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

Uma generalização do método de Newton escalar motivada pelo método do ponto proximal

Atécio Alves

Teresina - 2015

Atécio Alves

Dissertação de Mestrado:

Uma generalização do método de Newton escalar motivada pelo método do ponto proximal

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

Orientador:
Prof. Dr. Paulo Sérgio Marques dos Santos

Teresina - 2015

FICHA CATALOGRÁFICA

Serviço de Processamento Técnico da Universidade Federal do Piauí

Biblioteca Setorial do CCN

A474g Alves, A.

Uma generalização do método de Newton escalar motivada pelo método do ponto proximal. / Atecio Alves - Teresina, 2015.

44 fl.: il.

Dissertação (Mestrado) - Pós-Graduação em Matemática Universidade Federal do Piauí, 2015.

Orientador: Prof. Dr. Paulo Sérgio Marques dos Santos.

1. Análises. 2. Método de newton Clássico. 3. Distância de Bregman. 4. Método do Ponto Proximal. I. Título

CDD 515

Dedico este trabalho à minha mãe, Aldenora Alves Lima *(In Memoriam)*.

Agradecimentos

Primeiramente agradeço a Deus, por sempre nos dar força para continuar lutando, até que alcancemos a vitória.

Agradeço aos meus pais, Nilo Alves de Sousa e Aldenora Alves Lima*(In Memoriam)*, por todo o sacrifício feito em minha causa e por acreditarem que a Educação é o bem maior.

Agradeço às minhas avós, Maria José e Iranez, paterna e materna, respectivamente.

Agradeço aos meus irmãos, Marcos Jadiel Alves e Márcia Naiane Alves, por todos os momentos em que vivenciamos e vivenciaremos juntos.

Agradeço à minha namorada, Aline de Carvalho Silva(minha Mozão), por todo apoio e companherismo desde o dia que nos conhecemos até os dias de hoje.

Agraceço à Tia Luziene, ao Tio Júnior Mota e a todos que foram meus professores no colégio Educandário Sagrado Coração de Jesus, pois foram eles que me ensinaram a ler e a escrever.

Agradeço aos meus professores do ensino fundamental.

Agradeço aos meus professores do ensino médio, em especial ao professor Paulo Bola.

Agradeço aos meus professores da graduação em matemática: Jurandir, Barnabé, Mário Gomes, Gilvan Lima, João Benício, Marcondes Clark, Marcos Vinícios. Agradeço também ao professor Antônio José e outros tantos que aqui eu tenha deixado de citar.

Agradeço aos meus professores da Pós-Graduação em matemática: Roger Peres, Liane, Newton Santos, Carlos Humberto, João Xavier, Marcos Vinicio, Isaias. Agradeço também aos demais professores da Pós-Graduação em matemática.

Agradeço ao Professor Paulo Sérgio(Paulinho), meu orientador nos projetos de iniciação científica e no mestrado, e à professora Sissy, por todo apoio a mim despendido.

Agradeço ao professor Pedro Soares por compor a banca de defesa.

Agradeço à Dona Rosário por todo o apoio, pelos almoço e jantares promovidos e por ter me dado moradia quando precisei. E aproveito para agradecer também às suas filhas, Jacianny e Jeane, e ao seu pequeno neto João Antônio.

Agradeço aos meus cunhados Ismael, Marcelo, Jéssica, Arianna e Claudinha.

Agradeço aos meus amigos conterrâneos: André, compadre George, Nego Carlos, Luis, Helson, Janilson, Junior Bibição, Juniel, Railson, Jarin, Amaury, Galdino Bola, Marcos Nere, Kátia, Patrícia, Nilvânia, comadre Lucilane, Kellen*(In Memoriam)* e Fátima.

Agradeço aos meus padrinhos de batismo, Edmilson e Marizete.

Agradeço aos meus amigos Ytalo, Jaelson, Denylson, Igor, Marchezan, Laís, Maria Izabel e Anna Jéssica.

Agradeço à Dona Glória, ao senhor João, ao senhor Adalberto e à Dona Leila.

Agradeço ao meus amigos da UFPI: Elianderson, João Neto, Magari, Manário, Naftaly, Jailson, João Antônio, Wiumar, Rui Balizinha, Lucas, Jordan, Carlos Adriano, W. Bruno, Antônio Aguiar(Veludim), Sandoel, Rodolfo, Miguel, Vitaliano, Joel, Rafael(Gordiniz), Mário, Ray, Lucas Vidal, Lucas Machado, Thiago Esteves, Victor, Edilson, Jefferson, Antônio Luiz, Raul Kazan, Narielle, Suiany, Janine, Mariane, Lays Santana, Deborah e Andressa.

Agradeço aos meus Tios, Primos e demais familiares, em especial Moça, tia Dalvelina, tio Francisco, tia Nonata, Merlene, Thais, Eanes, Hélio, Fabi e Cariolanda.

Agradeço aos meus amigos da UESPI: João Paulo, Fabiano, Thiago, Gustavo, Geraldo, Jerfferson, Marcos, Lopes, Rondinelli, Dario, Cleiton, Amanda, Silvia e Eliane.

Agradeço a todos os que tive a oportunidade de conhecer e de conviver.

Agradeço à CAPES, à UFPI e ao IFPI pelo apoio financeiro.

"A árvore plantada pelo seu amor nunca muchará."

Marcos Jadiel.

Resumo

Neste trabalho é analisada uma generalização do método de Newton com uma variável, motivada pelo método do ponto proximal com distância de Bregman, para resolução de equações não-lineares e problemas de otimização. Prova-se a convergência quadrática do método generalizado, onde um caso especial é o método de Newton clássico. São ilustradas as vantagens do método generalizado com relação ao método clássico, através de testes numéricos. Os testes fornecem uma visão de como as instâncias do método generalizado podem ser escolhidas para uma dada equação não-linear. Por último, derivamos uma expressão fechada para o erro assintótico.

Abstract

In this thesis a generalization of the Newton Method is analyzed with one variable, motivated by the Proximal Point Method with Bregman Distance, to solve nonlinear equations and optimization problems. It is proven the quadratic convergence of the generalized method, with the Classical Newton's Method as special case. The advantages of the generalized method over the classic method are illustrated through numerical tests. The Numerical tests provide a view of how variations of the generalized method can be chosen for a given nonlinear equation. At last, it is derived a closed expression for the asymptotic error.

Sumário

Introdução

O Método de Newton é usado para encontar a solução aproximada de problemas de equações não-lineares e é um bom método devido a sua taxa de convergência ser quadrática. Problemas de equações não-lineares aparecem em muitas áreas, tais como otimização, engenharia, física, economia [6], [12], [13]. Sua descrição é simples e é a seguinte: dado um ponto inicial $x_0 \in I$, o método de Newton gera uma sequência $\{x_n\}_{n\in\mathbb{N}}$, através da iteração

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \tag{1}$$

esta chamada de iterada clássica de Newton.

Neste trabalho, considerar-se-á o problema de encontrar uma solução de uma equação não-linear de uma variável:

$$f(x) = 0, \tag{2}$$

onde $f\colon I \subset \mathbb{R} \to \mathbb{R}$ é uma função continuamente derivável.

Observou-se que resolver o problema de otimização

$$\min_{x\in I} \phi(x) \tag{3}$$

onde $\phi\colon \mathbb{R} \to \mathbb{R}$ é uma função duas vezes derivável, pode ser substituido por resolver o problema de encontrar um x tal que $\phi'(x) = 0$, ver [9].

Sob algumas hipóteses na função f, a sequência gerada por (1) está bem definida e converge com taxa quadrática, desde que tomamos um ponto incial x_0 suficientemente próximo da solução, ver [10], [9].

Uma generalização da iterada de Newton (1) é dada por

$$x_{n+1} = s^{-1}\left(s(x_n) - s'(x_n)\frac{f(x_n)}{f'(x_n)}\right), \tag{4}$$

onde s é uma função cuja derivada é lipschitziana em um intervalo que contenha a solução x^* de (2), ver [3]. Neste trabalho mostraremos que a sequência gerada por (4) está bem definida e que converge com taxa quadrática a partir de um ponto inicial x_0 suficientemente próximo a solução x^* de (2). Com a escolha de $s(x) = x$, a equação (4) é igual à equação (1). Por outro lado, (4) é um exemplo de (1), quando f é substituída por $F = f \circ s^{-1}$ e $y_n = s(x_n)$, pois quando aplicamos (1), obtém-se

$$y_{n+1} = y_n - \frac{F(x_n)}{F'(x_n)}. \tag{5}$$

Sendo s invertível numa vizinhança da solução, resultados clássicos do método de Newton implicam que as iterações (4) e (5) podem ser utilizadas para resolver (2), e ambas se reduzem a (1), quando se toma $s(x) = x$. É por isso que (4) é chamada de um método de Newton generalizado.

Para do desenvolvimento deste trabalho, usaremos a seguinte estrutura: No Capítulo 1, apresentaremos definições e resultados de funções contínuas, funções convexas e ponto fixo que serão ultilizados no desenvolvimento deste trabalho. No Capítulo 2 é definido o Método do Ponto Proximal com Distência de Bregman (MPPDB) e é discutido a boa definição e a convergência do método. No Capítulo 3, apresentamos o Método de Newton Generalizado, e provamos a convergência quadrática do método, e apresentamos implementaç ies do Método de Newton Generalizado com diferente funções instâncias s. No Capítulo 4, relatamos as considerações finais.

Capítulo 1

Noções Preliminares

Neste capítulo, serão apresentas as definições e os resultados básicos, com base em [10], [4] e [8], que serão utilizados no desenvolvimento deste trabalho.

1.1 Funções contínuas e resultados

Nesta seção, apresentar-se-á algumas definições sobre continuidade de uma função f e alguns resultados importantes.

Definição 1. *Uma função* $f\colon I \subset \mathbb{R} \to \mathbb{R}$*, diz-se contínua no ponto* $a \in I$ *quando,*

$$\forall\ \epsilon > 0\ \exists\ \delta > 0, \textit{tais que}\ x \in I,\ |x - a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon.$$

Diz-se que $f\colon I \subset \mathbb{R} \to \mathbb{R}$ é contínua quando f é contínua em todos os pontos $a \in I$.

Teorema 1. *Dada uma função* $f\colon I \subset \mathbb{R} \to \mathbb{R}$*. Se* f *é contínua num ponto* $a \in I$ *se, e somente se, para toda sequência de pontos* $x_n \in I$ *com* $\lim x_n = a$*, se tenha* $\lim f(x_n) = f(a)$*.*

Demonstração. Ver [10]. □

Teorema 2. (Teorema do valor intermediário) *Seja* $f\colon [a, b] \to \mathbb{R}$ *uma função contínua. Se* $f(a) < d < f(b)$*, então existe um* $c \in (a, b)$ *tal que* $f(c) = d$*.*

Demonstração. Ver [10]. □

Corolário 1. *Se* $I \subset \mathbb{R}$ *é um intervalo e* $f\colon I \to \mathbb{R}$ *é contínua, então a imagem de* f *do conjunto* I*,* $f(I)$*, é um intervalo.*

Demonstração. Ver [10]. □

Corolário 2. *Seja* $f\colon [a,b] \to \mathbb{R}$ *uma função contínua em* (a,b). *Se* $f(x) \neq 0$ *para qualquer* $x \in (a,b)$, *então* $f(x) > 0$ *ou* $f(x) < 0$ *para todo* $x \in (a,b)$.

Demonstração. Supondo que existem $x_1,\ x_2 \in (a,b)$, tais que $f(x_1) < 0 < f(x_2)$. Pelo Teorema do valor intermediário, existe um c entre x_1 e x_2 tal que $f(c) = 0$. Absurdo. Portanto, $f(x) > 0$ ou $f(x) < 0$ para todo $x \in (a,b)$. □

O próximo teorema assegura a existência de valores máximos e mínimos de uma função contínua quando o seu domínio é um conjunto compacto.

Teorema 3. **(Teorema de Weierstrass)** *Sejam* $f\colon I \to \mathbb{R}$ *uma função contínua e* I *um conjunto compacto. Então, existem* $x^*,\ \tilde{x} \in I$, *tais que* $f(x^*) \leqslant f(x) \leqslant f(\tilde{x})$ *para todo* $x \in I$.

Demonstração. Ver [8]. □

Através da seguinte definição e corolário pode-se garantir a existência de um minimizador x^* mesmo quando I não é compacto.

Definição 2. *O conjunto de nível da função* $f\colon I \subset \mathbb{R} \to \mathbb{R}$ *associado a* $c \in \mathbb{R}$, *é o conjunto dado por*

$$L_{f,I}(c) = \{x \in I | f(x) \leqslant c\}.$$

Corolário 3. *Seja* $f\colon I \subset \mathbb{R} \to \mathbb{R}$ *contínua em* I. *Se existe um* $c \in \mathbb{R}$ *tal que o conjunto de nível* $L_{f,I}(c)$ *é não-vazio e compacto, então existe* $x^* \in I$ *tal que* $f(x^*) \leqslant f(x)$ *para todo* $x \in I$.

Demonstração. Ver [8]. □

O Teorema 4, a seguir, afirma que se $I = [a,b]$ é um intervalo fechado, então $f(I)$ também é um intervalo fechado. Mas se I não for um intervalo do tipo $[a,b]$, $f(I)$ pode não ter as mesmas características de I.

Exemplo 1. *Dado* $I = (0,7)$ *e* $f\colon I \to \mathbb{R}$, *com* $f(x) = \operatorname{sen} x$, *temos que* $f(I) = [-1,1]$ *é um intervalo fechado, mas* I *é um intervalo aberto.*

Teorema 4. *Seja* $I \subset \mathbb{R}$ *um intervalo fechado, a imagem* $f(I)$ *por uma função contínua* $f\colon I \to \mathbb{R}$ *é também um intervalo fechado.*

Demonstração. Pelo Corolário 1, $f(I)$ é um intervalo. Tomando-se $\{y_n\}_{n\in\mathbb{N}}$ uma sequência de $f(I)$. Para cada $n \in \mathbb{N}$, temos que existe um $x_n \in I$, tal que $y_n = f(x_n)$. A sequência $\{x_n\}_{n\in\mathbb{N}}$ é limitada pois I é limitado por hipótese. Pelo Teorema de Bolzano-Weierstrass, ver [10], $\{x_n\}_{n\in\mathbb{N}}$ possui uma subsequência convergente. Seja $\{x_{n'}\}_{n'\in\mathbb{N}}$ uma subsequência de $\{x_n\}_{n\in\mathbb{N}}$, tal que $\lim x_{n'} = a$. Note que $a \in I$ pois I é fechado por hipótese. Sendo f contínua, então $\lim\limits_{n'\to\infty} f(x_{n'}) = f(a)$. Note que $f(a) \in f(I)$ pois $a \in I$. Então $\{y_{n'}\}_{n'\in\mathbb{N}}$, com $y_{n'} = f(x_{n'})$, é uma subsequência de $\{y_n\}_{n\in\mathbb{N}}$ convergente. Logo toda sequência de $f(I)$ possui um subsequência convergente e seu limite pertece a $f(I)$. Portanto $f(I)$ é fechado. □

Teorema 5. *Se $I \subset \mathbb{R}$ é um intervalo fechado, então toda bijeção contínua $f\colon I \to J \subset \mathbb{R}$ tem inversa $f^{-1}\colon J \to I$ contínua.*

Demonstração. Ver [10]. □

Definição 3. *Diz-se que $f\colon I \subset \mathbb{R} \to \mathbb{R}$ é derivável no ponto $a \in I$ se existir o seguinte limite*

$$\lim_{x\to a} \frac{f(x) - f(a)}{x - a} = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h} = f'(a).$$

Quando existe $f'(x)$, para todo $x \in I$ diz-se que $f\colon I \subset \mathbb{R} \to \mathbb{R}$ é derivável em I.

Teorema 6. *Se f é derivável em um ponto a, então é contínua em a.*

Demonstração. Ver [10]. □

Definição 4. *Diz-se que um ponto x^* pertencente ao domínio de uma função f é ponto crítico quando $f'(x^*) = 0$.*

Teorema 7. *Seja $f\colon I \subset \mathbb{R} \to \mathbb{R}$. Dado $a \in I$, se existe $f'(a)$ e a é um ponto de máximo ou mínimo de f, então $f'(a) = 0$.*

Demonstração. Ver [10]. □

A função derivada, mesmo sendo descontínua, goza da propriedade do valor intermediário, como é dado a seguir.

Teorema 8. (Darboux) *Seja $f\colon [a, b] \to \mathbb{R}$ derivável. Se $f'(a) < d < f'(b)$, então existe $c \in (a, b)$, tal que $f'(c) = d$.*

Demonstração. Ver [10]. □

O Teorema acima garante que, se $f'(x) \neq 0$, então ou $f'(x) > 0$ ou $f'(x) < 0$ para todo x do domínio da função f. De fato, se existissem a, b tais que $f'(a) < 0 < f'(b)$, pelo Teorema de Darboux existiria $c \in (a, b)$ tal que $f'(c) = 0$. Isso contradiz a hipótese de $f'(x) \neq 0$.

Teorema 9. (Teorema do valor médio) *Seja* $f\colon [a, b] \to \mathbb{R}$ *contínua. Se* f *é derivável em* (a, b)*, então exite um* $c \in (a, b)$*, tal que* $f(b) - f(a) = f'(c)(b - a)$*.*

Demonstração. Ver [10]. □

Teorema 10. *Sejam* $I \subset \mathbb{R}$ *um intervalo e* $f\colon I \to \mathbb{R}$ *uma função derivável. Então,* f *é monótona não-decrescente(não-crescente) se, e somente se,* $f'(x) \geqslant 0$*(*$f'(x) \leqslant 0$*) para qualquer* $x \in I$*.*

Demonstração. Ver [10]. □

Teorema 11. *Sejam* $I \subset \mathbb{R}$ *um intervalo, e* $f\colon I \to \mathbb{R}$ *uma função derivável. Se* $f'(x) \neq 0$ *para todo* $x \in I$*, então* f *é uma bijeção, crescente ou decrescente, de* I *sobre um intervalo* J *e sua inversa* $f^{-1}\colon J \to I$ *é derivável , com* $(f^{-1})'(y) = \dfrac{1}{f'(x)}$ *para todo* $y = f(x) \in J$*.*

Demonstração. Ver [10]. □

A n-ésima derivada de uma função f num ponto a será representada por $f^{(n)}(a)$. Para $n = 1,\ 2$ e 3 escreve-se $f'(a)$, $f''(a) = [f'(a)]'$ e $f'''(a) = [f''(a)]'$, respectivamente.

Definição 5. *Diz-se que* $f\colon I \to \mathbb{R}$ *é uma função de classe* C^n *quando* f *é* n *vezes derivável e, além disso,* $f^{(n)}\colon I \to \mathbb{R}$ *é contínua.*

Teorema 12. (Fómula de Taylor, com resto de Lagrange) *Seja* $f\colon [a, b] \to \mathbb{R}$ *n vezes derivável no intervalo* $[a, b]$*, com* $f^{(n-1)}$ *derivável em* (a, b)*. Existe um* $c \in (a, b)$ *tal que*

$$f(b) = f(a) + f'(a)(b-a) + \frac{f''(a)}{2!}(b-a)^2 + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} + \frac{f^{(n)}(c)}{n!}(b-a)^n.$$

Demonstração. Ver [10]. □

1.2 Funções convexas

Nesta seção, serão apresentadas definições e propriedades de funções convexas em um intervalo $I \subset \mathbb{R}$, as quais podem ser encontrada em [10], [8].

Definição 6. *Seja* $I \subset \mathbb{R}$ *um intervalo. Uma função* $f\colon I \to \mathbb{R}$ *é dita convexa quando para todo par de ponto* $x, y \in I$ *e* $\alpha \in [0,1]$ *tem-se*

$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y). \tag{1.1}$$

A função $f\colon I \to \mathbb{R}$ diz-se estritamente convexa quando a desigualdade (1.1) é estrita sempre que $x \neq y$ e $\alpha \in (0,1)$. A função $f\colon I \to \mathbb{R}$ diz-se fortemente convexa com módulo $\gamma > 0$, quando para quaisquer $x,\ y \in I$ e $\alpha \in [0,1]$, tem-se

$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y) - \gamma\alpha(1-\alpha)|x-y|^2.$$

Observa-se que uma função fortemente convexa é estritamente convexa, e uma estritamente convexa é convexa.

Exemplo 2. *A função* $f\colon \mathbb{R} \to \mathbb{R}$*,* $f(x) = x^2$*, é fortemente convexa.*

Exemplo 3. *A função* $f\colon \mathbb{R} \to \mathbb{R}$*,* $f(x) = e^x$*, é estritamente convexa, mas não é fortemente convexa.*

Exemplo 4. *A função* $f\colon \mathbb{R} \to \mathbb{R}$*,* $f(x) = x$*, é convexa, mas não é estritamente convexa.*

A convexidade de uma função é importante quando se quer encontrar um ponto $x^* \in I$, tal que $f(x^*) \leqslant f(x)$ para todo $x \in I$, ou seja, encontra a solução do seguinte problema

$$\min f(x) \text{ sujeito a } x \in I, \tag{1.2}$$

que é dito um problema de minimização convexa quando $f\colon I \to \mathbb{R}$ é uma função convexa.

Teorema 13. (Teorema de minimização convexa) *Seja* $I \subset \mathbb{R}$ *um intervalo e* $f\colon I \to \mathbb{R}$ *uma função convexa em* I*. Então todo minimizador local do problema (1.2) é global. Além disso, o conjunto de minimizadores é convexo. Se* f *é estritamente convexa, o minimizador é único.*

Demonstração. Ver [8], Teorema 3.1.5 $\square$

Teorema 14. *Se* $f\colon I \to \mathbb{R}$ *é uma função convexa, então o conjunto de nível*

$$L_{f,I}(c) = \{x \in I | f(x) \leqslant c\}$$

é convexo para todo $c \in \mathbb{R}$.

Demonstração. Ver [8], Teorema 3.4.1. □

Teorema 15. *Seja* $f\colon I \to \mathbb{R}$ *uma função convexa. Se existe um* $c \in \mathbb{R}$ *tal que* $L_{f,I}(c)$ *é limitado, então todo conjunto de nível de* f *é limitado.*

Demonstração. Segue a prova do teorema por contradição. Supondo que existe um $d \in \mathbb{R}$, tal que $L_{f,I}(d)$ é ilimitado. Pelo Teorema 14, $L_{f,I}(d)$ é convexo. Como $L_{f,I}(d) \subset \mathbb{R}$ é convexo e ilimitado, logo uma das semi retas $x+t$ ou $x-t$, $t > 0$ e qualquer $x \in L_{f,I}(d)$, está contida em $L_{f,I}(d)$. Tomando $v \in \{-1, 1\}$ tal que $x + tv \in L_{f,I}(d)$,para qualquer $x \in L_{f,I}(d)$ e $t > 0$. Demonstrar-se-á que, para qualquer $\lambda \in \mathbb{R}$ o conjunto $L_{f,I}(\lambda)$ é ilimitado. Se $\lambda > d$, tem-se que $L_{f,I}(d) \subset L_{f,I}(\lambda)$, logo $L_{f,I}(\lambda)$ é ilimitado pois $L_{f,I}(d)$ é ilimitado. Por outro lado, supondo que $\lambda < d$, daí temos que $L_{f,I}(\lambda) \subset L_{f,I}(d)$. Tomando um $x \in L_{f,I}(\lambda)$ e a semi reta $x + tv$, $t > 0$, logo $x + tv \in L_{f,I}(d)\ \ \forall t > 0$. Então

$$\begin{aligned} f(x+tv) &= f\left(\left(1-\frac{1}{\alpha}\right)x + \frac{1}{\alpha}(x+\alpha tv)\right) \\ f(x+tv) &\leqslant \left(1-\frac{1}{\alpha}\right)f(x) + \frac{1}{\alpha}f(x+\alpha tv) \\ f(x+tv) &\leqslant \left(1-\frac{1}{\alpha}\right)f(x) + \frac{1}{\alpha}d \end{aligned}$$

para qualquer $\alpha \geqslant 1$. Fazendo $\alpha \to \infty$, tem-se:

$$f(x+tv) \leqslant f(x) \leqslant \lambda \qquad \forall t > 0.$$

Logo, $x + tv \in L_{f,I}(\lambda)\ \ \forall t > 0$, assim $L_{f,I}(\lambda)$ é ilimitado, para qualquer $\lambda \in \mathbb{R}$. Isso contradiz a hipótese de existir um $c \in \mathbb{R}$, tal que $L_{f,I}(c)$ é limitado. Portanto, $L_{f,I}(d)$ é limitado para todo $d \in \mathbb{R}$. □

Teorema 16. *Se* $f\colon I \to \mathbb{R}$ *é convexa no interior de* I, *então existem as derivadas laterais* $f'_+(c)$ *e* $f'_-(c)$ *em todo ponto* $c \in \mathrm{int}(I)$.

Demonstração. Ver cap. 9 de [10]. □

Corolário 4. *Uma função convexa* $f\colon I \to \mathbb{R}$ *é contínua em todo ponto* c *no interior de* I.

Demonstração. Seja $c \in \text{int}(I)$. Logo,

$$\lim_{x \to c+} f(x) - f(c) = \lim_{x \to c+} \frac{f(x) - f(c)}{x - c}(x - c) = \lim_{x \to c+} \frac{f(x) - f(c)}{x - c} \lim_{x \to c+} (x - c) = 0$$

pois $\lim_{x \to c+} \dfrac{f(x) - f(c)}{x - c}$ existe, pelo teorema 16 , e $\lim_{x \to c+} (x - c) = 0$. Por outro lado

$$\lim_{x \to c-} f(x) - f(c) = \lim_{x \to c-} \frac{f(x) - f(c)}{x - c}(x - c) = \lim_{x \to c-} \frac{f(x) - f(c)}{x - c} \lim_{x \to c-} (x - c) = 0$$

pois $\lim_{x \to c-} \dfrac{f(x) - f(c)}{x - c}$ existe, pelo teorema 16, e $\lim_{x \to c-} (x - c) = 0$. Então $\lim_{x \to c} f(x) - f(c) = 0 \iff \lim_{x \to c} f(x) = f(c)$.

Portanto f é contínua para todo $c \in \text{int}(I)$. □

O próximo Teorema dará uma caracterização para uma função convexa derivável.

Teorema 17. *Seja $I \subset \mathbb{R}$ um intervalo aberto e $f\colon I \to \mathbb{R}$ uma função derivável em I. Então, as seguinte afirmações são equivalentes:*

(1) f *é convexa;*

(2) a derivada $f'\colon I \to \mathbb{R}$ *é monótona não-decrescente;*

(3) para quaisquer $a, x \in I$ *tem-se* $f(x) \geqslant f(a) + f'(a)(x - a)$.

Quando f' *é duas vezes derivável em* I*, as afirmações acima também são equivalentes a*

(4) $f''(x) \geqslant 0$ *para todo* $x \in I$.

Demonstração. Ver cap. 9 de [10]. □

Corolário 5. *Seja* $f\colon I \to \mathbb{R}$ *uma função convexa derivável. Então, todo ponto crítico de* f *é um ponto de mínimo absoluto.*

Demonstração. Seja x^* um ponto crítico, ou seja, $f'(x^*) = 0$. Pela condição (3) do Teorema 17, $f(x) \geqslant f(x^*)$, para todo $x \in I$. Portanto, x^* é mínimo absoluto. □

Teorema 18. *Seja $I \subset \mathbb{R}$ um interevalo aberto e $f\colon I \to \mathbb{R}$ uma função derivável em I. Então as seguinte afirmações são equivalentes:*

(1) f *é estritamente convexa;*

(2) a derivada $f' \colon I \to \mathbb{R}$ *é monótona crescente;*

(3) para quaisquer $a, x \in I$*, tal que* $a \neq x$ *tem-se* $f(x) > f(a) + f'(a)(x - a)$*.*

Demonstração. Ver [8]. □

O exemplo a seguir demonstra que convexidade estrita de f implica que f' é estritamente crescente, mas não implica que $f''(x) > 0$ para todo $x \in I$.

Exemplo 5. *Seja* $f \colon \mathbb{R} \to \mathbb{R}$ *uma função, com lei de formação* $f(x) = x^4$*. A função* f *é estritamente convexa, mas* $f''(x) = 12x^2$ *é nula no ponto* $x = 0$*.*

Teorema 19. *Toda função* f *duas vezes derivável em um intervalo aberto* $I \subset \mathbb{R}$ *tal que* $f''(x) > 0$ *para todo* $x \in I$ *é estritamente convexa. Além disso, se* f *possui um mínimo, este é único.*

Demonstração. Dados $x, a \in I$ com $a < x$, pelo Teorema de Taylor, com resto de Lagrange, existe um $c \in (a, x)$, tal que

$$f(x) = f(a) + f'(a)(x - a) + \frac{f''(c)}{2}(x - a)^2$$

como $f''(c) > 0$ e $x - a > 0$, tem-se

$$f(x) > f(a) + f'(a)(x - a).$$

Pelo Teorema 18, f é estritamente convexa. Supondo que x^* seja um ponto de mínimo, logo $f'(x^*) = 0$. Fazendo $a = x^*$ na desigualdade acima, tem-se que $f(x) > f(x^*)$ para todo $x \in I$ com $x \neq x^*$. Portanto, o ponto de mínimo é único. □

Para funções diferenciáveis fortemente convexas, existem caracterizações semelhantes as dos Teoremas 17 e 18.

Teorema 20. *Seja* $I \subset \mathbb{R}$ *um interevalo aberto e* $f \colon I \to \mathbb{R}$ *uma função derivável em* I*. Então, as seguinte afirmações são equivalentes:*

(1) f *é fortemente convexa em* I *com módulo* $\gamma > 0$*;*

(2) para quaisquer $a, x \in I$ *tem-se* $f(x) \geqslant f(a) + f'(a)(x - a) + \gamma|x - a|^2$*.*

(3) para quaisquer $a, x \in I$ *tem-se* $(f'(y) - f'(x))(y - x) \geqslant 2\gamma|y - x|^2$*.*

Quando f' *é duas vezes derivável em* I*, as afirmações acima também são equivalentes a*

(4) $f''(x) \geqslant 2\gamma$ *para todo* $x \in I$.

Demonstração. Ver [8]. □

O item *(4)*, do teorema acima diz que toda função fortemente convexa duas vezes derivável tem a segunda derivada positiva, mas nem toda função com segunda derivada positiva satisfaz a condição $f''(x) \geqslant 2\gamma$, para algum $\gamma > 0$. De fato, seja $f\colon \mathbb{R} \to \mathbb{R}$ uma função, com $f(x) = e^x$, temos que $f''(x) = e^x > 0$ para todo $x \in \mathbb{R}$, mas $\lim_{x \to -\infty} e^x = 0$, logo não existe $\gamma > 0$ tal que $f''(x) \geqslant 2\gamma$.

Definição 7. *Seja* $f\colon \mathbb{R} \to \mathbb{R}$ *uma função convexa. Diz-se que* $z \in \mathbb{R}$ *é um subgradiente de* f *no ponto* x *se*

$$f(y) \geqslant f(x) + z(y - x), \ \forall y.$$

O conjunto de todos os subgradientes de f no ponto x chama-se *subdiferencial* de f em x, que será denotado por $\partial f(x)$, ou seja,

$$\partial f(x) = \{z \in \mathbb{R} | \ f(y) \geqslant f(x) + z(y - x), \ \forall y\}.$$

O subgradiente define uma aproximação linear de f cujo gráfico fica abaixo daquele de f e cujo valor coincide com f no ponto x. No caso em que a função f é derivável o subgradiente é único, conforme o próximo resultado.

Teorema 21. *Seja* $f\colon \mathbb{R} \to \mathbb{R}$ *uma função convexa é derivável no ponto* $x \in \mathbb{R}$ *se, e somente se,* $\partial f(x)$ *contém apenas um elemento. Neste caso,* $\partial f(x) = \{f'(x)\}$.

Demonstração. Ver [8]. □

1.3 Ponto Fixo

Um ponto fixo de uma função é um número que quando aplicado na função o valor obtido é igual a este número, ver [4].

Definição 8. *Se* g *é uma função, um ponto fixo de* g *é um ponto* x^* *tal que* $g(x^*) = x^*$.

Problemas de encontrar raiz e problemas de ponto fixo são equivalentes no seguinte sentido:

• Seja f um função com x^* raiz, ou seja, $f(x^*) = 0$, podemos definir uma função g com ponto fixo em x^* da seguinte maneira

$$g(x) = x - f(x);$$

• Consequentemente, se a função g tem ponto fixo em x^*, então a função definida por

$$f(x) = x - g(x)$$

tem um zero em x^*.

O teorema a seguir dá condições suficientes para a existência e unicidade de um ponto fixo.

Teorema 22.

(i) Se $g \in C[a, b]$ e $g(x) \in [a, b]$ para todo $x \in [a, b]$, então g tem um ponto fixo em $[a, b]$.

(ii) Se, além disso, exitem a derivada $g'(x)$ para todo $x \in (a, b)$ e uma constante positiva $K < 1$, tal que

$$|g'(x)| \leqslant K < 1, \textit{para todo } x \in (a, b),$$

então g possui um único ponto fixo em $[a, b]$.

Demonstração.

(i) Se $g(a) = a$ ou $g(b) = b$, então g tem um ponto fixo. Caso contrário, tem-se que $g(a) > a \iff g(a) - a > 0$ e $g(b) < b \iff g(b) - b < 0$. A função $h \colon [a, b] \to \mathbb{R}$, definida por:

$$h(x) = g(x) - x$$

é contínua. Logo,

$$h(a) = g(a) - a > 0 \text{ e } h(b) = g(b) - b < 0 \iff h(b) < 0 < h(a).$$

Pelo Teorema do valor intermédiário existe um $x^* \in (a, b)$ tal que $h(x^*) = 0$. Assim, $h(x^*) = g(x^*) - x^* = 0 \Rightarrow g(x^*) = x^*$. Portanto, g tem um ponto fixo.

(ii) Sejam x^* e y^*, com $x^* \neq y^*$, pontos fixo de g. Pelo Teorema do valor médio, existe um c entre x^* e y^*, tal que

$$g(x^*) - g(y^*) = g'(c) \cdot (x^* - y^*).$$

Passando ao módulo e usando a hipótese que $|g'(x)| \leqslant K < 1,$ para todo $x \in (a, b)$. Tem-se

$$|x^* - y^*| = |g(x^*) - g(y^*)| = |g'(c)| \cdot |x^* - y^*| < |x^* - y^*|,$$

contradição, então $x^* = y^*$. Portanto, o ponto fixo de g é único.

□

Teorema 23. *Seja* $g \in C[a, b]$, *tal que* $g(x) \in [a, b]$ *para todo* $x \in [a, b]$. *Se, além disso, exitem* $g'(x)$ *em* (a, b) *e uma constante* $0 < K < 1$ *tal que*

$$|g'(x)| \leqslant K < 1, \textit{para todo } x \in (a, b),$$

então, g *possui um único ponto fixo* $x^* \in [a, b]$, *e para qualquer* $x_0 \in [a, b]$, *a sequência definida por* $x_{n+1} = g(x_n), n \geqslant 0$, *converge para* x^*.

Demonstração. Pelo Teorema 22, existe um único ponto $x^* \in [a, b]$, tal que $g(x^*) = x^*$. A sequência $\{x_n\}_{n \in \mathbb{N}}$ está bem definida, visto que $g \colon [a, b] \to [a, b]$. Pelo Teorema do valor médio, para cada $n \in \mathbb{N}$ existe c_n entre x_n e x^*, tal que

$$g(x_n) - g(x^*) = g'(c_n) \cdot (x_n - x^*).$$

Passando ao módulo na igualdade acima e tomando $x_{n+1} = g(x_n)$ e $g(x^*) = x^*$, obtemos a seguinte equação

$$|x_{n+1} - x^*| = |g'(c_n)| \cdot |x_n - x^*|.$$

Tem-se que $|g'(c_n)| \leqslant K$, $\forall n \in \mathbb{N}$, Logo

$$|x_{n+1} - x^*| \leqslant K \cdot |x_n - x^*|.$$

Então,

$$|x_1 - x^*| \leqslant K \cdot |x_0 - x^*|$$

$$|x_2 - x^*| \leqslant K \cdot |x_1 - x^*|$$

$$|x_3 - x^*| \leqslant K \cdot |x_2 - x^*|$$
$$\vdots$$
$$|x_{n+1} - x^*| \leqslant K \cdot |x_n - x^*|.$$

Então

$$|x_{n+1} - x^*| \leqslant K^{n+1} \cdot |x_0 - x^*|. \tag{1.3}$$

Passando ao limite na desigualdade (1.3) com $n \to \infty$, tem-se

$$0 \leqslant \lim_{n\to\infty} |x_{n+1} - x^*| \leqslant \lim_{n\to\infty} K^{n+1} \cdot |x_0 - x^*| = 0$$

pois $0 < K < 1$. Então $\lim\limits_{n\to\infty} |x_{n+1} - x^*| = 0 \iff \lim\limits_{n\to\infty} x_{n+1} = x^*$. Portanto, $\{x_n\}_{n\in\mathbb{N}}$ converge para o único ponto fixo x^*.

$\square$

Dada uma função $f\colon I \to \mathbb{R}$, de classe C^1 no intervalo I, com $f'(x) \neq 0$ para todo $x \in I$. Toma-se um valor inicial $x_0 \in I$ e a sequência $\{x_n\}_{n\in\mathbb{N}}$ gerada por

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \tag{1.4}$$

é o método de Newnton clássico, usado para a obtenção de valores aproximados da raiz da equação $f(x) = 0$. Se a sequência $\{x_n\}_{n\in\mathbb{N}}$ convergir, o seu limite x^* será uma raiz da equação $f(x) = 0$. De fato, passando ao limite com $n \to \infty$ na equação (1.4), tem-se

$$x^* = x^* - \frac{f(x^*)}{f'(x^*)} \Rightarrow f(x^*) = 0.$$

As raízes da equação $f(x) = 0$ são os pontos fixos da função $g\colon I \to \mathbb{R}$, definida por

$$g(x) = x - \frac{f(x)}{f'(x)}.$$

Definimos taxa de convergência de um método iterativo, visto que ela é um dos principais indicadores da eficiência de um método iterativo.

Supondo que $\{x_n\}_{n\in\mathbb{N}}$ é uma sequência que converge para x^*, com $x_n \neq x^*$ para todo $n \in \mathbb{N}$. Se existem duas constantes positivas α e λ, tais que

$$\lim_{n\to\infty} \frac{|x_{n+1} - x^*|}{|x_n - x^*|^{\alpha}} = \lambda,$$

então dizemos que a sequência $\{x_n\}_{n\in\mathbb{N}}$ converge para x^* com ordem α e com erro assintótico λ. Destaca-se dois casos especiais:

(i) Se $\alpha = 1$ e $\lambda < 1$, então a sequência é dita linearmente convergente ;

(ii) Se $\alpha = 2$, então a sequência é dita quadraticamente convergente.

O resultado a seguir dará condições suficientes para que a sequência gerada pela iteração do ponto fixo convirja linearmente.

Teorema 24. . *Seja* $g \in C^1[a,b]$, *tal que* $g(x) \in [a,b]$ *para todo* $x \in [a,b]$, *com* $|g'(x)| \leqslant K$ *para todo* $x \in (a,b)$, *onde* K *é uma constante tal que* $0 < K < 1$. *Se* $g'(x^*) \neq 0$, *então para qualquer número* $x_0 \neq x^*$ *em* $[a,b]$, *a sequência* $\{x_n\}_{n\in\mathbb{N}}$ *que é gerada por* $x_{n+1} = g(x_n)$ *converge linearmente para o único ponto fixo* x^* *em* $[a,b]$.

Demonstração. Pelos Teoremas 22 e 23, g tem um único ponto fixo x^*, e a sequência $\{x_n\}_{n\in\mathbb{N}}$ que é gerada por $x_{n+1} = g(x_n)$ converge para x^*. Proseguindo, será demonstrado que $\{x_n\}_{n\in\mathbb{N}}$ é linearmente convergente para x^*. Pelo Teorema do valor médio, para cada $n \in \mathbb{N}$ existe um c_n entre x_n e x^*, tais que

$$|g(x_n) - g(x^*)| = |g'(c_n)| \cdot |x_n - x^*|.$$

Sabendo que $x_{n+1} = g(x_n)$ e $g(x^*) = x^*$, tem-se

$$|x_{n+1} - x^*| = |g'(c_n)| \cdot |x_n - x^*|$$

$$\frac{|x_{n+1} - x|}{|x_n - x^*|} = |g'(c_n)|. \tag{1.5}$$

Afirmação 1. *A sequência* $\{c_n\}_{n\in\mathbb{N}}$ *converge para* x^*.

De fato, c_n está entre x_n e x^* para todo $n \in \mathbb{N}$. Logo,

$$0 \leqslant |c_n - x^*| \leqslant |x_n - x^*|, \ \forall\, n \in \mathbb{N}. \tag{1.6}$$

Passando ao limite na desigualdade (1.6) com $n \to \infty$, tem-se

$$0 \leqslant \lim_{n\to\infty} |c_n - x^*| \leqslant \lim_{n\to\infty} |x_n - x^*| = 0$$

pois $\lim\limits_{n\to\infty} x_n = x^*$. Então, $\lim\limits_{n\to\infty} |c_n - x^*| = 0 \iff \lim\limits_{n\to\infty} c_n = x^*$.
Agora passando-se ao limite na equação (1.5) e usando a hipótese da continuidade de g', tem-se

$$\lim_{n\to\infty} \frac{|x_{n+1} - x|}{|x_n - x^*|} = \lim_{n\to\infty} |g'(c_n)| = |g'(x^*)| \leqslant K < 1.$$

Portanto, a convergência é linear. $\square$

O próximo resultado apresenta condições adicionais que garantem a convergência quadrática.

Teorema 25. *Seja* $g \in C^2[a,b]$ *e* x^* *a solução da equação* $g(x) = x$*. Supondo que* $g'(x^*) = 0$*, e que existe uma constante positiva* M*, tal que* $|g''(x)| < M$ *em um intervalo* $I \subset [a,b]$ *contendo* x^**. Então exite um* $\delta > 0$*, tal que, para qualquer* $x_0 \in [x^* - \delta, x^* + \delta]$*, a sequência definida por* $x_{n+1} = g(x_n)$*, para todo* $n \geqslant 0$*, converge quadraticamente para* x^**. Além disso, para* n *suficientemente grande, vale a seguinte desigualdade:*

$$|x_{n+1} - x^*| < \frac{M}{2}|x_n - x^*|^2.$$

Demonstração. Por hipótese g' é contínua em (a,b), então para todo $\varepsilon > 0 \ \exists \ \delta > 0$, tal que $|x - x^*| < \delta \Rightarrow |g'(x) - g'(x^*)| < \epsilon$. Tomando $\varepsilon \in (0,1)$, existe um $\delta > 0$, tal que para $x \in (x^* - \delta, x^* + \delta)$ implica que $|g'(x)| < \varepsilon < 1$.

Dado $x \in (x^* - \delta, x^* + \delta)$, pelo Teorema do valor médio, existe um c entre x e x^*, tais que

$$\begin{aligned} |g(x) - g(x^*)| &= |g'(c)| \cdot |x - x^*| \\ |g(x) - x^*| &\leqslant |x - x^*| \\ |g(x) - x^*| &< \delta. \end{aligned}$$

Logo, $g(x) \in (x^* - \delta, x^* + \delta)$. Restrigindo a função g ao intervalo $[x^* - \delta, x^* + \delta]$, pelo Teorema 23, para qualquer ponto inicial $x_0 \in (x^* - \delta, x^* + \delta)$ a sequência $\{x_n\}_{n \in \mathbb{N}}$ converge para x^*.

Mostrar-se-á que a convergência é quadrática. Tem-se $g_{|[x^* - \delta, x^* + \delta]}$ duas vezes derivável com g' contínua no intervalo aberto $(x^* - \delta, x^* + \delta)$. Pela fórmula de Taylon, com resto de Lagrange, para cada $n \in \mathbb{N}$ existe um c_n entre x_n e x^*, tal que

$$g(x_n) = g(x^*) + g'(x^*) \cdot (x_n - x^*) + \frac{g''(c_n)}{2}(x_n - x^*)^2.$$

Sabendo que $g(x_n) = x_{n+1}$, $g(x^*) = x^*$ e $g'(x^*) = 0$, tem-se

$$\begin{aligned} x_{n+1} &= x^* + \frac{g''(c_n)}{2}(x_n - x^*)^2 \\ x_{n+1} - x^* &= \frac{g''(c_n)}{2}(x_n - x^*)^2. \end{aligned}$$

Assim

$$\frac{|x_{n+1} - x^*|}{|x_n - x^*|^2} = \frac{|g''(c_n)|}{2}. \tag{1.7}$$

Portanto, pela afirmação 1 $c_n \to x^*$.

Agora, passando ao limite na equação (1.7) com $n \to \infty$ e sabendo que g'' é contínua, tem-se

$$\lim_{n\to\infty} \frac{|x_{n+1} - x^*|}{|x_n - x^*|^2} = \lim_{n\to\infty} \frac{|g''(c_n)|}{2} = \frac{|g''(x^*)|}{2}.$$

Logo a convergência é quadrática. Por outro lado, como $|g''(x)| < M,\ \forall\, x \in (x^*-\delta, x^*+\delta)$, tem-se

$$\begin{aligned} |x_{n+1} - x^*| &= \frac{|g''(c_n)|}{2}|x_n - x^*|^2 \\ |x_{n+1} - x^*| &< \frac{M}{2}|x_n - x^*|^2. \end{aligned}$$

□

Conforme os Teoremas 24 e 25, para o método do ponto fixo convergir quadraticamente é preciso ter $g(x^*) = x^*$ e $g'(x^*) = 0$. Um exemplo de problema de ponto fixo associado a um problema de encontrar raiz da equação $f(x) = 0$ é adicionar ou subtrair um múltiplo de $f(x)$. Considerando a sequência

$$x_{n+1} = g(x_n), \text{ para todo } n \geqslant 0,$$

para g definida da seguinte forma:

$$g(x) = x - \phi(x)f(x), \tag{1.8}$$

onde ϕ é uma função contínua e diferenciável e escolhida a seguir. Derivando a equação (1.8), obtém-se

$$g'(x) = 1 - \phi'(x)f(x) - \phi(x)f'(x).$$

Tomando $x = x^*$, e sabendo que $g'(x^*) = 0$ e $f(x^*) = 0$, tem-se

$$0 = g'(x^*) = 1 - \phi'(x^*)f(x^*) - \phi(x^*)f'(x^*) = 1 - \phi(x^*)f'(x^*) \iff \phi(x^*) = \frac{1}{f'(x^*)}.$$

Escolhendo $\phi(x) = \dfrac{1}{f'(x)}$, então se é garantido que $\phi(x^*) = \dfrac{1}{f'(x^*)}$ e produz convergência quadrática da iteração clássica de Newton

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Se $f(x^*) = 0$ e $f'(x^*) \neq 0$, então, para valores suficientemente próximos de x^*, o método de Newton converge quadraticamente.

A convergência quadrática pode ser obtida sob outras hipóteses, de acordo com os próximos resultados.

Lema 1. *Seja* (a,b) *um intervalo, e* f *uma função derivável em* (a,b). *Se* f' *é lipschitz* $\gamma \geqslant 0$, *então para quaisquer* x, $y \in (a,b)$, *vale*

$$|f(x) - f(y) - f'(y)(x-y)| \leqslant \frac{\gamma}{2}(x-y)^2. \tag{1.9}$$

Demonstração. Se f' é lipschitz, ou seja, dados x, $y \in (a,b)$ tem-se que $|f(x) - f(y)| \leqslant \gamma|x-y|$, então é f' contínua. Pelo Teorema fundamental do cálculo, tem-se

$$f(x) - f(y) = \int_y^x f'(\theta)d\theta.$$

Assim

$$\begin{aligned}
|f(x) - f(y) - f'(y)(x-y)| &= \left|\int_y^x f'(\theta)d\theta - \int_y^x f'(y)d\theta\right| = \left|\int_y^x (f'(\theta) - f'(y))d\theta\right| \\
&\leqslant \left|\int_y^x |f'(\theta) - f'(y)|d\theta\right| \leqslant \left|\int_y^x \gamma|\theta - y|d\theta\right| \\
&= \gamma\left|\int_y^x |\theta - y|d\theta\right| = \gamma\left|\int_y^x \gamma(\theta - y)d\theta\right| \\
&= \gamma\left|\frac{\theta^2}{2} - y\theta\Big|_y^x\right| = \frac{\gamma}{2}(x-y)^2.
\end{aligned}$$

Usadas as propriedades de funções integráveis adequadas foi demonstrada a validade da desigualdade (1.9). □

Teorema 26. (Taxa de convergência quadrática sob hipótese Lipschitiziana) *Sejam* a *e* b *números reais, tal que* $a < b$. *Dada uma função* $F\colon (a,b) \to \mathbb{R}$, *com* F' *lipschitziana com constante* γ. *Supondo que, para algum* $\rho > 0$, $|F'(x)| \geqslant \rho$ *para todo* $x \in (a,b)$. *Se* $x^* \in (a,b)$ *é uma solução da equação* $F(x) = 0$, *então existe um* $\eta > 0$, *tal que, para* $x_0 \in (x^* - \eta, x^* + \eta)$, *a sequência gerada por*

$$x_{n+1} = x_n - \frac{F(x_n)}{F'(x_n)}, \quad n = 0,\ 1,\ 2, \cdots$$

está bem definida e converge para x^*. *Além disso, para* $n = 0,\ 1,\ 2, \cdots$,

$$|x_{n+1} - x^*| \leqslant \frac{\gamma}{2\rho}|x_n - x^*|^2. \tag{1.10}$$

Demonstração. Definindo $\eta = \min\left\{\{x^* - a, b - x^*, \frac{\rho}{2\gamma}\right\}$. Tomando um ponto inicial $x_0 \in (x^* - \eta, x^* + \eta) \iff |x_0 - x^*| < \eta$, e usando que $F(x^*) = 0$,e o Lema 1, $|F'(x_0)| \geqslant$

$\rho \iff \dfrac{1}{|F'(x_0)|} \leqslant \dfrac{1}{\rho}$, $|x_0 - x^*| < \eta$ e $|x_0 - x^*| < \dfrac{\rho}{2\gamma}$.

$$\begin{aligned}
|x_1 - x^*| &\leqslant \left| x_0 - \frac{F(x_0)}{F'(x_0)} - x^* \right| = \left| (x_0 - x^*) - \frac{F(x_0)}{F'(x_0)} \right| \\
&= \frac{1}{|F'(x_0)|} \cdot |F(x^*) - F(x_0) - F'(x_0)(x^* - x_0)| \\
&\leqslant \frac{\gamma}{2\rho}|x_0 - x^*|^2 = \frac{\gamma}{2\rho}|x_0 - x^*| \cdot |x_0 - x^*| \\
&\leqslant \frac{\gamma}{2\rho} \cdot \frac{\rho}{2\gamma}\eta = \frac{\eta}{4} < \eta.
\end{aligned}$$

Portanto, $x_1 \in (x^* - \eta, x^* + \eta)$. Agora, supondo que $x_n \in (x^* - \eta, x^* + \eta)$, demonstrar-se-á que $x_{n+1} \in (x^* - \eta, x^* + \eta)$.

$$\begin{aligned}
|x_{n+1} - x^*| &\leqslant \frac{\gamma}{2\rho}|x_n - x^*|^2 = \frac{\gamma}{2\rho}|x_n - x^*| \cdot |x_n - x^*| \\
|x_{n+1} - x^*| &\leqslant \frac{\gamma}{2\rho} \cdot \frac{\rho}{2\gamma}\eta = \frac{\eta}{4} \\
|x_{n+1} - x^*| &< \eta,
\end{aligned}$$

logo, $x_{n+1} \in (x^* - \eta, x^* + \eta)$, pelos mesmo argumentos usados para concluir que $x_1 \in (x^* - \eta, x^* + \eta)$.Portando, $x_n \in (x^* - \eta, x^* + \eta)$, $\forall\ n \geqslant 0$. Por outro lado, como $x_n \in (a, b)\ \forall\ n \geqslant 0$, tem-se

$$\begin{aligned}
|x_{n+1} - x^*| &= \left| x_n - \frac{F(x_n)}{F'(x_n)} - x^* \right| \\
&= \left| (x_n - x^*) - \frac{F(x_n)}{F'(x_n)} \right| \\
&= \left| \frac{1}{F'(x_n)} \cdot [F'(x_n)(x_n - x^*) - F(x_n)] \right|.
\end{aligned}$$

Adicionando $F(x^*) = 0$ dentro dos cochetes, tem-se

$$|x_{n+1} - x^*| = \frac{1}{|F'(x_n)|} \cdot |F(x^*) - F(x_n) - F'(x_n)(x^* - x_n)|. \tag{1.11}$$

Como $|F'(x)| \geqslant \rho\ \forall\ x \in (a, b)$ em particular $|F'(x_n)| \geqslant \rho\ \forall\ n \geqslant 0$, então $\dfrac{1}{|F'(x_n)|} \leqslant \dfrac{1}{\rho}$. Aplicando esta desigualdade e o Lema 1 na equação (1.11), tem-se que

$$|x_{n+1} - x^*| \leqslant \frac{\gamma}{2\rho}|x_n - x^*|^2, \quad n = 0,\ 1,\ 2, \cdots. \tag{1.12}$$

Sendo $|x_1 - x^*| \leqslant \dfrac{\gamma}{2\rho}|x_0 - x^*|^2$ e $|x_2 - x^*| \leqslant \dfrac{\gamma}{2\rho}|x_1 - x^*|^2$, então

$$\begin{aligned}
|x_2 - x^*| &\leqslant \frac{\gamma}{2\rho}\left(\frac{\gamma}{2\rho}|x_0 - x^*|^2\right)^2 \\
|x_2 - x^*| &\leqslant \left(\frac{\gamma}{2\rho}\right)^3 |x_0 - x^*|^4.
\end{aligned}$$

Usando este raciocínio, obtém-se

$$
\begin{aligned}
|x_3 - x^*| &\leqslant \left(\frac{\gamma}{2\rho}\right)^7 |x_0 - x^*|^8 \\
|x_4 - x^*| &\leqslant \left(\frac{\gamma}{2\rho}\right)^{15} |x_0 - x^*|^{16} \\
&\vdots \\
|x_{n+1} - x^*| &\leqslant \left(\frac{\gamma}{2\rho}\right)^{2^{(n+1)}-1} |x_0 - x^*|^{2^{(n+1)}} \\
|x_{n+1} - x^*| &\leqslant \left(\frac{\gamma}{2\rho}|x_0 - x^*|\right)^{2^{(n+1)}-1} |x_0 - x^*|.
\end{aligned}
$$

Como $|x_0 - x^*| < \dfrac{\rho}{2\gamma}$, então

$$|x_{n+1} - x^*| \leqslant \left(\frac{1}{4}\right)^{2^{(n+1)}-1} |x_0 - x^*|. \tag{1.13}$$

Passando ao limite na desigualdade (1.13) com $n \to \infty$,

$$0 \leqslant \lim_{n\to\infty} |x_{n+1} - x^*| \leqslant \lim_{n\to\infty} \left(\frac{1}{4}\right)^{2^{(n+1)}-1} |x_0 - x^*| = 0$$

então $\lim\limits_{n\to\infty} |x_{n+1} - x^*| = 0 \iff \lim\limits_{n\to\infty} x_n = x^*$.

Portando a sequência gerada pelo método de Newton converge para a solução da equação $F(x) = 0$. Além disto, a convergência é quadrática por (1.12). □

Capítulo 2

Método do Ponto Proximal com Distância de Bregman

Neste capítulo, serão apresentadas as definições e as propriedades de funções e distância de Bregman e o método do ponto proximal com distância de Bregman. As definições e propriedades deste capítulo serão de muita importância para o desenvolvimento deste trabalho, visto que surge disto a motivação do método de Newton generalizado.

Este capítulo foi baseado em [1], [2], [5], [7].

2.1 Funções e distância de Bregman

As noções de funções e distância de Bregman fora mencionada pioneiramente por Bregman, em 1967, ver [1].

Seja $I \subset \mathbb{R}$ um intervalo fechado, não necessariamente limitado. Considere uma função convexa $h : I \to \mathbb{R}$, derivável em $\operatorname{int}(I)$ e seja $D_h : I \times \operatorname{int}(I) \to \mathbb{R}$ dada por

$$D_h(x, y) = h(x) - h(y) - h'(y)(x - y). \tag{2.1}$$

Diz-se que h é uma função de Bregman (e D_h uma distância de Bregman induzida por h) com zona $\operatorname{int}(I)$, se as seguintes condições forem válidas:

B1) h é derivável em $\operatorname{int}(I)$;

B2) h é estritamente convexa e contínua em I;

B3) Para todo $\delta \in \mathbb{R}$ os conjuntos de níveis parciais $\Gamma_1(y,\delta) \equiv \{x \in I : D_h(x,y) \leqslant \delta\}$ e $\Gamma_2(x,\delta) \equiv \{y \in \text{int}(I) : D_h(x,y) \leqslant \delta\}$ são limitados para todo $y \in \text{int}(I)$ e para todo $x \in I$ respetivamente;

B4) Se $\{y_k\} \subset \text{int}(I)$ converge para $y^* \in \mathbb{R}$, então $y^* \in I$ e $D_h(y^*, y_k)$ converge a 0;

B5) Se $\{x_k\} \subset I$ e $\{y_k\} \subset \text{int}(I)$ são sequências, tal que $\{x_k\}$ é limitada, $\lim\limits_{k\to\infty} y_k = y^*$ e $\lim\limits_{k\to\infty} D_h(x_k, y_k) = 0$, então $\lim\limits_{k\to\infty} x_k = y^*$.

Observação 1. *Sendo* h *estritamente convexa, pelo Teorema 18, dados* $x \in I$ *e* $y \in \text{int}(I)$*,* $x \neq y$*, segue que* $D_h(x,y) \geqslant 0$*. Além disso,* $D_h(x,y) = 0$ *se, e somente se,* $x = y$*. Por esta razão,* D_h *é chamada de distância generalizada.*

A seguir, serão definidas duas importantes subclasses de funções de Bregman: fronteira coerciva e zona coerciva.

Uma função de Bregman h é dita ser *fronteira coerciva* se:

B6) Se $\{y_k\} \subset \text{int}(I)$ é tal que $\lim\limits_{k\to\infty} y_k = y \in \partial I$ (Fronteira de I), então para todo $x \in \text{int}(I)$, tem-se $\lim\limits_{k\to\infty} h'(y_k)(x - y_k) = -\infty$.

Uma função de Bregman h é dita ser *zona coerciva* se

B7) Para todo $y \in \mathbb{R}$, existe $x \in \text{int}(I)$ tal que $h'(x) = y$.

Observação 2. *Se* h *satisfaz B6), tem-se* $\lim\limits_{k\to\infty} D_h(x, y_k) = +\infty$ *para todo* $x \in \text{int}(I)$*. Em outras palavras,* $D_h(x,y)$ *se aproxima do infinito quando* y *se aproxima da fronteira de* I.

Exemplo 6. *Seja* $I = \mathbb{R}$ *e* $h(x) = x^2$*. A função* h *é estritamente convexa, assim*

$$\begin{aligned} D_h(x,y) &= x^2 - y^2 - 2y(x-y) \\ &= x^2 - 2xy - y^2 + 2y^2 \\ &= |x-y|^2 \end{aligned}$$

observa-se que as condições B1)-B5) são verificadas. Portanto, D_h *é uma distância de Bregman.*

Exemplo 7. *Seja* $I = \mathbb{R}_+$ *e* $h\colon \mathbb{R}_+ \to \mathbb{R}$*, com* $h(x) = \begin{cases} x\log x - x, & \textit{se } x > 0, \\ 0, & \textit{se } x = 0. \end{cases}$ *é uma função de Bregman com fronteira e zona coerciva.*

Tem-se que $h'(x) = \log x$, de fato

$$h'(x) = [x \log x - x]' = \log x + x\frac{1}{x} - 1 = \log x.$$

Então,

$$\begin{aligned} D_h(x,y) &= x\log x - x - (y\log y - y) - \log y(x-y) \\ &= x\log x - x - y\log y + y - x\log y + y\log y \\ &= x(\log x - \log y) + y - x \\ &= x\log\left(\frac{x}{y}\right) + y - x. \end{aligned}$$

A seguir vamos deduzir algumas propriedades operatórias da distância de Bregman.

Proposição 1. *Se* h *é uma função de Bregman com zona* $\operatorname{int}(I)$*, então:*

(i) $D_h(x,y) - D_h(x,z) - D_h(z,y) = (h'(y) - h'(z))(z-x)$*, para todos* $x \in I$ *e* $y, z \in \operatorname{int}(I)$.

(ii) $\frac{d}{dx}D_h(x,y) = h'(x) - h'(y)$ *para todos* $x, y \in \operatorname{int}(I)$.

(iii) $D_h(\cdot, y)$ *é estritamente convexa para todo* $y \in \operatorname{int}(I)$.

Demonstração. (i) Da definição de D_h, tem-se

$$D_h(x,y) = h(x) - h(y) - h'(y)(x-y) \tag{2.2}$$

$$D_h(x,z) = h(x) - h(z) - h'(z)(x-z) \tag{2.3}$$

$$D_h(z,y) = h(z) - h(y) - h'(y)(z-y) \tag{2.4}$$

Subtraindo (2.3) e (2.4) de (2.2), o resultado é obtido.

(ii) Por B1), o resultado segue derivando (2.1).

(iii) Seja $y \in \operatorname{int}(I)$ fixo,$r,\ s \in I$, com $r \neq s$, e $t \in (0,1)$, então por B2) tem-se

$$\begin{aligned} D_h((1-t)r + ts, y) &= h((1-t)r + ts) - h(y) - h'(y)((1-t)r + ts - y) \\ &< (1-t)h(r) + th(s) - [(1-t)h(y) + th(y)] - \\ &- (1-t)h'(y)(r-y) - th'(y)(s-y) \\ &= (1-t)D_h(r,y) + tD_h(s,y). \end{aligned}$$

Portanto, $D_h((1-t)r + ts, y) < (1-t)D_h(r,y) + tD_h(s,y)$. □

2.2 Boa definição do método

Sejam $I \subset \mathbb{R}$ um intervalo fechado e $\phi\colon I \to \mathbb{R}$ uma função convexa, considerar-se-á o problema de mininização convexa

$$\min_{x \in I} \phi(x). \tag{2.5}$$

O Método do Ponto Proximal com distância de Bregman (MPPDB) gera uma sequência $\{x_n\}_{n\in\mathbb{N}}$, onde $x_0 \in \text{int}(I)$ é escolhido inicialmente e os pontos seguinte são dados pela iteração

$$x_{n+1} = \text{argmin}\{\phi(x) + \alpha_n D_h(x, x_n)\}, \tag{2.6}$$

onde h é uma função de Bregman com zona int(I) e α_n é um parâmetro que satisfaça

$$0 < \alpha_n < \tilde{\alpha}, \text{ para algum } \tilde{\alpha} > 0.$$

Observação 3. *Quando usamos a função de Bregman do Exemplo 6, temos o Método do Ponto Proximal clássico, este introduzido inicialmente por Martinet [11], para a resolução de problemas de otimização convexa, e posteriormente desenvolvido para problemas de ponto fixo por Rockafellar [15].*

O lema a seguir estabelece a boa definição do método (2.6).

Lema 2. *Sejam ϕ limitada inferiormente em I e h uma distância de Bregman com fronteira coerciva, então a sequência gerada por (2.6) está bem definida e contida no* int(I).

Demonstração. Seja $M \in \mathbb{R}$, tal que $M \leqslant \phi(x)$, para todo $x \in I$. Definindo $\phi_n(x) = \phi(x) + \alpha_n D_h(x, x_n)$, $\phi_n(x)\colon I \to \mathbb{R}$, temos

$$M + \alpha_n D_h(x, x_n) \leqslant \phi_n(x).$$

Seja $L_{\phi_n,I}(c)$ o conjunto de nível de ϕ_n no ponto c. Tomando $z \in L_{\phi_n,I}(c)$, logo

$$\phi_n(z) \leqslant c \Rightarrow D_h(z, x_n) \leqslant \frac{c - M}{\alpha_n},$$

então $z \in \Gamma_1(x_n, \delta)$ com a escolha $\delta = \dfrac{c-M}{\alpha_n}$. Por B3), $L_{\phi_n,I}(c)$ é limitado. Como o conjunto de nível de ϕ_n no ponto c é compacto, então existe um ponto de mínimo $x_{n+1} \in I$, pelo Corolário 3, e o Teorema 13 garante que x_{n+1} é único.

Se x_{n+1} o ponto de mínimo de $\phi_n(x)$, logo $\forall x \in I$

$$\begin{aligned}
\phi_n(x_{n+1}) &\leqslant \phi_n(x) \\
\phi(x_{n+1}) + \alpha_n D_h(x_{n+1}, x_n) &\leqslant \phi(x) + \alpha_n D_h(x, x_n) \\
\phi(x_{n+1}) + \alpha_n h(x_{n+1}) - \alpha_n h'(x_n)x_{n+1} &\leqslant \phi(x) + \alpha_n h(x) - \alpha_n h'(x_n)x \\
\phi(x) + \alpha_n h(x) &\geqslant \phi(x_{n+1}) + \alpha_n h(x_{n+1}) + \alpha_n h'(x_n)(x - x_{n+1}),
\end{aligned}$$

então $\alpha_n h'(x_n) \in \partial(\phi + \alpha_n h)(x_{n+1})$.

A fim de demonstrar que $x_n \in \text{int}(I)$, para todo n, demonstrar-se-á que $\partial(\phi + \alpha_n h)(x) = \emptyset$ para todo $x \in \partial I$. Supondo que existe um $\xi \in \partial(\phi + \alpha_n h)(x)$, onde $x \in \partial I$, tomando $z \in \text{int}I$ e $y_l \in \text{int}I$ definido por

$$y_l = (1 - \epsilon_l)x + \epsilon_l z \iff \epsilon_l(z - x) = y_l - x, \tag{2.7}$$

onde $\epsilon_l \in (0, 1)$. Tem-se que $y_l \to x$, quando $\epsilon_l \to 0$. Se

$$z - y_l = (1 - \epsilon_l)(z - x) \iff (z - x) = \frac{1}{1 - \epsilon_l}(z - y_l), \tag{2.8}$$

usando (2.8) em (2.7), tem-se que

$$y_l - x = \frac{\epsilon_l}{1 - \epsilon_l}(z - y_l). \tag{2.9}$$

Por definição de subdiferencial,

$$\phi(x) + \alpha_n h(x) + \xi(y_l - x) \leqslant \phi(y_l) + \alpha_n h(y_l) \tag{2.10}$$

$$\epsilon_l \xi(z - x) = \xi(y_l - x) \leqslant \phi(y_l) - \phi(x) + \alpha_n(h(y_l) - h(x)), \tag{2.11}$$

Sendo ϕ e h funções convexas, então

$$\phi(y_l) \leqslant (1 - \epsilon_l)\phi(x) + \epsilon_l \phi(z) \iff \phi(y_l) - \phi(x) \leqslant \epsilon_l(\phi(z) - \phi(x)) \tag{2.12}$$

e

$$h(x) \geqslant h(y_l) + h'(y_l)(x - y_l) \iff h(y_l) - h(x) \leqslant h'(y_l)(y_l - x). \tag{2.13}$$

Usando (2.12), (2.13) e (2.9) em (2.11), obtém-se

$$\epsilon_l \xi(z - x) \leqslant \epsilon_l(\phi(z) - \phi(x)) + \alpha_n \frac{\epsilon_l}{1 - \epsilon_l} h'(y_l)(z - y_l). \tag{2.14}$$

De (2.14),

$$\frac{1 - \epsilon_l}{\alpha_n}[\phi(z) - \phi(x) + \xi(z - x)] \leqslant h'(y_l)(z - y_l), \tag{2.15}$$

passando ao limite em (2.15), com $\epsilon_l \to 0$, e por B6,

$$\frac{1}{\alpha_n}[\phi(z) - \phi(x) + \xi(z - x)] \leqslant -\infty.$$

O que contraria o fato de por um lado ser uma constante e por outro lado ser $-\infty$. Logo, $\partial(\phi + \alpha_n h)(x) = \emptyset$ para todo $x \in \partial I$. Portanto, $x_{n+1} \in \mathrm{int}I$, o que mostra a boa definição da sequência gerada por (2.6). □

2.3 Análise de convergência

Assumir-se-á que o problema (2.5) tenha solução x^* e que ϕ é limitada inferiormente em I, ver [7].

Teorema 27. *Se* h *é uma função de Bregman de fronteira coerciva com respeito ao* int(I) *e o problema* (2.5) tem solução x^*, então a sequência $\{x_n\}_{n\in\mathbb{N}}$ gerada por (2.6) converge para x^*.

Demonstração. Pelo Lema 2, a sequência $\{x_n\}$ está bem definida e contida no $\mathrm{int}I$. Seja x^* uma soluçao do problema (2.5) e x_{n+1} o mínimo de ϕ_n, tomando $x = x^*$, $y = x_n$ $z = x_{n+1}$ na Proposição 1*(i)*, tem-se

$$D_h(x^*, x_n) - D_h(x^*, x_{n+1}) - D_h(x_{n+1}, x_n) = (h'(x_n) - h'(x_{n+1}))(x_{n+1} - x^*). \quad (2.16)$$

Sendo $D_h(x, y)$ derivável, pelo Teorema 21, $\partial\alpha_n D(x_{n+1}, x_n) = \{\alpha_n(h'(x_{n+1}) - h'(x_n))\}$, usando as propriedades de subdiferencial, ver [8], tem-se que

$$\begin{aligned} 0 \in \partial[\phi + \alpha_n D(\cdot, x_n)](x_{n+1}) \quad &= \quad \partial\phi(x_{n+1}) + \partial\alpha_n D(x_{n+1}, x_n) \\ 0 = \xi + \alpha_n(h'(x_{n+1}) - h'(x_n)) \quad &\Longleftrightarrow \quad \alpha_n(h'(x_n) - h'(x_{n+1})) = \xi, \quad \xi \in \partial\phi(x_{n+1}). \end{aligned}$$

Tomando

$$y_n = \alpha_n(h'(x_n) - h'(x_{n+1})) \iff \frac{1}{\alpha_n} y_n = h'(x_n) - h'(x_{n+1}), \quad (2.17)$$

e

$$\phi(x^*) \geqslant \phi(x_{n+1}) + y_n(x^* - x_{n+1}) \iff y_n(x_{n+1} - x^*) \geqslant \phi(x_{n+1}) - \phi(x^*). \quad (2.18)$$

Substituindo (2.17) e (2.18) em (2.16),

$$D_h(x^*, x_n) - D_h(x^*, x_{n+1}) - D_h(x_{n+1}, x_n) \geqslant \frac{1}{\alpha_n}(\phi(x_{n+1}) - \phi(x^*)) \geqslant 0. \quad (2.19)$$

Considerando a sequência $\{D_h(x^*, x_n)\}$, de (2.19) tem-se

$$D_h(x^*, x_{n+1}) \leqslant D_h(x^*, x_n) - D_h(x_{n+1}, x_n) \leqslant D_h(x^*, x_n),$$

tem-se que a sequência $\{D_h(x^*, x_n)\}$ é monótona não-crescente e limitada inferiormente, logo é convergente. Fazendo $\delta = D_h(x^*, x_0)$, por B3) a sequencia $\{x_n\}$ é limitada, logo possui uma subsequência $\{x_j\}$ convergente. Por (2.19),

$$0 \leqslant D_h(x_{j+1}, x_j) \leqslant D_h(x^*, x_j) - D_h(x^*, x_{j+1}), \tag{2.20}$$

passando ao limite em (2.20), com $j \to \infty$

$$\lim D_h(x_{j+1}, x_j) = 0.$$

Se $\lim x_j = \hat{x}$, então por B5 $\lim x_{j+1} = \hat{x}$. Agora, fazendo $x_n = x_j$ e $x_{n+1} = x_{j+1}$ em (2.19),

$$\alpha_n[D_h(x^*, x_j) - D_h(x^*, x_{j+1}) - D_h(x_{j+1}, x_j)] \geqslant \phi(x_{j+1}) - \phi(x^*) \geqslant 0,$$

passando ao limite, com $j \to \infty$, tem-se

$$\lim \phi(x_{j+1}) - \phi(x^*) = 0 \iff \lim \phi(x_{j+1}) = \phi(x^*).$$

Pela continuidade de ϕ temos que $\phi(\hat{x}) = \phi(x^*)$, logo $\hat{x}$ é uma solução de (2.5) e é um minimizador global. Tem-se que $x_j \to \hat{x}$, por B4) $D_h(\hat{x}, x_j) \to 0$. Sendo $\hat{x}$ uma solução de (2.5), usando passos anteriores, $\{D_h(\hat{x}, x_n)\}$ é não-crescente e tem um subsequência convergindo para zero, então $\{D_h(\hat{x}, x_n)\}$ também converge zero. Tomando outra subsequência $\{x_i\}$ de $\{x_n\}$, tal que $x_i \to \tilde{x}$. Em particular, $\{D_h(\hat{x}, x_i)\}$ é uma subsequência de $\{D_h(\hat{x}, x_n)\}$, logo

$$\lim_{i\to\infty} D_h(\hat{x}, x_i) = 0,$$

por B5), $\hat{x} = \tilde{x}$. Portanto a sequência gerada pelo MPPDB converge para a solução de (2.5). □

Supondo a derivabilidade na função que se quer minimizar, a taxa de convergência com α_n positivos pode ser no máximo linear, como segue a proposição.

Proposição 2. [Taxa de convergência] *Seja* $I \subset \mathbb{R}$ *um intervalo fechado não-vazio e* $h\colon I \to \mathbb{R}$ *uma função estritamente convexa duas vezes continuamente derivável, tal que* $h''(x) > 0$ *para todo* $x \in I$. *Seja* $\phi\colon \mathbb{R} \to \mathbb{R}$ *uma função convexa duas vezes continuamente derivável. Considere a sequência* $\{x_n\}_{n\in\mathbb{N}}$ *definida por* (2.6), com $0 < \bar{\alpha} \leqslant \alpha_n \leqslant \tilde{\alpha}$ para todo n. Se $\{x_n\}_{n\in\mathbb{N}}$ converge para a solução x^*, então converge no máximo linearmente.

Demonstração. Uma condição necessária e suficiente para ponto de mínimo de uma função convexa é ser ponto crítico desta, ou seja,

$$x_{n+1} \text{ é solução de (2.6)} \iff (\phi + \alpha_n D_h(\cdot, x_n))' (x_{n+1}) = 0.$$

Logo, derivando $\phi(x) + \alpha_n D_h(x, x_n)$, temos

$$\begin{aligned} [\phi(x) + \alpha_n D_h(x, x_n)]' &= [\phi(x) + \alpha_n h(x) - \alpha_n h(x_n) - \alpha_n h'(x_n)(x - x_n)]' \\ &= \phi'(x) + \alpha_n h'(x) - \alpha_n h'(x_n). \end{aligned}$$

Agora aplicando x_{n+1}:

$$\phi'(x_{n+1}) + \alpha_n h'(x_{n+1}) - \alpha_n h'(x_n) = 0 \iff \frac{\phi'(x_{n+1})}{\alpha_n} + h'(x_{n+1}) - h'(x_n) = 0, \quad (2.21)$$

que pode ser resolvida implicitamente para x_{n+1}, para todo $n \geqslant 0$.

Por hipótese $\left(\dfrac{\phi''(x)}{\alpha_n} + h''(x)\right) > 0$ para todo $x \in \operatorname{int}(I)$. Pelo Teorema 11, $\left(\dfrac{\phi'(x)}{\alpha_n} + h'(x)\right)$ é invertível e derivável.

A função iteração $\eta_n(x) = \left[\left(\frac{\phi'}{\alpha_n} + h'\right)^{-1} \circ h'\right](x)$ associada com a iteração proximal (2.21) está bem definida. Isolando x_{n+1},

$$x_{n+1} = \eta_n(x_n) = \left[\left(\frac{\phi'}{\alpha_n} + h'\right)^{-1} \circ h'\right](x_n), \quad \forall n. \quad (2.22)$$

Seja $x^* = \lim_{n\to\infty} x_n$, sem perda de generalidade assumir-se-á que $\alpha_n \to \alpha > 0$. Seja

$$\bar{\eta}(x) = \left[\left(\frac{\phi'}{\alpha} + h'\right)^{-1} \circ h'\right](x). \quad (2.23)$$

Passando ao limite na equação (2.22) temos

$$x^* = \left[\left(\frac{\phi'}{\alpha} + h'\right)^{-1} \circ h'\right](x^*) = \bar{\eta}(x^*),$$

logo x^* é ponto fixo de $\bar{\eta}$. Além disso, $\bar{\eta}$ é derivável e com inversa derivável.

$$\begin{aligned} [\bar{\eta}(\bar{\eta}^{-1}(x))]' &= [x]' \\ \bar{\eta}'(\bar{\eta}^{-1}(x)) \cdot (\bar{\eta}^{-1})'(x) &= 1 \end{aligned}$$

substituindo x por x^*, temos $\bar{\eta}'(x^*) \cdot (\bar{\eta}^{-1})'(x^*) = 1 \Rightarrow \bar{\eta}'(x^*) \neq 0$.

Usando os argumentos da demosntração do Teorema 24, concluir-se que

$$\lim_{n\to\infty} \frac{|x_{n+1} - x^*|}{|x_n - x^*|} = \bar{\eta}(x^*) \neq 0,$$

se $\bar{\eta}(x^*) < 1$, então $\{x_n\}$ converge com taxa linear para x^*. $\square$

Capítulo 3

O Método de Newton Generalizado

No capítulo anterior, viu-se que o MPPDB gera uma sequência que converge no máximo linearmente. Para obter uma formulação que permita taxa de convergência quadrática, considera-se a função de iteração

$$\tilde{g}(x) = (h')^{-1}\left(h'(x) + \frac{1}{\alpha}\phi'(x)\right). \tag{3.1}$$

Supondo que $h''(x) > 0$ para todo $x \in \text{int}(I)$, pode ser notado que a função $\tilde{g}$ está bem definida para qualquer $\alpha \neq 0$, e para um α positivo $\tilde{g}$ é a inversa da função iteração proximal dada por (2.23).

Encontrar a solução do problema de minimização convexa (2.5) é equivalente a encontrar um ponto fixo de $\tilde{g}$. De fato, seja x^* um ponto fixo de $\tilde{g}$:

$$\begin{aligned}
\tilde{g}(x^*) &= (h')^{-1}\left(h'(x^*) + \frac{1}{\alpha}\phi'(x^*)\right) \\
x^* &= (h')^{-1}\left(h'(x^*) + \frac{1}{\alpha}\phi'(x^*)\right) \\
h'(x^*) &= h'(x^*) + \frac{1}{\alpha}\phi'(x^*),
\end{aligned}$$

logo $\phi'(x^*) = 0$. Pelo Corolário 5, x^* é solução do problema (2.5). Pelos Teoremas 24 25, a convergência quadrática só pode ser obtida quando $\tilde{g}'(x^*) = 0$. Supondo que ϕ, h são fortemente convexas, então $\tilde{g}$ é derivável, pois $(h')^{-1}$ e $h' + \dfrac{1}{\alpha}\phi'$ são deriváveis. Derivando ambos os lados de (3.1),

$$\begin{aligned}
[\tilde{g}(x)]' &= \left[(h')^{-1}\left(h'(x) + \frac{1}{\alpha}\phi'(x)\right)\right]' \\
\tilde{g}'(x) &= [(h')^{-1}]'\left(h'(x) + \frac{1}{\alpha}\phi'(x)\right) \cdot \left(h''(x) + \frac{1}{\alpha}\phi''(x)\right).
\end{aligned}$$

Tomando $x = x^*$, na equação acima, tem-se que

$$\begin{aligned}
\tilde{g}'(x^*) &= [(h')^{-1}]'\left(h'(x^*) + \frac{1}{\alpha}\phi'(x^*)\right) \cdot \left(h''(x^*) + \frac{1}{\alpha}\phi''(x^*)\right) \\
\tilde{g}'(x^*) &= [(h')^{-1}]'(h'(x^*)) \cdot \left(h''(x^*) + \frac{1}{\alpha}\phi''(x^*)\right).
\end{aligned}$$

Pelo Teorema 11, tem-se que $[(h')^{-1}]'(h'(x^*)) = \dfrac{1}{h''(x^*)}$. Agora substituindo $[(h')^{-1}]'(h'(x^*))$ por $\dfrac{1}{h''(x^*)}$ na expressão de $\tilde{g}'$, obtem-se a igualdade

$$\tilde{g}'(x^*) = \frac{1}{h''(x^*)}\left(h''(x^*) + \frac{1}{\alpha}\phi''(x^*)\right).$$

Da igualdade acima, pode ser observado que $\tilde{g}'(x^*) = 0$ se, e somente se, $\alpha = -\dfrac{\phi''(x^*)}{h''(x^*)}$. De fato, se $\tilde{g}'(x^*) = 0$,

$$\frac{1}{h''(x^*)}\left(h''(x^*) + \frac{1}{\alpha}\phi''(x^*)\right) = 0 \Rightarrow h''(x^*) + \frac{1}{\alpha}\phi''(x^*) = 0 \Rightarrow$$

$$\alpha = -\frac{\phi''(x^*)}{h''(x^*)}.$$

Por outro lado, se $\alpha = -\dfrac{\phi''(x^*)}{h''(x^*)}$,

$$\begin{aligned}
\tilde{g}'(x^*) &= \frac{1}{h''(x^*)}\left(h''(x^*) - \frac{\phi''(x^*)}{h''(x^*)}\phi''(x^*)\right) \\
\tilde{g}'(x^*) &= \frac{1}{h''(x^*)}\left(h''(x^*) - h''(x^*)\right) \\
\tilde{g}'(x^*) &= 0.
\end{aligned}$$

Assim, a escolha do α acima garante a convergência quadrática. No entanto, substituindo α, pela forma funcional, em (3.1), obtém-se o que será denominado de função de Newton generalizada:

$$g(x) = (h')^{-1}\left(h'(x) - h''(x)\frac{\phi'(x)}{\phi''(x)}\right). \tag{3.2}$$

O nome, dada a função g, é motivado quando se considera $h(x) = \dfrac{x^2}{2}$, pois $h'(x) = x$, $(h')^{-1}(x) = x$ e $h''(x) = 1$ e quando é substituído em (3.2), tem-se a função de Newton clássica:

$$g(x) = x - \frac{\phi'(x)}{\phi''(x)}.$$

A função g depende apenas das primeiras e segundas derivadas de ϕ e h. Por notação conveniente, sejam

$$f = \phi' \ \text{ e } \ s = h'$$

Reescreve-se a função de Newton generalizada g como

$$g(x) = s^{-1}\left(s(x) - s'(x)\frac{f(x)}{f'(x)}\right). \tag{3.3}$$

As hipóteses de convexidade das funções ϕ e h são colocadas para relacionar o método proposto com regularização proximal com distância de Bregman associada ao problema (2.5). De fato, a hipótese de convexidade forte para ϕ e h é usada para deduzir a fórmula (3.3), que pode ser substituída por condições mais fracas $\phi''(x^*) \neq 0$ e $h''(x^*) \neq 0$, ou seja, $f'(x^*) \neq 0$ e $s'(x^*) \neq 0$.

Lema 3. *Sejam $\{x_n\}_{n\in\mathbb{N}}$ uma sequência gerada por $x_{n+1} = g(x_n)$ e $y_n = s(x_n)$. Então a iteração*

$$x_{n+1} = s^{-1}\left(s(x_n) - s'(x_n)\frac{f(x_n)}{f'(x_n)}\right) \tag{3.4}$$

pode ser reescrita em termos da sequência y_n da seguinte forma

$$y_{n+1} = y_n - \frac{(f\circ s^{-1})(y_n)}{(f\circ s^{-1})'(y_n)}. \tag{3.5}$$

Demonstração. A iteração (3.4) pode ser reescrita como

$$s(x_{n+1}) = s(x_n) - s'(x_n)\frac{f(x_n)}{f'(x_n)}.$$

Sendo $y_n = s(x_n) \iff x_n = s^{-1}(y_n)$. Substituindo na igualdade acima, tem-se

$$\begin{aligned}
y_{n+1} &= y_n - s'(s^{-1}(y_n))\frac{f(s^{-1}(y_n))}{f'(s^{-1}(y_n))} \\
y_{n+1} &= y_n - \frac{f(s^{-1}(y_n))}{f'(s^{-1}(y_n))/s'(s^{-1}(y_n))} \\
y_{n+1} &= y_n - \frac{(f\circ s^{-1})(y_n)}{(f\circ s^{-1})'(y_n)}.
\end{aligned}$$

Pois $(f\circ s^{-1})'(y_n) = f'(s^{-1}(y_n)) \cdot (s^{-1})'(y_n)$ e $(s^{-1})'(y_n) = \dfrac{1}{s'(s^{-1}(y_n))}$, pelo Corolário 11, $(f\circ s^{-1})'(y_n) = \dfrac{f'(s^{-1}(y_n))}{s'(s^{-1}(y_n))}$. □

A iteração (3.5) nada mais é que a iteração clássica de Newton quando aplicada para encontrar a raiz da equação

$$F(y) = (f\circ s^{-1})(y) = 0.$$

Observação 4. *O método generalizado pode ser visto como o método clássico de Newton aplicado para encontrar a solução transformada y^*, tal que $(f\circ s^{-1})(y^*) = 0$, e que a solução do problema original é obtida por $x^* = s^{-1}(y^*)$.*

Estabelece-se uma taxa de convergência quadrática para o método generalizado utilizando o Teorema 26, e o lema a seguir dá condições sobre as quais $F := (f \circ s^{-1})'$ seja Lipschitziana.

Lema 4. *Sejam* $f,\ s \in C^1[a,b]$. *Supondo que para algum* $\theta > 0$, *tem-se* $\theta < |s'(x)|$ *para qualquer* $x \in (a,b)$. *Se* f' *e* s' *são Lipschitz em* (a,b), *então* $(f \circ s^{-1})'$ *está bem definida e é Lipschitz em* $I = s([a,b])$. *Além disso, se* L_f *e* L_s *são as constantes de Lipschitz de* f' *e* s', *respectivamente, então a constante de Lipschitz de* $(f \circ s^{-1})'$ *em* I *é*

$$M_{fs} = \frac{M_f L_s + M_s L_f}{\theta^3}, \tag{3.6}$$

com $M_f \geqslant \max\{|f'(x)|\colon x \in [a,b]\}$ *e* $M_s \geqslant \max\{|s'(x)|\colon x \in [a,b]\}$.

Demonstração. Sendo $f,\ s \in C^1[a,b]$, então f' e s' são contínuas em $[a,b]$, logo M_f e M_s são finitos. Das hipóteses sob s', garante-se que s é bijeção de $[a,b]$ em sua imaguem ($Im_{(s)} = s([a,b])$). De fato, sejam $x_1,\ x_2 \in [a,b]$ e s derivável em (x_1,x_2), por hipótese. Pelo Teorema do valor médio existe um $c \in (x_1,x_2)$, tal que

$$|s(x_1) - s(x_2)| = |s'(c)| \cdot |x_1 - x_2|.$$

Supondo $x_1 \neq x_2$ e sabendo que $0 < \theta < |s'(x)|\ \ \forall\, x \in (a,b)$, tem-se

$$|s(x_1) - s(x_2)| \neq 0 \iff s(x_1) \neq s(x_2),$$

então s é injetiva, e $s\colon [a,b] \to s([a,b])$ é bijeção. Pelo Teorema 11 s é invertível com $s^{-1} \in C^1$. Portanto, $f \circ s^{-1}$ é bem definida em I e derivável.

Passa-se então a demonstar que $(f \circ s^{-1})'$ é Lipschitz. Sejam $y_1,\ y_2 \in I$ e $J = f([a,b])$. Tomando a função $(f \circ s^{-1})\colon I \to J$, logo existem únicos $z_1,\ z_2 \in [a,b]$, tais que $s(z_1) = y_1$ e $s(z_2) = y_2$. Derivando $(f \circ s^{-1})$, tem-se que

$$(f \circ s^{-1})'(y) = f'(s^{-1}(y)) \cdot (s^{-1})'(y) \ \text{ e } \ (s^{-1})'(y) = \frac{1}{s'(s^{-1}(y))}.$$

Logo, para todo $y \in I$,

$$(f \circ s^{-1})'(y) = \frac{f'(s^{-1}(y))}{s'(s^{-1}(y))}.$$

$$
\begin{aligned}
|(f\circ s^{-1})'(y_1)-(f\circ s^{-1})'(y_2)| &= \left|\frac{f'(s^{-1}(y_1))}{f'(s^{-1}(y_1)}-\frac{f'(s^{-1}(y_2))}{f'(s^{-1}(y_2)}\right|\\
&= \left|\frac{f'(z_1)s'(z_2)-f'(z_2)s'(z_1)}{s'(z_1)s'(z_2)}\right|\\
&= \left|\frac{f'(z_1)s'(z_2)-f'(z_1)s'(z_1)+f'(z_1)s'(z_1)-f'(z_2)s'(z_1)}{s'(z_1)s'(z_2)}\right|\\
&= \left|\frac{f'(z_1)(s'(z_2)-s'(z_1))+s'(z_1)(f'(z_1)-f'(z_2))}{s'(z_1)s'(z_2)}\right|\\
&\leqslant \frac{|f'(z_1)|\cdot|s'(z_2)-s'(z_1)|+|s'(z_1)|\cdot|f'(z_1)-f'(z_2)|}{|s'(z_1)||s'(z_2)|}.
\end{aligned}
$$

Sabendo que $0<\theta<|s'(x)| \iff \dfrac{1}{|s'(x)|}<\dfrac{1}{\theta}$ para todo $x\in(a,b)$, f' e s' são lipschitziana, ou seja, $|f'(z_2)-f'(z_1)|\leqslant L_f|z_2-z_1|$ e $|s'(z_2)-s'(z_1)|\leqslant L_s|z_2-z_1|$, e que $M_f\geqslant \max\{|f'(x)|\colon x\in[a,b]\}$ e $M_s\geqslant \max\{|s'(x)|\colon x\in[a,b]\}$, tem-se

$$
\begin{aligned}
|(f\circ s^{-1})'(y_1)-(f\circ s^{-1})'(y_2)| &\leqslant \frac{1}{\theta^2}[|f'(z_1)|\cdot|s'(z_2)-s'(z_1)|+|s'(z_1)|\cdot|f'(z_1)-f'(z_2)|]\\
&\leqslant \frac{1}{\theta^2}[M_fL_s|z_2-z_1|+M_sL_f|z_2-z_1|]\\
&= \frac{M_fL_s+M_sL_f}{\theta^2}|z_2-z_1|.
\end{aligned}
$$

Usando o Teorema do valor médio em s, obtém-se

$$
\begin{aligned}
|y_1-y_2|=|s(z_1)-s(z_2)| &= |s'(\psi)|\cdot|z_1-z_2|\geqslant\theta|z_1-z_2|\\
|z_1-z_2| &\leqslant \frac{|y_1-y_2|}{\theta},
\end{aligned}
$$

para algum $\psi\in(a,b)$. Substituindo:

$$
|(f\circ s^{-1})'(y_1)-(f\circ s^{-1})'(y_2)|\leqslant\frac{M_fL_s+M_sL_f}{\theta^3}|y_1-y_2|=M_{fs}|y_1-y_2|.
$$

□

O próximo resultado garante que a sequência $\{x_n\}_{n\in\mathbb{N}}$ gerada por $x_{n+1}=g(x_n)$ converge para um zero da função f, pelo menos de forma quadrática.

Teorema 28. *Seja* $f,\ s\in C^1[a,b]$*, tais que as hipóteses do Lema 4 são satisfeitas. Existe um* $x^*\in(a,b)$*, tal que* $f(x^*)=0$ *e para algum* $\rho>0$, $|f'(x)|\geqslant\rho$ *para todo* $x\in(a,b)$*. Então, existe um* $\eta>0$*, tal que se* $x_0\in(x^*-\eta,x^*+\eta)$*, a sequência* $\{x_n\}_{n\in\mathbb{N}}$ *definida por*

$$
x_{n+1}=g(x_n),\ n=0,1,2,\cdots,
$$

onde g é dada em (3.3), é bem definida. Além disso, a sequência $\{x_n\}_{n\in\mathbb{N}}$ converge quadraticamente para x^*, e vale

$$|x_{n+1} - x^*| \leqslant \frac{M_{fs}(M_s)^3}{2\rho\theta}|x_n - x^*|^2, \tag{3.7}$$

onde θ, M_{fs} e M_s são como no Lema 4.

Demonstração. Pelo Lema 4, $(f \circ s^{-1})'$ é bem definida e lipschitziana em (a, b) com constanTte de Lipschitz M_{fs}. Sendo $|f'(x)| \geqslant \rho$ e $0 \neq |s'(x)| \leqslant M_s \Rightarrow \dfrac{1}{|s'(x)|} \geqslant \dfrac{1}{M_s}$, $\forall\, x \in (a, b)$, tem-se que, com $y = s(x) \iff x = s^{-1}(y)$,

$$|(f \circ s^{-1})'(y)| = |f'(s^{-1}(y)) \cdot (s^{-1})'(y)| = \left|\frac{f'(s^{-1}(y))}{s'(s^{-1}(y))}\right| = \left|\frac{f'(x)}{s'(x)}\right| \geqslant \frac{\rho}{M_s},$$

para todo $y \in s^{-1}(a, b)$. Definindo $F(y) = (f \circ s^{-1})(y)$, tem-se que
$0 = f(x^*) = f(s^{-1}(y^*)) = (f \circ s^{-1})(y^*) = F(y^*) = 0$, pelo Teorema 26, existe um $\gamma > 0$, tais que, se $y_0 \in (y_* - \gamma, y^* + \gamma)$ a sequência definida por (3.5) é bem definida e converge quadraticamente para y^* e vale

$$|y_{n+1} - y^*| \leqslant \frac{M_{fs}M_s}{2\rho}|y_n - y^*|^2 \quad n = 0, 1, 2, \cdots .$$

Pelo Lema 3, a sequência $\{x_n\}_{n\in\mathbb{N}}$, onde $x_n = s^{-1}(y_n)$ é gerada por $x_{n+1} = g(x_n)$. Pela continuidade de s^{-1}, o $\lim x_n = \lim s^{-1}(y_n) = s^{-1}(y^*) = x^*$. Como $y_n = s(x_n)$, a desigualdade acima poderá ser reescrita em termos de x_n, assim,

$$|s(x_{n+1}) - s(x^*)| \leqslant \frac{M_{fs}M_s}{2\rho}|s(x_n) - s(x^*)|^2, \quad n = 0, 1, 2, \cdots . \tag{3.8}$$

Pelo Teorema do valor médio, existem $\psi_1,\ \psi_2 \in (a, b)$ tais que

$$|s(x_{n+1}) - s(x^*)| = |s'(\psi_1)| \cdot |x_{n+1} - x^*| \geqslant \theta|x_{n+1} - x^*| \tag{3.9}$$

e

$$|s(x_n) - s(x^*)| = |s'(\psi_2)| \cdot |x_n - x^*| \leqslant M_s|x_{n+1} - x^*|. \tag{3.10}$$

Usando as desigualdades (3.9) e (3.10) em (3.8), tem-se

$$\begin{aligned}
\theta \cdot |x_{n+1} - x^*| &\leqslant |s(x_{n+1}) - s(x^*)| \leqslant \frac{M_{fs}M_s}{2\rho}M_s^2|x_n - x^*|^2 \\
|x_{n+1} - x^*| &\leqslant \frac{M_{fs(M_s)^3}}{2\rho\theta}|x_n - x^*|^2.
\end{aligned}$$

□

3.1 Instâncias da função de Newton generalizada

Várias funções h são escolhidas e as funções de Newton generalizada g, dada por (3.3), são listadas na Tabela 3.1, igualmente com [3].

Na Tabela 3.1, foi escolhida h, de tal forma que sua inversa seja calculada rapidamente. Foram usadas funções fortemente convexas para motivar a função de Newton generalizada, mas a seguinte proposição mostra que h não precisa ser uma função convexa.

Proposição 3. *Ambas as funções* h *e* $-h$ *proporcionam uma função de Newton generalizada.*

Demonstração. Supõe-se que g é obtida usando $-h$. Então $s(x) = -h'(x)$, e

$$\begin{aligned} h'(x) &= -s(x) \\ h'(s^{-1}(y)) &= -s(s^{-1}(y)) = -y \\ (h')^{-1}(h'(s^{-1}(y))) &= (h')^{-1}(-y) \\ s^{-1}(y) &= (h')^{-1}(-y). \end{aligned}$$

Logo, $s^{-1}(x) = (h')^{-1}(-x)$ e $s'(x) = -h''(x)$. Substituindo em (3.3), tem-se

$$g(x) = (h')^{-1}\left(-\left(-h'(x) - (-h''(x))\frac{f(x)}{f'(x)}\right)\right) = (h')^{-1}\left(h'(x) - h''(x)\frac{\phi'(x)}{\phi''(x)}\right),$$

que é a mesma expressão em (3.2), ou seja, a expressão de $g(x)$ usando $h(x)$. □

Na Tabela 3.1, também foi listado o domínio das funções h. Observou-se que a função $g(x) = x + \log\left(1 - \dfrac{f(x)}{f'(x)}\right)$, na segunda linha da tabela, tem domínio diferente da função $h(x) = e^x$, onde domínio de g depende da função f: g é bem definida quando $1 - \dfrac{f(x)}{f'(x)} > 0$, ou seja, quando $f(x) < f'(x)$ e $f'(x) > 0$. A função $h(x) = \log x$ não é uma função convexa, mas $-h$ é uma função estritamente convexa no domínio de h.

$h(x)$	$\operatorname{dom} h$	$s(x)$	$s^{-1}(x)$	$s'(x)$	$g(x)$
$\frac{x^2}{2}$	$\mathbb{R}$	x	x	1	$x - \frac{f(x)}{f'(x)}$
$x \log x - x$	$\mathbb{R}_+$	$\log x$	e^x	$\frac{1}{x}$	$x e^{-f(x)/(x f'(x))}$
e^x	$\mathbb{R}$	e^x	$\log x$	e^x	$x + \log\left(1 - \frac{f(x)}{f'(x)}\right)$
$\log x$	$\mathbb{R}_{++}$	$\frac{1}{x}$	$\frac{1}{x}$	$-\frac{1}{x^2}$	$x\left(1 + \frac{f(x)}{x f'(x)}\right)^{-1}$
$\cosh x$	$\mathbb{R}$	$\sinh x$	$\operatorname{arcsinh} x$	$\cosh x$	$\operatorname{arcsinh}\left(\sinh x - \cosh x \frac{f(x)}{f'(x)}\right)$
$\log(\cos x)$	$\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$	$\tan x$	$\arctan x$	$\sec^2 x$	$\arctan\left(\tan x - \sec^2 x \frac{f(x)}{f'(x)}\right)$
$x \arctan x - \frac{\log(x^2+1)}{2}$	$\mathbb{R}$	$\arctan x$	$\tan x$	$\frac{1}{x^2+1}$	$\tan\left(\arctan x - \frac{f(x)}{(1+x^2) f'(x)}\right)$

Tabela 3.1: Funções de Newton generalizadas para cada função h escolhida.

3.2 Erro assintótico constante

Sendo f e s duas vezes continuamente diferenciáveis, pode ser fornecida uma fórmula para o erro assintótico λ, ver [3].

Proposição 4. *Sejam* $f,\ s \in C^3[a,b]$. *Então*

$$\lambda = \frac{1}{2}\left|\frac{f''(x^*)}{f'(x^*)} - \frac{s''(x^*)}{s'(x^*)}\right|. \tag{3.11}$$

Demonstração. Por hipótese, $f',\ s' \in C^2[a,b]$. Assim,

$$g(x) = s^{-1}\left(s(x) - s'(x)\frac{f(x)}{f'(x)}\right)$$

é também de classe C^2 em $[a,b]$, pois s^{-1} e $s(x) - s'(x)\frac{f(x)}{f'(x)}$ são de classe C^2 em $[a,b]$. Sendo x^* um zero da função f, tais que $f'(x^*) \neq 0$ e $s'(x^*) \neq 0$, tem-se que

$$g(x^*) = s^{-1}\left(s(x^*) - s'(x^*)\frac{f(x^*)}{f'(x^*)}\right) = s^{-1}(s(x^*)) = x^*.$$

Tomando

$$s(g(x)) = s(x) - s'(x)\frac{f(x)}{f'(x)},$$

e derivando ambos os lados da equação acima, tem-se

$$\begin{aligned}
[s(g(x))]' &= \left[s(x) - s'(x)\frac{f(x)}{f'(x)}\right]' \\
s'(g(x))g'(x) &= s'(x) - s''(x)\frac{f(x)}{f'(x)} - s'(x)\frac{f'(x)f'(x) - f(x)f''(x)}{(f'(x))^2} \\
s'(g(x))g'(x) &= s'(x) - s''(x)\frac{f(x)}{f'(x)} - s'(x)\left(1 - \frac{f(x)f''(x)}{(f'(x))^2}\right) \\
s'(g(x))g'(x) &= s'(x) - s''(x)\frac{f(x)}{f'(x)} - s'(x) + s'(x)\frac{f(x)f''(x)}{(f'(x))^2} \\
s'(g(x))g'(x) &= -s''(x)\frac{f(x)}{f'(x)} + s'(x)\frac{f(x)f''(x)}{(f'(x))^2}.
\end{aligned}$$

Fazendo $x = x^*$, tem-se

$$s'(g(x^*))g'(x^*) = -s''(x^*)\frac{f(x^*)}{f'(x^*)} + s'(x^*)\frac{f(x^*)f''(x^*)}{(f'(x^*))^2} = 0.$$

Como $g(x^*) = x^*$ e $s'(x^*) \neq 0$, conclui-se que $g'(x^*) = 0$. Pela demonstração do Teorema 25, o erro assintótico constante é:

$$\lambda = \frac{1}{2}|g''(x^*)|.$$

Mais uma vez, será efetuada a derivação a fim de que seja encontada a expressão para $g''(x^*)$.

$$\begin{aligned}[s'(g(x))g'(x)]' &= \left[-s''(x)\frac{f(x)}{f'(x)} + s'(x)\frac{f(x)f''(x)}{(f'(x))^2}\right]' \\ s''(g(x))g'(x)g'(x) + s'(g(x))g''(x) &= -\left[s'''(x)\frac{f(x)}{f'(x)} + s''(x)\frac{f'(x)f'(x) - f(x)f''(x)}{(f'(x))^2}\right]\end{aligned}$$

$$+\frac{[s''(x)f(x)f''(x) + s'(x)f'(x)f''(x) + s'(x)f(x)f'''(x)]f'(x)^2 - 2s'(x)f(x)f''(x)f'(x).f''(x)}{f'(x)^4}$$

Tomando $x = x^*$, $f(x^*) = 0$, $s'(x^*) \neq 0$, $g(x^*) = x^*$ e $g'(x^*) = 0$ na equação acima, tem-se

$$s'(x^*)g''(x^*) = -s''(x^*)\frac{f'(x^*)f'(x^*)}{(f'(x^*))^2} + \frac{[s'(x^*)f'(x^*)f''(x^*)]f'(x)^2}{f'(x^*)^4} = -s''(x^*) + \frac{s'(x^*)f''(x^*)}{f'(x^*)}.$$

Dividindo por $s'(x^*)$, obtém-se

$$g''(x^*) = \frac{f''(x^*)}{f'(x^*)} - \frac{s''(x^*)}{s'(x^*)}.$$

Portanto, o erro assintótico é dado por (3.11). □

Nota-se que o erro assintótico constante para a iteração (3.5) é

$$\tilde{\lambda} = \frac{1}{2}\left|\frac{1}{s'(x^*)}\left(\frac{f''(x^*)}{f'(x^*)} - \frac{s''(x^*)}{s'(x^*)}\right)\right|, \tag{3.12}$$

diferentemente de λ em (3.11). De fato, seja

$$G(y) = y - \frac{F(y)}{F'(y)},$$

com $F(y) := (f \circ s^{-1})(y)$ e y^* uma raíz de F. Pela demonstração do Teorema 25, tem-se que

$$\tilde{\lambda} = \frac{1}{2}|G''(y^*)|.$$

Ao derivar G, duas vezes e aplicando $y^* = s^{-1}(x^*)$, tem-se

$$\begin{aligned} G'(y) &= 1 - \frac{[F'(y)]^2 - F(y)F''(y)}{[F'(y)]^2} = \frac{F(y)F''(y)}{[F'(y)]^2} \\ G''(y) &= \frac{[F'(y)F''(y) + F(y)F'''(y)]F'(y)^2 - 2F'(y)F''(y)F(y)F''(y)}{[F'(y)]^4} \\ G''(y^*) &= \frac{[F'(y^*)F''(y^*)]F'(y^*)^2}{[F'(y^*)]^4} = \frac{F''(y^*)}{F'(y^*)}.\end{aligned}$$

A etapa sequinte será derivar duas vezes a função $F(y)$:

$$\begin{aligned}
F'(y) &= f'(s^{-1}(y))(s^{-1})'(y) = \frac{f'(s^{-1}(y))}{s'((s^{-1}(y))} \\
F''(y) &= \frac{\frac{f''(s^{-1}(y))}{s'((s^{-1}(y))}s'((s^{-1}(y)) - f'(s^{-1}(y))\frac{s''(s^{-1}(y))}{s'((s^{-1}(y))}}{[s'((s^{-1}(y))]^2} \\
F''(y) &= \frac{f''(s^{-1}(y))}{[s'((s^{-1}(y))]^2} - f'(s^{-1}(y))\frac{s''(s^{-1}(y))}{[s'((s^{-1}(y))]^3}.
\end{aligned}$$

Assim, aplicando $x^* = s^{-1}(y^*)$, em F' e F'', tem-se

$$\begin{aligned}
F'(y^*) &= \frac{f'(s^{-1}(y^*))}{s'((s^{-1}(y^*))} = \frac{f'(x^*)}{s'(x^*)} \\
F''(y^*) &= \frac{f''(s^{-1}(y^*))}{[s'((s^{-1}(y^*))]^2} - f'(s^{-1}(y^*))\frac{s''(s^{-1}(y^*))}{[s'((s^{-1}(y^*))]^3} = \frac{f''(x^*)}{[s'(x^*)]^2} - f'(x^*)\frac{s''(x^*)}{[s'(x^*)]^3},
\end{aligned}$$

Logo

$$G''(y^*) = \frac{\frac{f''(x^*)}{[s'(x^*)]^2} - f'(x^*)\frac{s''(x^*)}{[s'(x^*)]^3}}{\frac{f'(x^*)}{s'(x^*)}} = \frac{1}{s'(x^*)}\left(\frac{f''(x^*)}{f'(x^*)} - \frac{s''(x^*)}{s'(x^*)}\right).$$

Portanto, o erro assintótico da sequência dada por (3.5) é (3.12).

3.3 Experimentos numéricos

Nesta seção, será ilustrado o comportamento da iteração gerada pela fórmula de Newton generalizada atrevés de três exemplos, ver [3], para a função ϕ, e usando as funções s escolhida como na Tabela 3.1. Também será tabelado o erro assintótico constante λ, obtido através de itereções numéricas para cada função s. O Algoritmo, o método de Newton generalizado, foi codificado em SCILAB.

Ao se utilizar $|x_n - x^*| < 10^{-8}$ ou $n > 10$ como condição de parada, as Tabelas 3.2 - 3.4, apresentam as funções $s(x)$ escolhidas, os intervalos $[a, b]$ - tal que a solução está no intervalo, no qual se vê que o método de Newton generalizado converge com taxa quadrática para a solução x^* com diferentes funções $s(x)$ -, o número de iterações n, o ponto inicial x_0, escolhido de forma randômica, o ponto x_n e o valor da função f aplicado no ponto x_n, ou seja, $f(x_n)$.

Exemplo 8. *Considerando a função* $\phi(x) = (x-1)e^x - x$, *então* $f(x) = \phi'(x) = xe^x - 1$ *e* $f'(x) = \phi''(x) = (x+1)e^x$. *Nota-se que* $\phi(x)$ *não é uma função convexa em* $\mathbb{R}$*; de fato,* $x = -1$ *é um ponto de inflexão, ou seja,* $\phi''(-1) = 0$*, é onde a concavidade da função muda. Para* $x > -1$*,* ϕ *é estritamente convexa, logo é convexa, e para* $x < -1$*,*

ϕ *é côncava. Se* x *é o ponto ótimo, então* $xe^x = 1$ *tem solução única, que é* $x^* =$ 0.56714329040 *(correto em* 11 *casa decimais).*

$s(x)$	$[a,b]$	n	x_0	x_n	$f(x_n)$
x	$[-0.39, 2.71]$	7	-0.0876248	0.56711433	0
$\log x$	$[0.51, 0.64]$	4	0.5462048	0.56711433	0
e^x	$[0.25, 0.92]$	4	0.6164106	0.56711433	$2,220D-16$
$\frac{1}{x}$	$[0.35, 2.27]$	9	2.0702383	0.56711433	0
$\sinh x$	$[0.39, 0.89]$	5	0.8752964	0.56711433	$2,220D-16$
$\tan x$	$[-0.31, 1.22]$	7	1.1544654	0.56711433	$2,220D-16$
$\arctan x$	$[-0.25, 1.15]$	6	1.1235536	0.56711433	0

Tabela 3.2: Experimento numérico para $f(x) = xe^x - 1 = 0$.

Exemplo 9. *Considerando a função* $\phi(x) = 8(x\log x - x) - \frac{x^2}{2}$*, então* $f(x) = \phi'(x) = 8\log x - x$ *e* $f'(x) = \phi''(x) = \frac{8}{x} - 1$*. Nota-se que* $\phi(x)$ *não é uma função convexa em* $\mathbb{R}_{++}$*; de fato,* $x = 8$ *é um ponto de inflexão, ou seja,* $\phi''(8) = 0$*, e onde ocorre a mudança de concavidade função. Para* $x \in (0, 8)$*,* ϕ *é estritamente convexa, logo é convexa, e para* $x > 8$*,* ϕ *é côncava. O problema* $f(x) = 0$ *tem duas soluções,* $x^* = 1.15537082510$*, ponto de mínimo, e* $\hat{x} = 26.09348547661$*, ponto de máximo, ambos com* 11 *casas decimais corretas.*

$s(x)$	$[a,b]$	n	x_0	x_n	$f(x_n)$
x	$[0.42, 5.95]$	6	2.0764868	1.1553708	0
$\log x$	$[0.05, 7.98]$	6	7.3117827	1.1553708	0
e^x	$[0.17, 2.05]$	4	1.0923074	1.1553708	$1.332D-15$
$\frac{1}{x}$	$[0.35, 2.72]$	5	1.9227451	1.1553708	0
$\sinh x$	$[0.39, 1.98]$	8	1.7840443	1.1553708	0
$\tan x$	$[0.01, 1.32]$	5	1.241497	1.1553708	0
$\arctan x$	$[0.25, 5.15]$	4	4.7947149	1.1553708	0

Tabela 3.3: Experimento numérico para $f(x) = 8\log x - x = 0$.

Exemplo 10. *Considerando a função* $\phi(x) = \frac{x^4}{4} + \frac{x^2}{2} - 3x + 1$, *então* $f(x) = \phi'(x) = x^3 + x - 3$ *e* $f'(x) = \phi''(x) = 3x^2 + 1$. *A função* ϕ *é estritamente convexa em* $\mathbb{R}$, *pois* $\phi''(x) > 0 \; \forall \, x \in \mathbb{R}$ *e o problema tem uma única solução,* $x^* = 1.21341166276$*(correto em* 11 *casa decimais).*

$s(x)$	$[a, b]$	n	x_0	x_n	$f(x_n)$
x	$[0.00, 5.00]$	8	4.0736185	1.2134117	$-4.44D-16$
$\log x$	$[0.56, 1.98]$	7	0.8068885	1.2134117	$8.882D-16$
e^x	$[-0.62, 2.14]$	5	-0.274956	1.2134117	$-4.441-16$
$\frac{1}{x}$	$[0.76, 2.72]$	6	0.9932355	1.2134117	8.882D-16
$\sinh x$	$[-1.35, 3.22]$	5	-0.667764	1.2134117	$-4.441-16$
$\tan x$	$[-0.31, 3.22]$	7	3.061212	1.2134117	$-4.441-16$
$\arctan x$	$[0.65, 1.95]$	8	0.6654727	1.2134117	$-1.776D-15$

Tabela 3.4: Experimento numérico para $f(x) = x^3 + x - 3 = 0$.

Na Tabela a seguir, apresentam-se as funções instâncias s, o erro assintótico constante λ e a variação do erro assintótico constante $|\lambda_n - \lambda|$. O valor de λ_n é obtido com a substituição de x^* por x_n em (3.11).

	$f(x) = xe^x - 1$		$f(x) = 8\log x - x$		$f(x) = x^3 + x - 3$	
$s(x)$	λ	$\lvert\lambda_n - \lambda\rvert$	λ	$\lvert\lambda_n - \lambda\rvert$	λ	$\lvert\lambda_n - \lambda\rvert$
x	0.8190519	$1.992D-12$	0.5058115	$2.820D-14$	0.6719892	$7.788D-13$
$\log x$	1.7006633	$1.720D-11$	0.0730500	$8.882D-16$	1.0840505	$1.536D-12$
e^x	0.3190519	$1.992D-12$	1.0058115	$2.798D-14$	0.1719892	$7.790D-13$
$\frac{1}{x}$	2.5822747	$3.241D-11$	0.3597115	$3.009D-14$	1.4961118	$2.293D-12$
$\sinh x$	0.5624228	$5.595D-12$	0.9155742	$1.554D-14$	0.2531392	$1.111D-12$
$\tan x$	0.1821063	$1.575D-11$	2.7728869	$4.490D-13$	2.0059613	$1.900D-11$
$\arctan x$	1.248169	$1.808D-12$	0.0109809	$2.365D-14$	1.1627785	$9.512D-13$

Tabela 3.5: Experimento numérico diferentes valores para λ.

Capítulo 4

Considerações finais

Neste trabalho, foi analisada uma generalização do método de Newton com uma variável, motivada pelo método do ponto proximal com distância de Bregman, para resolução de equações não-lineares e problemas de otimização. Mostrou-se a convergência quadrática do método generalizado, onde um caso especial, $s(x) = x$, é o método de Newton clássico. Foram ilustradas as vantagens do método generalizado com relação ao método clássico, através de testes numéricos. Os testes forneceram uma visão de como as instâncias s do método generalizado podem ser escolhidas para uma dada equação não-linear. Por último, derivamos uma expressão fechada para o erro assintótico. Nos experimentos numéricos, foi visto o comportamento do método generalizado e observado que, para algumas funções s escolhidas, o erro assintótico λ é bem mais próximo de zero.

Referências Bibliográficas

[1] Bregman, L. M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. Zh. Vychisl. Mat. Mat. Fiz, v. 7, n. 3, p. 620-631, 1967.

[2] Burachik, R. S., Iusem, A. N., A generalized proximal point algorithm for the variational inequality problem in a Hilbert space. SIAM J. Optim. 8, 197-216 (1998).

[3] Burachik, R. S., Kaya, C. Y. and Sabach, S.: Generalized univariate Newton methods motivated by proximal regularization, Journal of Optimization Theory and Applications 155 (2012), 923–940.

[4] Burden, R. L., Faires, J. D.: Numerical Analysis, 9th edition. Thompson Brooks/Cole, Belmont, CA, USA (2011).

[5] Chen, G., Teboulle, M. Convergence analysis of a proximal-like minimization algorithm using Bregman functions, SIAM Journal on Control and Optimization, v.3, n. 3, p. 538-543, 1993.

[6] Dennis, J. E., Schnabel, R. B. Numerical methods for unconstrained optimization and nonlinear equations, New York, Prentice-Hall, 1983.

[7] Iusem, A, N,. Métodos de Ponto Proximal em Otimização, 20º Colóquio Brasileiro de Matemática, IMPA, Rio de Janeiro, 1995.

[8] Izmailov, A., Solodov, M. Otimização, Volume 1, Rio de Janeiro: IMPA, 2009.

[9] Izmailov, A., Solodov, M. Otimização, volume II, Rio de Janeiro, IMPA, 2007.

[10] Lima, Elon L.: Análise real, Volume 1, Funções de uma variável. 11a. ed. Rio de Janeiro: IMPA,2012.

[11] Martinet, B. Régularisation, d'inequations variationelles por approximations successives. (French) Rev. Française Imformat. Recherche Opérationnelle 4, Ser. R-3, 154-158, (1970).

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

[13] Polyak, B. T.: Newton's method and its use in optimization. European J. Operational Research 181, 1086-1096 (2007).

[14] Rockafellar, R. T.: Convex Analysis. Princeton Univ. Press (1970) .

[15] Rockafellar, R, T. Monotone Operator and the Proximal Point Algorithm. SIAM SIAM J. Optim. 14, 877-898 (1976).