$\begin{aligned} G_F(x) - x_* &= x - x_* - [F'(x)^*F'(x)]^{-1}F'(x)^*F(x) \\ &= [F'(x)^*F'(x)]^{-1}F'(x)^*[F'(x)(x - x_*) - F(x) + F(x_*)]. \end{aligned}$$

Logo, da equação acima, usando a definição da inversa de Moore-Penrose (veja Definição 1.4.2) e a definição do erro linear de $F$ (veja (1.23)), segue que

$$\|G_F(x) - x_*\| \leq \|F'(x)^\dagger\| \|E_F(x, x_*)\|.$$

Então, dos Lemas 1.6.1 e 1.6.2, temos

$$\|G_F(x) - x_*\| \leq \frac{\beta e_f(\|x - x_*\|, 0)}{1 - \beta(f'(\|x - x_*\|) + 1)}. \tag{2.10}$$

Daí, usando a definição do erro da linearização de $f$ dada em (1.24) e ($h1$), concluímos que

$$\|G_F(x) - x_*\| \leq \frac{\beta[f'(\|x - x_*\|)\|x - x_*\| - f(\|x - x_*\|)]}{1 - \beta(f'(\|x - x_*\|) + 1)}. \tag{2.11}$$

Logo,

$$\|G_F(x) - x_*\| \leq \frac{\beta[f'(\|x - x_*\|)\|x - x_*\| - f(\|x - x_*\|)]}{\|x - x_*\|^2[1 - \beta(f'(\|x - x_*\|) + 1)]}\|x - x_*\|^2,$$

o que conclui a primeira parte do lema.

De (2.11) temos:

$$\begin{aligned} \|G_F(x) - x_*\| &\leq \frac{\beta[f'(\|x - x_*\|)\|x - x_*\| - f(\|x - x_*\|)]}{1 - \beta(f'(\|x - x_*\|) + 1)} \\ &= \frac{\beta[f'(\|x - x_*\|)\|x - x_*\| - f(\|x - x_*\|)]}{[1 - \beta(f'(\|x - x_*\|) + 1)]\|x - x_*\|}\|x - x_*\| \\ &\leq \|x - x_*\|, \end{aligned}$$

onde a última desigualdade ocorre por (1.15) com $t = \|x - x_*\|$ (pois $x \in B(x_*, r)/\{x_*\}$, isto é, $0 < \|x - x_*\| < r \leq \rho$).

Portanto,

$$\|G_F(x) - x_*\| \leq \|x - x_*\|.$$

□

## 2.2 Análise de convergência

Esta seção é dedicada à prova do Teorema 2.0.1, conforme [32]. Destacamos que esse é um importante resultado por ser uma generalização dos resultados do Método de Newton para encontrar uma solução para o problema (2.2) (veja [29]), isto é, resolver o sistema de equações não-lineares $F(x) = 0$.

Primeiro, observe que a equação em (2.3) junto com (2.4) implica que a sequência $\{x_k\}$ satisfaz

$$x_{k+1} = G_F(x_k), \quad k = 0, 1, .... \tag{2.12}$$

*Demonstração.* Como $x_0 \in B(x_*, r)/\{x_*\}$, isto é, $0 < \|x_0 - x_*\| < r$, provaremos por um argumento indutivo que $\{x_k\}$ está bem definida e contida em $B(x_*, r)$. Como $\|x_0 - x_*\| < r$, supondo que $\|x_k - x_*\| < r$, devemos mostrar que $\|x_{k+1} - x_*\| < r$. De fato, pois o Lema 1.6.2 garante a boa definição da $G_F$, daí, por (2.12) e a última desigualdade do Lema 2.1.3, temos

$$\|x_{k+1} - x_*\| = \|G(x_k) - x_*\| < \|x_k - x_*\| < r,$$

concluindo assim a prova por indução.

Agora, provaremos que $\{x_k\}$ converge para $x_*$. Como $\{x_k\}$ está bem definida e contida em $B(x_*, r)$, combinando a desigualdade (2.9) do Lema 2.1.3 com (2.12), temos

$$\|x_{k+1} - x_*\| \leq \frac{\beta[f'(\|x_k - x_*\|)\|x_k - x_*\| - f(\|x_k - x_*\|)]}{\|x_k - x_*\|^2[1 - \beta(f'(\|x_k - x_*\|) + 1)]}\|x_k - x_*\|^2 \quad \forall k = 0, 1, .(2.13)$$

Por outro lado, usando (2.12) e a última parte do Lema 2.1.3

$$\|x_k - x_*\| = \|G(x_{k-1}) - x_*\| < \|x_{k-1} - x_*\| < ... < \|x_0 - x_*\|, \quad k = 1, 2, .... \quad (2.14)$$

Daí, combinando (2.13), (2.14) e o segundo item da Proposição 1.5.3, concluímos que

$$\|x_{k+1} - x_*\| \leq \frac{\beta[f'(\|x_0 - x_*\|)\|x_0 - x_*\| - f(\|x_0 - x_*\|)]}{\|x_0 - x_*\|^2[1 - \beta(f'(\|x_0 - x_*\|) + 1)]}\|x_k - x_*\|^2 \quad \forall k = 0, 1, .(2.15)$$

Logo, a desigualdade (2.1) do teorema está provada.

Considerando (2.14) e a desigualdade (2.15) temos que, para $k = 0, 1, ...$

$$\|x_{k+1} - x_*\| \leq \left[\frac{\beta[f'(\|x_0 - x_*\|)\|x_0 - x_*\| - f(\|x_0 - x_*\|)]}{\|x_0 - x_*\|[1 - \beta(f'(\|x_0 - x_*\|) + 1)]}\right]\|x_k - x_*\|. \quad (2.16)$$

Da desigualdade (1.15) com $t = \|x_0 - x_*\|$ e de (2.16), obtemos:

$$\|x_{k+1} - x_*\| < \|x_k - x_*\|.$$

Então, $\{\|x_k - x_*\|\}$ converge para zero, isto é, $\{x_k\}$ converge para $x_*$. O raio de convergência ótimo e unicidade do zero de $F$ foram provados nos Lemas 2.1.1 e 2.1.2, respectivamente. $\square$

A inequação (2.15) implica que o método de Gauss-newton converge quuadraticamente para $x_*$. Então seu comportamento é semelhante ao método de Newton ([28], [36]).

# Capítulo 3

# Convergência local do método de Gauss-Newton sob uma condição majorante relaxada

Neste capítulo apresentaremos um método de Gauss-Newton para resolver o problema $F(x) = 0$, estudado em [27] que é uma generalização do Corolário 8 de [32], onde foi visto no Capítulo 2, bem como do Teorema 2 em [29] . Veremos uma análise de convergência local do método de Gauss-Newton para resolução do Problema (2.2) sob uma condição majorante relaxada, onde a hipótese $(h2)$ da definição de função majorante é enfraquecida apenas para $f'$ estritamente crescente. Veremos que, sob esta condição, o método está bem definido e converge com uma taxa superlinear, conforme [27]. Além disso, também apresentaremos o raio ideal de convergência.

## 3.1 Considerações Iniciais

O objetivo desta seção é apresentar o seguinte teorema:

**Teorema 3.1.1.** *Seja $X$ e $Y$ espaços de Hilbert, $\Omega \subset X$ um conjunto aberto e $F : \Omega \to Y$ uma função continuamente diferenciável tal que $F'$ tem imagem fechada em $\Omega$. Seja $x_* \in \Omega$, $R > 0$ e $B := \|F'(x_*)^\dagger\|$, $\kappa := sup\{t \in [0, R) : B(x_*, t) \in \Omega\}$ Suponha que $F(x_*) = 0$, $F'(x_*)$ é injetiva e existe uma $f : [0, R) \to \mathbb{R}$ continuamente diferenciável tal que*

$$\|F'(x) - F'(x_* + \tau(x - x_*))\| \leq f'(\|x - x_*\|) - f'(\tau\|x - x_*\|),$$

*para todo $\tau \in [0, 1]$, $x \in B(x_*, \kappa)$ e* *(h1)* $f(0) = 0$ *e* $f'(0) = -1$,

*(h2)* $f'$ *é estritamente crescente.*

*Sejam*

$$\upsilon := \{t \in [0, R) : 1 - \beta[f'(t) + 1] > 0\},$$
$$\rho := sup\ \{\ \delta \in (0, \upsilon) : \frac{\beta[tf'(t) - f(t)]}{t[1 - \beta(f'(t) + 1)]} < 1,\ t \in (0, \delta)\ \}, e$$
$$r := \min\ \{\kappa, \rho\}.$$

*Então as sequências com ponto inicial* $x_0 \in B(x_*, r) - \{x_*\}$ *e* $t_0 = \|x_0 - x_*\|$, *respectivamente, a saber:*

$$x_{k+1} = x_k - F'(x_k)^\dagger F(x_k), \quad t_{k+1} = \frac{\beta[t_k f'(t_k) - f(t_k)]}{1 - \beta(f'(t_k) + 1)}, \quad k = 0, 1, ... \tag{3.1}$$

*estão bem definidas,* $\{t_k\}$ *é estritamente decrescente, está contida em* $(0, r)$ *e, converge para* $0$. *Além disso,* $\{x_k\} \subset B(x_*, r)$ *e converge para o ponto* $x_*$, *que é único zero de* $F$ *em* $B(x_*, \sigma)$, *onde* $\sigma := \sup\{t \in (0, \kappa) : \beta[f(t)/t + 1] < 1\}$ *e*

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0, \quad \lim_{k\to\infty} \frac{t_{k+1}}{t_k} = 0.$$

*Além disso, se* $\frac{\beta(\rho f'(\rho) - f(\rho))}{\rho(1 - \beta(f'(\rho) + 1))} = 1$ *e* $\rho < \kappa$, *então* $r = \rho$ *é o melhor raio de convergência.*
*Se* $0 \leq p \leq 1$, *e ocorre* $(h3)$, *onde*

$(h3)$ *A função* $(0, \upsilon) \ni t \longmapsto [tf'(t) - f(t)]\ /t^{p+1}$ *é estritamente crescente, então a sequência* $\left\{\frac{t_{k+1}}{t_k^{p+1}}\right\}$ *é estritamente decrescente e temos:*

$$\|x_{k+1} - x_*\| \leq \left[\frac{t_{k+1}}{t_k^{p+1}}\right] \|x_k - x_*\|^{p+1} \leq \left[\frac{t_1}{t_0}^{p+1}\right] \|x_k - x_*\|^{p+1}, \quad k = 0, 1, ....$$

*Consequentemente, para* $k \geq 0$

$$\|x_k - x_*\| \leq \begin{cases} t_0 \left[\frac{t_1}{t_0}\right]^k, & se \quad p = 0 \\ t_0 \left[\frac{t_1}{t_0}\right]^{\frac{(p+1)^k - 1}{p}}, & se \quad p \neq 0 \end{cases}.$$

Ferreira [28], apresenta uma analise de convergência local do método de Newton para resolver sistemas de equações não-lineares sob uma condição majorante com taxa quadrática, onde a hipóetese da convexidade da derivada da função majorante é assumida. Por outro lado observou-se em [29] que a hipótese de convexidade da derivada da função majorante é necessária apenas para a obtermos uma taxa de convergência quadrática, visto que sem a convexidade da $f'$, a sequencia gerada pelo método de newton continua convergindo, porém agora com taxa de convergengia superlinear como foi observado no Teorema 2 em [29]. Agora, em relação ao método de Gauss-Newton, as mesmas caracteristicas que são observadas, isto é, a convexidade da $f'$ é necessária apenas para obtermos uma taxa de convergência quadrática, confome vimos no Capítulo 2, se $f'$ não for convexa temos que a taxa de convergência será superlinear, observado por Gonçalves em [27].

Novamente consideraremos os seguinte fatos vistos no Capítulo 2:
Sejam $X$ e $Y$ espaços de Hilbert real ou complexo, $\Omega \subseteq X$ um conjunto aberto e $F : \Omega \to Y$ uma função continuamente diferenciável. Consideraremos o seguinte problema

$$F(x) = 0. \tag{3.2}$$

O método de Gauss-Newton usado para encontrar zeros do problema acima é descrito como:

Dado $x_0 \in \Omega$, defina

$$x_{k+1} = x_k - [F'(x_k) * F'(x_k)]^{-1} F'(x_k)^* F(x_k), \quad k = 0, 1, .... \tag{3.3}$$

Consideraremos aqui a iteração de Gauss-Newton para $F$ em $B(x_*, r)$, tal como definida no Capítulo 2, a seguinte aplicação.

Seja $G_F : B(x_*, r) \to Y$, definida por:

$$G_F(x) = x - F'(x)^\dagger F(x). \tag{3.4}$$

Agora, observe que a equação em (3.3) junto com (3.4) implica que a sequência $\{x_k\}$ satisfaz

$$x_{k+1} = G_F(x_k), \quad k = 0, 1, .... \tag{3.5}$$

Lembremos ainda que $g_f$ dada em (1.19) por:

$$g_f(t) = \frac{\beta[tf'(t) - f(t)]}{1 - \beta(f'(t) + 1)}. \tag{3.6}$$

Visto que para prova do Lema 1.6.2 não utilizamos a hipótese da convexidade de $f'$, então o mesmo resultado obtido no Capítulo 2 temos também aqui, ou seja, tal lema garante a boa definição de uma única iteração Gauss-Newton.

Para assegurarmos que a iteração de Gauss-Newton pode ser repetida infinitamente necessitamos de alguns resultados, dentre eles Bola ótima de convergência e unicidade da solução de $F$ em $B(x_*, \sigma)$, que serão apresentados na próxima seção.

## 3.2 Bola ótima de convergência e unicidade

A seguir apresentaremos os lemas que nos dão a bola ótima de convergência e a unicidade da solução de $F$ em $B(x_*, \sigma)$, As demonstrações serão ocultadas pois são similares a aquelas encontradas no Capítulo 2, visto que a prova desses resultados não depende da convexidade da derivada da função majorante e nem de $(h3)$.

**Lema 3.2.1.** *Se* $\frac{\beta(\rho f'(\rho)-f(\rho))}{\rho(1-\beta(f'(\rho)+1))} = 1$ *e* $\rho < k$, *então* $r = \rho$ *é o melhor raio de convergência possível.*

*Demonstração.* Similar à prova do Lema 2.1.1. □

**Lema 3.2.2.** *O ponto* $x_*$ *é o zero de* $F$ *em* $B(x_*, \sigma)$.

*Demonstração.* Similar à prova do Lema 2.1.2. □

## 3.3 Resultados auxiliares

Veremos a seguir mais resultados que ajudam a assegurar que a iteração de Gauss-Newton possa se repetir infinitamente.

**Lema 3.3.1.** *Se* $\|x - x_*\| < r$, *então* $\|G_F(x) - x_*\| < g_f(\|x - x_*\|)$ *e, consequentemente,* $G_F(B(x_*, r)) \subset B(x_*, r)$.

*Demonstração.* Primeiro, vejamos que para $x = x_*$, como $F'(x_*)^\dagger F(x_*) = 0$, temos que
$\|G_F(x_*) - x_*\| < g_f(\|x_* - x_*\|)$.
Agora, assuma que $0 < \|x - x_*\| < r$. Do Lema 1.6.2 temos $F'(x)^* F(x)$ é invertível. Usando (3.4) e que $F(x_*) = 0$, obtemos:

$$\begin{aligned} G_F(x) - x_* &= x - x_* - [F'(x)^* F'(x)]^{-1} F'(x)^* F(x) \\ &= [F'(x)^* F'(x)]^{-1} F'(x)^* [F'(x)(x - x_*) - F(x) + F(x_*)]. \end{aligned}$$

Usando a equação acima, a definição da inversa de Moore-Penrose, a definição de $E_F$ em (1.23), os Lemas 1.6.1 e 1.6.2 e, a definição da $g_f$ em (3.6), temos:

$$\begin{aligned}
\|G_F(x) - x_*\| &\leq \|F'(x)^\dagger\|\|E_F(x, x_*)\| \\
&\leq \frac{\beta e_f(\|x - x_*\|, 0)}{1 - \beta(f'(t) + 1)} \\
&= \frac{\beta(f(0) - [f(\|x - x_*\|) - f'(\|x - x_*\|)(\|x - x_*\|)]}{1 - \beta(f'(\|x - x_*\|) + 1)} \\
&= \frac{\beta[f'(\|x - x_*\|)(\|x - x_*\|) - f(\|x - x_*\|)]}{1 - \beta(f'(\|x - x_*\|) + 1)} \\
&= g_f(\|x - x_*\|), \qquad (3.7)
\end{aligned}$$

o que prova a primeira parte do lema.

Para concluir a prova, tomando $x \in B(x_*, r)$, onde $0 < \|x - x_*\| < r \leq \rho$, por (3.7) e (1.22) obtemos

$$\|G_F(x) - x_*\| \leq g_f(\|x - x_*\|) < \|x - x_*\|.$$

Logo,

$$G_F(B(x_*, r)) \subset B(x_*, r),$$

o que conclui a prova do lema. □

**Lema 3.3.2.** *Suponha* $(h3)$*. Se* $\|x - x_*\| \leq t < r$ *então*

$$\|G_F(x) - x_*\| \leq \left[\frac{g_f(t)}{t^{p+1}}\right] \|x - x_*\|^{p+1}.$$

*Demonstração.* Se $x = x_*$, pela definição da $G_F$, temos

$$\|G_F(x_*) - x_*\| = \|F'(x_*)^+ F(x_*)\| = 0 = \frac{g_f(t)}{t^{p+1}} \|x_* - x_*\|^{p+1}.$$

Agora, como $f'$ é estritamente crescente em $[0, R)$, então a função $[0, \upsilon) \ni t \to \frac{1}{[1-\beta(f'(t)+1)]}$, também possui essa mesma propriedade. Consideraremos $(h3)$ e mostraremos que $\left\{\frac{t_{k+1}}{t_k^{p+1}}\right\}$ é estritamente decrescente.

Por $f'$ ser estritamente crescente em $[0, R)$, obtemos que a função:

$$[0, \upsilon) \ni\mapsto 1/[1 - \beta(f'(t) + 1)] \qquad (3.8)$$

também é.

De $(h3)$ temos que a função:

$$(0,\upsilon) \ni t \longmapsto [tf'(t) - f(t)] / t^{p+1} \tag{3.9}$$

é estritamente crescente. Daí, multiplicando (3.8) e (3.9) obtemos que

$$\frac{[tf'(t) - f(t)]}{t^{p+1}[1 - \beta(f'(t) + 1)]}, \quad t \in (0,\upsilon) \tag{3.10}$$

é estritamente crescente em $(0,\upsilon)$.

Agora, usando a definição da $g_f$ e (3.10), temos que

$$\frac{g_f(t)}{t^{p+1}} = \frac{[tf'(t) - f(t)]}{t^{p+1}[1 - \beta(f'(t) + 1)]}, \quad t \in (0,\upsilon),$$

a qual é estritamente crescente.

Daí, usando $0 < \|x - x_*\| \leq t$, obtemos

$$\frac{g_f(\|x - x_*\|)}{\|x - x_*\|^{p+1}} \leq \frac{g_f(t)}{t^{p+1}} \Rightarrow g_f(\|x - x_*\|) \leq \frac{g_f(t)}{t^{p+1}} \|x - x_*\|^{p+1}.$$

A última desigualdade e o Lema 3.3.1 nos diz que

$$\|G_f(x) - x_*\| \leq g_f(\|x_- x_*\|) \leq \frac{g_f(t)}{t^{p+1}} \|x - x_*\|^{p+1}.$$

Portanto,

$$\|G_f(x) - x_*\| \leq \frac{g_f(t)}{t^{p+1}} \|x - x_*\|^{p+1},$$

provando o lema. □

**Corolário 3.3.1.** *A sequência $\{t_k\}$ está bem definida, é estritamente decrescente e está contida em $(0,\rho)$. Além disso, $\{t_k\}$ converge superlinearmente para 0, isto é, $\lim_{k\to\infty} \frac{t_{k+1}}{t_k} = 0$. Se adicionarmos $(h3)$, a sequência $\left\{\frac{t_{k+1}}{t_k^{p+1}}\right\}$ é estritamente decrescente.*

*Demonstração.* Afim de mostrar a boa definição da sequência $\{t_k\}$ dada em (3.1), vejamos que $t_{k+1} = \dfrac{\beta[t_k f'(t_k) - f(t_k)]}{1 - \beta(f'(t_k) + 1)}$ e usando a definição da $g_f$ dada (3.6), temos facilmente que a sequência $\{t_k\}$ é equivalentemente definida como:

$$t_0 = \|x_0 - x_*\|, \quad t_{k+1} = g_f(t_k), \quad k = 0, 1, .... \tag{3.11}$$

Temos que a sequência $\{t_k\}$ está bem definida. De fato, mostraremos que $t_k \in (0, \rho)$. Para isto, usaremos um argumento de indução. Observe que $0 < t_0 = \|x_k - x_*\| < r \leq \rho$, assumindo que $0 < t_k < \rho$ para algum $k > 0$, mostraremos que $0 < t_{k+1} < \rho$. De fato, usando (1.22) e (3.11), obtemos

$$0 < \frac{g_f(t_k)}{t_k} < 1, \quad t_k \in (0, \rho) \quad \Longrightarrow 0 < \frac{g_f(t_k)}{t_k} = \frac{t_{k+1}}{t_k} < 1.$$

Daí,

$$0 < t_{k+1} < t_k < \rho, \tag{3.12}$$

isto é, $0 < t_{k+1} < \rho$. Portanto, $\{t_k\}$ está contida em $(0, \rho)$. De (3.12) ganhamos ainda que $\{t_k\}$ é estritamente decrescente.

Logo, $\{t_k\}$ é limitada e monótona, e consequentemente convergente. Seja $t_*$ tal que $\lim_{k\to\infty} t_k = t_*$, temos $0 \leq t_* < \rho$. Usando a continuidade da $g_f$ em $[0, \rho)$ e (3.11) segue que

$$t_* = \lim_{k\to\infty} t_k = \lim_{k\to\infty} g_f(t_{k-1}) = g_f\left(\lim_{k\to\infty} t_{k-1}\right) = g_f(t_*).$$

Assim, $0 \leq t_* = g_f(t_*)$.

Se $t_* \neq 0$ então por (1.22) (veja Proposição 1.5.6) temos $g_f(t_*) < t_*$, o qual é uma contradição. Logo $\lim_{k\to\infty} t_k = t_* = 0$, ou seja, $\{t_k\}$ converge para 0.

Agora, usando a definição de $\{t_k\}$ e $\lim_{t\to 0} \frac{g_f(t)}{t} = 0$ (veja Proposição 1.5.6 ), temos

$$\lim_{k\to\infty} \frac{t_{k+1}}{t_k} = \lim_{k\to\infty} \frac{g_f(t_k)}{t_k} = 0.$$

Portanto, $\displaystyle\lim_{k\to\infty} \frac{t_{k+1}}{t_k} = 0$.

Agora, como $\{t_k\}$ é estritamente decrescente e $\dfrac{g_f(t)}{t^{p+1}}$ é estritamente crescente (visto no Lema 3.3.2), obtemos

$$t_k < t_{k-1} \quad \Rightarrow \quad \frac{g_f(t_k)}{t_k^{p+1}} < \frac{g_f(t_{k-1})}{t_{k-1}^{p+1}} \quad \Rightarrow \quad \frac{t_{k+1}}{t_k^{p+1}} < \frac{t_k}{t_{k-1}^{p+1}}.$$

Então, a sequência $\left\{\frac{t_{k+1}}{t_k^{p+1}}\right\}$ é estritamente decrescente. $\square$

**Corolário 3.3.2.** *A sequência $\{x_k\}$ está bem definida, contida em $B(x_*, r)$, converge para $x_*$ e*

$$\lim_{k\to+\infty} \frac{\|x_{k+1}-x_*\|}{\|x_k - x_*\|} = 0. \tag{3.13}$$

*Se, adicionalmente, (h3) vale, então as sequências $\{x_k\}$ e $\{t_k\}$ satisfazem*

$$\|x_{k+1}-x_*\| \leq \left[\frac{t_{k+1}}{t_k^{p+1}}\right] \|x_k - x_*\|^{p+1} \leq \left[\frac{t_1}{t_0^{p+1}}\right] \|x_k - x_*\|^{p+1}, \quad k = 0, 1, 2, .... \tag{3.14}$$

*Consequentemente, para $k > 0$*

$$\|x_{k+1}-x_*\| \leq \begin{cases} t_0 \left[\frac{t_1}{t_0}\right]^k, & se \quad p = 0 \\ t_0 \left[\frac{t_1}{t_0}\right]^{\frac{(p+1)^k - 1}{p}}, & se \quad p \neq 0. \end{cases}$$

*Demonstração.* Visto que $x_0 \in B(x_*, r)/\{x_*\}$, isto é, $0 < \|x_0 - x_*\| < r$, e $r \leq \upsilon$, provaremos por um argumento indutivo que $\{x_k\}$ está bem definida e contida em $B(x_*, r)$. Como $\|x_0 - x_*\| < r$, supondo que $\|x_k - x_*\| < r$, devemos mostrar que $\|x_{k+1} - x_*\| < r$. De fato, pois o Lema 1.6.2 garante a boa definição da $G_F$, daí, por (3.5) e do Lema 3.3.1, temos

$$\|x_{k+1}-x_*\| = \|G(x_k) - x_*\| < \|x_k - x_*\| < r,$$

concluindo a prova por indução.

Agora, como

$$\|x_k - x_*\| < r, \quad k = 0, 1, 2, ...,$$

provaremos que $\{x_k\}$ converge para $x_*$. Como $\{x_k\}$ está bem definida e contida em $B(x_*, r)$, obtemos de (3.5), Lema 3.3.1 e Proposição 1.5.6, segue que

$$0 \leq \|x_{k+1}-x_*\| = \|G_F(x_k) - x_*\| \leq g_f(\|x_k - x_*\|) < \|x_k - x_*\| < r \leq \rho. \tag{3.15}$$

De (3.15) obtemos $\{\|x_k - x_*\|\}$ limitada e estritamente decrescente e, portanto, convergente.

Seja $l_* = \lim_{k\to\infty} \|x_k - x_*\|$. Como $\{\|x_k - x_*\|\}$ esta contida em $(0, \rho)$ e é estritamente decrescente, temos $0 \leq l_* < \rho$.

Agora, da desigualdade (3.15), temos

$$0 \leq \|x_{k+1}-x_*\| \leq g_f(\|x_k - x_*\|) < \|x_k - x_*\|.$$

Daí, passando ao limite na desigualdade acima com $k \to \infty$ e usando a continuidade da $g_f$ em $[0, \rho)$, obtemos $0 \le l_* \le g_f(l_*) \le l_*$, o que por sua vez implica $0 \le l_* = g_f(l_*)$. Note que, se $l_* \ne 0$, temos pela Proposição 1.5.6 que $g_f(l_*) < l_*$, contradição. Logo, $l_* = 0$.

Então, $\lim_{k\to\infty} \|x_k - x_*\| = 0$ e, consequentemente, $x_k \to x_*$.

Para provar a igualdade (3.17), note que a desigualdade (3.15) implica

$$0 \le \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} \le \frac{g_f(\|x_k - x_*\|)}{\|x_k - x_*\|}, \quad k = 0, 1, ....$$

Como $\lim_{k\to\infty} \|x_k - x_*\| = 0$, usando a desigualdade acima e Proposição 1.5.6, temos que

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0.$$

Agora, considerando $(h3)$, mostraremos (3.14). Primeiro, através de um argumento indutivo mostraremos que as sequências $\{x_k\}$ e $\{t_k\}$ definidas, respectivamente em (3.5) e (3.11), satisfazem

$$\|x_k - x_*\| \le t_k, \quad k = 0, 1, .... \tag{3.16}$$

De fato, basta tomar $t_0 = \|x_k - x_*\|$. Assim a desigualdade (3.16) é valida para $k = 0$. Assumindo $\|x_k - x_*\| \le t_k$ para $k > 0$, com $\{t_k\} \subset (0, \rho)$, provaremos que

$$\|x_{k+1} - x_*\| \le t_{k+1}$$

.

Usando (3.5), Lema 3.3.2, a hipótese de indução e (3.11), obtemos

$$\|x_{k+1} - x_*\| = \|G_F(x_k) - x_*\| \le \frac{g_f(t_k)}{t_k^{p+1}} \|x_k - x_*\|^{p+1} \le g_f(t_k) = t_{k+1}.$$

Então $\|x_{k+1} - x_*\| \le t_{k+1}$, o que conclui a prova por indução.

A primeira desigualdade em (3.14) será obtida através da combinação de (3.5) e (3.16), e usaremos também o Lema 3.3.2 e a equação (3.11):

$$\|x_{k+1} - x_*\| = \|G_F(x_k) - x_*\| \le \frac{g_f(t_k)}{t_k^{p+1}} \|x_k - x_*\|^{p+1} = \frac{t_{k+1}}{t_k^{p+1}} \|x_k - x_*\|^{p+1}.$$

Para mostrar a segunda desigualdade em (3.14), observe que, do Corolário 3.3.1, $\left\{ \frac{t_{k+1}}{t_k^{p+1}} \right\}$ é estritamente decrescente. Logo:

$$\frac{t_{k+1}}{t_k^{p+1}} < \frac{t_1}{t_0^{p+1}}, \quad k = 0, 1, ....$$

Portanto,

$$\left[\frac{t_{k+1}}{t_k^{p+1}}\right] \|x_k - x_*\|^{p+1} \leq \left[\frac{t_1}{t_0^{p+1}}\right] \|x_k - x_*\|^{p+1}, \quad k = 0, 1, 2, ....$$

Por (3.14), obtemos:

$$\|x_{k+1} - x_*\| \leq \left[\frac{t_1}{t_0^{p+1}}\right] \|x_k - x_*\|^{p+1}, \quad k = 0, 1, 2, ....$$

**i)** Se $p = 0$, temos

$$\|x_{k+1} - x_*\| \leq \left(\frac{t_1}{t_0}\right) \|x_k - x_*\|.$$

Aplicando essa desigualdade $k$ vezes, segue que

$$\begin{aligned}
\|x_k - x_*\| &\leq \left(\frac{t_1}{t_0}\right) \|x_{k-1} - x_*\| \\
&\leq \left(\frac{t_1}{t_0}\right)\left(\frac{t_1}{t_0}\right) \|x_{k-2} - x_*\| \\
&\leq \left(\frac{t_1}{t_0}\right)\left(\frac{t_1}{t_0}\right)\left(\frac{t_1}{t_0}\right) \|x_{k-3} - x_*\| \\
&\vdots \\
&\leq \left(\frac{t_1}{t_0}\right)\left(\frac{t_1}{t_0}\right) \cdots \left(\frac{t_1}{t_0}\right) \|x_0 - x_*\| \\
&= \left(\frac{t_1}{t_0}\right)^k \|x_0 - x_*\|,
\end{aligned}$$

que é o resultado desejado.

**ii)** Se $p \neq 0$, então

$$\begin{aligned}
\|x_k - x_*\| &\leq \frac{t_1}{t_0^{p+1}} \|x_{k-1} - x_*\|^{p+1} \\
&\leq \frac{t_1}{t_0^{p+1}} \left( \frac{t_1}{t_0^{p+1}} \|x_{k-2} - x_*\|^{p+1} \right)^{p+1} \\
&\leq \frac{t_1}{t_0^{p+1}} \left( \frac{t_1}{t_0^{p+1}} \right)^{p+1} \|x_{k-2} - x_*\|^{(p+1)^2} \\
&\leq \frac{t_1}{t_0^{p+1}} \left( \frac{t_1}{t_0^{p+1}} \right)^{p+1} \left( \frac{t_1}{t_0^{p+1}} \|x_{k-3} - x_*\|^{p+1} \right)^{(p+1)^2} \\
&\leq \frac{t_1}{t_0^{p+1}} \left( \frac{t_1}{t_0^{p+1}} \right)^{p+1} \left( \frac{t_1}{t_0^{p+1}} \right)^{(p+1)^2} \|x_{k-3} - x_*\|^{(p+1)^3} \\
&\vdots \\
&\leq \frac{t_1}{t_0^{p+1}} \left( \frac{t_1}{t_0^{p+1}} \right)^{p+1} \left( \frac{t_1}{t_0^{p+1}} \right)^{(p+1)^2} \cdots \left( \frac{t_1}{t_0^{p+1}} \right)^{(p+1)^{k-1}} \|x_0 - x_*\|^{(p+1)^k} \\
&= \left( \frac{t_1}{t_0} \right)^{\left(1+(p+1)^2+\ldots+(p+1)^{k-1}\right)} \left( \frac{1}{t_0^p} \right)^{\left(1+(p+1)^2+\ldots+(p+1)^{k-1}\right)} \|x_0 - x_*\|^{(p+1)^k} \\
&= \left( \frac{t_1}{t_0} \right)^{\frac{(p+1)^k-1}{p}} \left( \frac{1}{t_0^p} \right)^{\frac{(p+1)^k-1}{p}} t_0^{(p+1)^k} \\
&= \left( \frac{t_1}{t_0} \right)^{\frac{(p+1)^k-1}{p}} \frac{1}{t_0^{(p+1)^k-1}} t_0^{(p+1)^k} \\
&= \left( \frac{t_1}{t_0} \right)^{\frac{(p+1)^k-1}{p}} \frac{1}{t_0^{-1}} \\
&= t_0 \left( \frac{t_1}{t_0} \right)^{\frac{(p+1)^k-1}{p}},
\end{aligned}$$

onde a segunda igualdade segue da soma de $k$ termos de uma série geométrica de razão $(p+1)$. Assim:

$$\|x_k - x_*\| \leq t_0 \left( \frac{t_1}{t_0} \right)^{\frac{(p+1)^k-1}{p}} \quad \forall\ k \in \mathbb{N}.$$

Isto conclui a prova do corolário.

$\square$

## 3.4 Análise de convergência

Nesta seção apresentaremos a demonstração o Teorema 3.1.1, principal resultado desta
dissertação. É importante ressaltar que este teorema é similar ao Teorema 2.0.1 do Capítulo 2 desta dissertação, porém, com hipótese sobre a $f$ relaxada.

*Demonstração.* Do Corolário 3.3.1 obtemos que a sequência $\{t_k\}$ está bem definida, é estritamente decrescente e está contida em $(0, r)$. Além disso, $\{t_k\}$ converge superlinearmente para 0, isto é, $\lim_{k\to\infty} \frac{t_{t+1}}{t_k} = 0$. Agora, do Corolário 3.3.2 temos que a sequência $\{x_k\}$ está bem definida, contida em $B(x_*, r)$, converge para $x_*$ e

$$\lim_{k\to+\infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0. \tag{3.17}$$

Do Lema 3.2.2 ganhamos que existe um único zero de $F$ em $B(x_*, \sigma)$. Além disso, Se $\frac{\beta(\rho f'(\rho) - f(\rho))}{\rho(1-\beta(f'(\rho)+1))} = 1$ e $\rho < k$, o Lema 3.2.1 nos garante que $r = \rho$ é o melhor raio de convergência.

Por fim, se adicionarmos $(h3)$, dos Corolários 3.3.1 e 3.3.2 temos que a sequência $\left\{\frac{t_{k+1}}{t_k^{p+1}}\right\}$ é estritamente decrescente e

$$\|x_{k+1} - x_*\| \leq \left[\frac{t_{k+1}}{t_k^{p+1}}\right] \|x_k - x_*\|^{p+1} \leq \left[\frac{t_1}{t_0}^{p+1}\right] \|x_k - x_*\|^{p+1}, \quad k = 0, 1, ....$$

Consequentemente, para $k \geq 0$

$$\|x_k - x_*\| \leq \begin{cases} t_0 \left[\frac{t_1}{t_0}\right]^k, & \text{se} \quad p = 0 \\ t_0 \left[\frac{t_1}{t_0}\right]^{\frac{(p+1)^k - 1}{p}}, & \text{se} \quad p \neq 0 \end{cases}.$$

Concluindo a prova do teorema.

$\square$

# Referências Bibliográficas

[1] A. A. Ribeiro, E. W. Karas. Um curso de otimização. *Cengage Learning Editora* (a ser lançado em 2013).

[2] A. Ben-Israel, T. N. E.Greville. Generalized Inverses. Theory and Applications. 2nd ed. Springer-Verlag, New York, 2003.

[3] A. Izmailov, M. Solodov. Otimização: Condições de Otimalidade, Elementos de Análise Convexa e Dualidade, volume 1. IMPA, Rio de Janeiro, 2005.

[4] C. A. Floudas, P. M. Pardalos. (Eds.). Encyclopedia of Optimization, Second Edition. Springer, 2009. ISBN: 978-0-387-74758-3.

[5] C. Li, K. F. NG. Majorizing functions and convergence of the Gauss-Newton method for convex composite optimization, *SIAM J. Optim*, pp. 613-642, 2007.

[6] C. Li, X. WANG. On convergence of the Gauss-Newton method for convex composite optimization, *Math. Program.*, v. 91, pp. 349-356, 2002. ISSN: 0025-5610. 10.1007/s101070100249.

[7] C. Li, W. Zhang, X. Jin. Convergence and uniqueness properties of Gauss-Newton's method. Comput. Math. Appl. 47(6-7), 1057-1067 (2004).

[8] C. W. Groetsch. Generalized Inverses of Linear Operators, Repersentation and Approximation, Mercel Dekker, Inc, New York and Basel, 1977.

[9] E. L. Lima. *Espaços métricos*, volume 1. IMPA, Rio de Janeiro, Brasil, 2011.

[10] E. L. Lima. *Curso de Análise*, volume 1. IMPA, Rio de Janeiro, Brasil, 2009.

[11] E. Zeidler. *Applied Functional Analysis: Applications to Mathematical Physics.* Applied Mathematical Sciences. Vol. 108. Springer-Verlag, New York. 1995.

[12] E. Zeidler. *Applied Functional Analysis: main principles and their applications*. Applied Mathematical Sciences Vol. 109. Springer-Verlag, New York. 1995.

[13] G. W. Stewart. On the continuity of the generalized inverse, SIAM J. *Appl. Math.*, v. 17, pp. 33-45, 1969. ISSN: 0036-1399.

[14] I. Argyros. On the semilocal convergence of the Gauss-Newton method. *Adv. Nonlinear Var. Inequal.* 8(2), 93-99 (2005).

[15] I. Argyros, S. Hilout. On the local convergence of the Gauss-Newton method. *Punjab Univ. J. Math.* 41, 23-33 (2009).

[16] I. Argyros, S. Hilout. On the Gauss-Newton method. *J. Appl. Math. Comput*. 1-14 (2010).

[17] I. K. Argyros, S. Hilout. Improved generalized diferentiability conditions for Newton-like methods, *J. Complexity*, v. In Press, Corrected Proof, pp., 2010. ISSN: 0885-064X. doi: DOI: 10.1016/j.jco.2009.12.001.

[18] J. Chen, W. Li. Convergence of Gauss-Newton's method and uniqueness of the solution, *Appl. Math. Comput*., v. 170, n. 1, pp. 686-705, 2005. ISSN: 0096-3003.

[19] J. Chen, W. Li. Local convergence results of Gauss-Newton's like method in weak conditions, *J. Numer. Appl. Anal.*, v. 324, n. 2, pp. 1381-1394, 2006. ISSN: 0022-247X. doi: DOI: 10.1016/j.jmaa.2006.01.032.

[20] J. B. Hiriart-Urruty, C. Lemaréchal. *Convex analysis and minimization algorithms I*. Springer-Verlag (1993).

[21] J. E. Dennis, R. B. Schnabel. Numerical methods for unconstrained optimization and nonlinear equations, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[22] J. M. Ortega, W. C. Rheinboldt. Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York,1970.

[23] J. Nocedal, S. J. Wright. Numerical optimization. Springer Series in Operations Research. New York, Springer-Verlag, 1999. ISBN: 0-387- 98793-2.

[24] J. V. Burke, M. C. Ferris. A Gauss-Newton method for convex composite optimization, *Math. Programming*, v. 71, n. 2, Ser. A, pp. 179-194, 1995. ISSN: 0025-5610. doi: 10.1007/BF01585997.

[25] L. U. Kantorovich, V. G. P. Akilo. *Functional Analysis in normed spaces*,Oxford, Pergamon(1964)

[26] L. V. Kantorovich. The principle of the majorant and Newton's method, *Doklady Akad. Nauk SSSR (N.S.)*, v. 76, pp. 17-20, 1951.

[27] M. L. N. Gonçalves. Local convergence of Gauss-Newton method for overdetermined nonlinear systems of equations under a majorant condition. *prepint. http://arxiv.org/pdf/1201.1265.*

[28] O. P. Ferreira. Local convergence of Newton's method in banach space from the viewpoint of the majorant principle. IMA *J. Numer. Appl. Anal.*, 29(3):746-759, 2009.

[29] O. P. Ferreira. Local convergence of Newton's method under majorant condition. *J. Comput. Appl. Math.*, 235(5):1515-1522, 2011.

[30] O. P. Ferreira, B. F. Svaiter. Kantorovich's majorants principle for Newton's method. *Comput. Optim. Appl.* 42(2), 213-229 (2009).

[31] O. P. Ferreira, M. L. N. Gonçalves. Local convergence analysis of inexact Newton-like methods under majorant condition, *Comput. Optim. Appl.*, v. Article in Press, pp. 1-21, 2009. ISSN: 0926-6003.

[32] O. P. Ferreira, M. L. N. Gonçalves, P. R. Oliveira. Local convergence analysis of the Gauss-Newton method under a majorant condition. *J. Complexity*, 27(1):111-125, 2011.

[33] P. A . Wedin. Perturbation theory for pseudo-inverses, *Nordisk Tidskr. Informations behandling (BIT).*' V. 13, pp. 217-232, 1973.

[34] S.L. Campbell, C.D Meyer, Jr. *Generalized Inverses of Linear Transformations*, Dover Publications,Inc, 1991.

[35] W. M. Häussler. A Kantorovich-type convergence analysis for the Gauss-Newton method.*J. Numer. Appl. Anal.* 48(1), 119-125 (1986).

[36] X. Wang. Convergence of Newton's method and uniqueness of the solution of equations in Banach space, IMA *J. Numer. Appl. Anal.*, v. 20, n. 1, pp. 123-134, 2000. ISSN: 0272-4979.