UNIVERSIDADE FEDERAL DO TOCANTINS
CAMPUS UNIVERSITÁRIO DE PALMAS
CURSO DE CIÊNCIA DA COMPUTAÇÃO
TRABALHO DE CONCLUSÃO DE CURSO II

O ALGORITMO DE PONTO PROXIMAL PARA OTIMIZAÇÃO DE FUNÇÕES CONVEXAS DE UMA VARIÁVEL REAL A VALORES REAIS

Herinson Barbosa Rodrigues

Orientador: Dr. Rogério Azevedo Rocha

Palmas
Dezembro de 2016

O ALGORITMO DE PONTO PROXIMAL PARA OTIMIZAÇÃO DE FUNÇÕES CONVEXAS DE UMA VARIÁVEL REAL A VALORES REAIS

Herinson Barbosa Rodrigues

Trabalho de Conclusão de Curso II apresentado ao Curso de Ciência da Computação, CUP, da Universidade Federal do Tocantins, como parte dos requisitos necessários à obtenção do título de Bacharel em Ciência da Computação.

Comissão Julgadora:

Prof. Dr. Rogério Azevedo Rocha

Prof. Dr. Nome do Segundo Examinador

Prof. Dr. Nome do Terceiro Examinador

PALMAS, TO – BRASIL
DEZEMBRO DE 2016

"Err and err and err again,
but less and less and less"
- Piet Hein

Resumo

Neste trabalho, apresentamos o algoritmo de ponto proximal para otimização de funções convexas de uma variável real a valores reais. Mais especificamente, propomos ilustrações algébricas e geométricas do procedimento iterativo do algoritmo, com o rigor matemático através da análise de convergência do algoritmo. Por fim, testamos o algoritmo através de experimentos numéricos utilizando GNU Octave.

Palavras-chave: funções reais de variável real, algoritmo de ponto proximal, otimização, convexidade.

Abstract

In this work, we present the proximal point algorithm for optimization of convex functions of a real variable to real values. Particularly, we propose algebraic and geometric illustrations of the algorithm's iterative procedure, with mathematical accuracy through the algorithm's convergence analysis. Lastly, we test the algorithm with numerical experiments using GNU Octave.

Keywords: real function of a real variable, proximal point algorithm, optimization, convexity.

Lista de Tabelas

Lista de Figuras

Sumário

1 Introdução

O Algoritmo de Ponto Proximal (APP) ou Método de Ponto Proximal teve sua origem no trabalho *Brève communication. Régularisation d'inéquations variationnelles par approximations successives* [**?**] de Martinet e posteriormente foi desenvolvido e estudado por Rockafellar no trabalho *Monotone Operators and the Proximal Algorithm* [**?**]. Como divulgação deste APP, destaca-se o trabalho *Métodos de Ponto Proximal em Otimização* de Iusem [**?**] que foi apresentado no 20º Colóquio Brasileiro de Matemática[1].

O interesse especial para se utilizar um algoritmo proximal é encontrar um valor ótimo global para uma determinada função. O campo da matemática aplicada responsável por lidar com problemas desse tipo é a otimização global.

Otimização global é o processo de encontrar o mínimo de uma função de n parâmetros, possivelmente com os valores dos parâmetros sujeitos à certas restrições. Quando não há restrições, podemos definir o problema como

$$\underset{x}{\text{minimize}} \quad f(x) \tag{1.1}$$

onde f é uma função objetivo e x é um vetor que representa os n parâmetros. Se f é uma função de $\mathbb{R}^n$ em $\mathbb{R}$, de forma que os elementos x_i do vetor de entrada x e o valor de saída são números reais, o problema de otimização global é contínuo [**?**].

Na computação científica, otimização é uma das tarefas mais recorrentes encontradas. Qualquer problema de otimização é essencialmente constituído por uma função objetivo e, dependendo da natureza da função objetivo, há uma necessidade de maximizá-la ou minimizá-la. Há várias aplicações reais onde a otimização global pode ser utilizada. Por exemplo, quando deseja-se maximizar o lucro de um produto, maximizar a confiabilidade de um equipamento, minimizar riscos ou minimizar o peso de uma estrutura de metal.

Quando são impostas restrições, temos um *problema de otimização restrito*. Quando não há restrições, temos um *problema de otimização irrestrito*. Se a função objetivo e suas restrições são lineares, tais problemas podem ser facilmente resolvidos com métodos de programação linear utilizando, por exemplo, o algoritmo simplex [**?**]. Por outro lado, métodos de programação não linear são mais adequados quando a função objetivo é não linear com restrições lineares, não linear com restrições não lineares ou não linear e irrestrita.

No cenário industrial e empresarial existem inúmeras aplicações de programação não linear, tais como aquisição, fabricação, análise de equilíbrio e logística. Algumas inovações recentes dependem quase inteiramente da programação não linear, como redes neurais e

[1] O Colóquio Brasileiro de Matemática é uma ampla reunião científica organizada anualmente pelo IMPA - Instituto de Matemática Pura e Aplicada

sistemas adaptativos. Dentre os algoritmos existentes na programação não linear, destaca-se o clássico Algoritmo de Ponto Proximal.

Um APP é um algoritmo para resolver um problema de otimização convexa que utiliza operadores proximais dos termos objetivos.

Há várias razões para o estudo dos algoritmos proximais. Primeiramente, eles funcionam sob extremas condições gerais. Também são rápidos já que podem ser simplesmente operadores proximais para funções mais complexas. Uma outra razão é que eles são apropriados para otimização distribuída. E por fim, são conceitualmente e matematicamente simples, sendo assim fáceis de entender e implementar para um determinado problema.

Como já mencionado, este algoritmo consiste no problema clássico de minimização irrestrita onde o objetivo é minimizar uma função convexa f através de aproximações aplicadas repetidamente a partir de um ponto inicial x^0. Considere o problema:

$$\text{minimize}\{f(x) \mid x \in \mathbb{R}^n\} \tag{1.2}$$

onde a função objetivo $f : \mathbb{R}^{\mathrm{n}} \to \mathbb{R}$ é convexa e continuamente diferenciável [**?**]. O APP gera uma sequência $\{x^k\} \subset \mathbb{R}^{\mathrm{n}}$ através do seguinte procedimento iterativo:

1. Escolha $x^0 \in \mathbb{R}^{\mathrm{n}}$.

2. Dado x^k, se $x^k \in \arg\min\{f(x) \mid x \in \mathbb{R}^{\mathrm{n}}\}$, então $x^{k+p} = x^k, \forall p \geq 1$.

3. Dado x^k, se $x^k \notin \arg\min\{f(x) \mid x \in \mathbb{R}^{\mathrm{n}}\}$, então tomo como próxima iterada (x^{k+1}) o único mimimizador, em $\mathbb{R}^{\mathrm{n}}$, da função $f(x) + \lambda_k||x - x^k||^2$, isto é,

$$x^{k+1} = \arg\min\{f(x) + \lambda_k\|x - x^k\|^2 \mid x \in \mathbb{R}^{\mathrm{n}}\} \tag{1.3}$$

   onde λ_k é um número real satisfazendo

$$0 < \lambda_k \leq \tilde{\lambda} \tag{1.4}$$

   para algum $\tilde{\lambda} > 0$ (incluindo o caso constante) e $||.||$ é a distância euclidiana.

. Nas últimas décadas, pesquisadores de vários países tem se interessado pelo estudo deste algoritmo. Assim, surgiram variações e generalizações deste algoritmo no sentido de englobar casos em que a função objetivo $f : \mathbb{R}^n \to \mathbb{R}$ é não convexa e casos em que o termo quadrático é substituído por uma distância generalizada, como a distância de Bregman, distância proximal, φ-divergências e quase-distâncias. Dentre os trabalhos que contém estas generalizações destacam-se os trabalhos de Teboulle [**?**], Eckstein [**?**], Iusem [**?**], Auslender [**?**] e Moreno et al. [**?**].

Esta classe de algoritmos proximais foi estendida para otimização vetorial e multiobjetivo e, neste sentido, destaca-se o trabalho *Proximal Methods in Vector Optimization* de Bonnel et al. [?]. Ainda nessa linha de pesquisa, temos os trabalhos de Rocha [?] e Apolinário [?]. Um ponto importante também são as generalizações desta classe de algoritmos proximais para espaços não euclidianos, como espaços de Banach e variedades Riemannianas e, neste caso, temos os trabalhos de Alber et al. [?] e Ferreira et al. [?].

1.1 Objetivos

Destaca-se que o APP surgiu inicialmente aplicado a funções $f : \mathbb{R}^n \to \mathbb{R}$ e, assim, para sua análise de convergência se faz necessário o conhecimento de conteúdos avançados como análise no $\mathbb{R}^n$, análise convexa, entre outros. Portanto, no meio acadêmico, somente discentes a nível de mestrado e/ou doutorado possuem os pré-requisitos necessários para um desejável entendimento deste algoritmo.

Neste sentido, este trabalho tem como objetivo estudar os conceitos teóricos, explorando a parte geométrica (quando viável), que são necessários para um desejável entendimento do APP para otimização de funções convexas de uma variável real a valores reais, ou seja, vamos considerar $f : \mathbb{R}^{\mathrm{n}} \to \mathbb{R}$ com $n = 1$.

Observamos que este trabalho servirá de base para o Trabalho de Conclusão de Curso II (TCC II) cujo foco principal será realizar a análise de convergência, a implementação e experimentos numéricos do APP aplicado a funções $f : \mathbb{R} \to \mathbb{R}$.

1.2 Contribuições

Por se tratar de um tema de alto nível de abstração, o estudo do APP requer vários pré-requisitos que a maioria dos cursos de exatas não contemplam durante um curso de graduação. Com isso, a principal contribuição deste trabalho é divulgar o algoritmo entre os acadêmicos de graduação de diversas áreas como matemática, computação e engenharias, de uma forma mais acessível e didática.

Este trabalho, em conjunto com o TCC II, propõe um estudo completo do APP aplicado a funções $f : \mathbb{R} \to \mathbb{R}$. Neste sentido, os pré-requisitos para um bom entendimento do algoritmo são funções convexas, cálculo diferencial e integral e sequências em $\mathbb{R}$, ou seja, temas que são lecionados nos cursos de graduação das áreas citadas no parágrafo anterior.

Além da análise de convergência e interpretações geométricas, o TCC II irá complementar o estudo do algoritmo com uma implementação utilizando a linguagem computacional GNU Octave [?], que é uma linguagem de alto nível desenvolvida para computação matemática. Por se tratar de um estudo quase que puramente matemático, boa parte dos trabalhos sobre algoritmos proximais deixa essa lacuna e não trazem implementações em ambientes computacionais. Assim, este complemento possibilita a comparação dos resultados de diversos métodos de otimização global que podem ser utilizados para encontrar mínimos globais de funções.

2 Preliminares

Neste capítulo, vamos expor os conceitos e resultados básicos que são imprescindíveis para a análise de convergência do algoritmo de ponto proximal aplicado a funções convexas de uma variável real a valores reais. Mais detalhes podem ser conferidos em diversos livros de análise real, análise convexa e otimização tais como: Lima [**?**], Àvila [**?**], Ribeiro [**?**] e Karas [**?**] e Rockafellar [**?**].

2.1 Sequência e subsequência

Definição 2.1.1 (Sequência em $\mathbb{R}$)**.** *Uma sequência de números reais é uma função* $x : \mathbb{N} \to \mathbb{R}$*. O valor de* $x(k)$*, para todo* $k \in \mathbb{N}$*, será representado por* x^k *e chamado o termo de ordem* k*, ou* $k-$*ésimo termo da sequência. Escrevemos* $\{x^1, ..., x^k, ...\}$*, ou* $\{x^k\}_{k\in\mathbb{N}}$*, ou simplesmente* $\{x^k\}$ *para indicar a sequência* x*.*

Definição 2.1.2 (Subsequência em $\mathbb{R}$)**.** *Dada uma sequência* $x = \{x^k\}_{k\in\mathbb{N}}$ *de números reais, uma subsequência de* x *é a restrição da função* x *a um subconjunto infinito* $\mathbb{N}' = \{k_1 < ... < k_i < ...\}$ *de* $\mathbb{N}$*. Neste caso, escreve-se* $x' = \{x^k\}_{k\in\mathbb{N}'}$ *ou* $\{x^{k_1}, ..., x^{k_i}, ...\}$ *ou* $\{x^{k_i}\}_{i\in\mathbb{N}}$ *para indicar a subsequência* $x' = x|\,\mathbb{N}'$*.*

Examplo 2.1.1. *Os números naturais que formam os números primos é um exemplo de sequência.*

$$X_n = (2, 3, 5, 7, 11, 13, 17, 19, ..., 97)$$

Diz-se que a sequência $\{x^k\}$ *é limitada se existem constantes* $a, b \in \mathbb{R}$ *tais que, para todo* $k \in \mathbb{N}$*,* $a \leq x^k \leq b$ *. Quando* $x^k \leq b$ *(respect.* $a \leq x^k$ *) para todo* $k \in \mathbb{N}$ *diz-se que* $\{x^k\}$ *é limitada superiormente (respect. inferiormente). Uma sequência* $\{x^k\}$ *chama-se crescente (respect. decrescente) quando* $x^k < x^{k+1}$ *(respect.* $x^k > x^{k+1}$ *) para todo* $k \in \mathbb{N}$*. Se vale* $x^k \leq x^{k+1}$ *(respect.* $x^k \geq x^{k+1}$*) a sequência diz-se não-decrescente (respect. não-crescente). As sequências crescentes, não-decrescentes, decrescentes e não-crescentes são chamadas sequências monótonas.*

Examplo 2.1.2. *A subsequência da sequência* $(1/n)$*,*

$$1, \frac{1}{2}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, ...$$

é a sequência $(1/k^2)$

$$1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \frac{1}{25}, ...$$

Aqui, $n_k = k^2$. Por outro lado, as sequências

$$1, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots \qquad \frac{1}{2}, 1, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots$$

não são subsequências de $(1/n)$ já que n_k não é uma função estritamente crescente de k em nenhum caso.

Definição 2.1.3. *Diz-se que um número real a é limite de uma sequência $\{x^k\}$ de números reais, e escreve-se $a = \lim\limits_{k\to\infty} x^k$ (ou $a = \lim x^k$), quando para cada número real $\varepsilon > 0$, dado arbitrariamente, for possível obter um natural $k_0 \in \mathbb{N}$ tal que $|x^k - a| < \varepsilon$, sempre que $k > k_0$.*

Uma sequência que possui limite chama-se convergente. Um ponto $\bar{a} \in \mathbb{R}$ é dito ser um ponto de acumulação de uma sequência $\{x^k\}$ se existe uma subsequência $\{x^{k_j}\}$ tal que $\lim\limits_{j\to\infty} x^{k_j} = \bar{a}$.

Examplo 2.1.3. *Provaremos que a sequência*

$$(x_n) = \left(\frac{n}{n+12}\right) = \left(\frac{1}{13}, \frac{2}{14}, \frac{3}{15}, \ldots, \frac{n}{n+12}\right)$$

converge para o número $L = 1$.

Observe que:

$$|x_n - 1| = |\frac{n}{n+12} - 1| = \frac{12}{n+12} < \epsilon \iff n > \frac{12}{\epsilon} - 12$$

Isso quer dizer que, dado qualquer $\epsilon > 0$, existe $N = max(1, 12/\epsilon - 12)$ tal que

$$n > N \implies |x_n - 1| < \epsilon$$

Ou seja, quanto menor for o valor de ϵ, maior será o valor de N. Assim, se $\epsilon = 10^{-k}$, então $N = 12 \cdot 10^k - 12$.

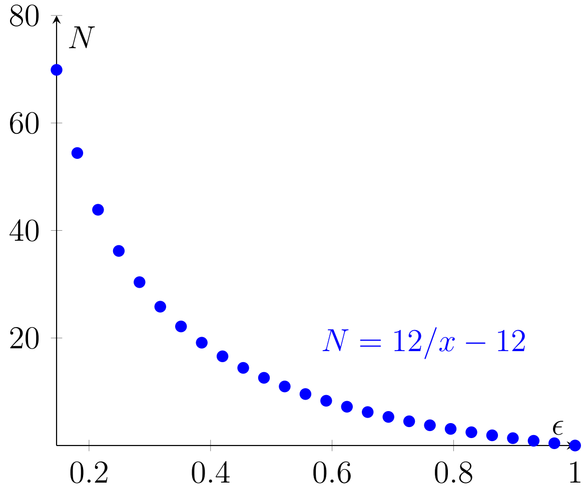


Figura 2.1: *Sequência de números naturais*

Na figura **??**, *podemos notar que para ϵ arbitrariamente pequeno, o valor de N cresce exponencialmente.*

Para as próximas quatro Proposições, $\{x^k\}$ é uma sequência em $\mathbb{R}$.

Proposição 2.1.1. *Se $\lim_{k\to\infty} x^k = a$ então toda subsequência de $\{x^k\}$ converge para o limite a. Em particular, se $\{x^k\}$ for convergente e possuir uma subsequência que converge para a então $\lim_{k\to\infty} x^k = a$.*

Demonstração. Como $x^k \to a$, dado qualquer valor $\epsilon > 0$, existe K tal que $k > K \implies |x^k - a| < \epsilon$. Mas, $k_j \geq j$, de forma que $j > K \implies k_j > K$; portanto, $j > K \implies |x^{k_j} - a| < \epsilon$. o que completa a demonstração. □

Proposição 2.1.2. *Se $\{x^k\}$ é convergente então $\{x^k\}$ é limitada.*

Demonstração. Dado qualquer $\epsilon > 0$, existe um índice K tal que $k > K \implies L - \epsilon < x^k < L + \epsilon$.

Isto quer dizer que a sequência é limitada à direita por $L + \epsilon$ e à esquerda por $L - \epsilon$, a partir de $k = K + 1$. Considere:

$$M = max\{|x^1|, |x^2|, |x^3|..., |x^K|, |L - \epsilon|, |L + \epsilon|\}$$

Então $|x^k| \leq M$ para todo k, logo a sequência é limitada. □

A sequência da figura **??** *não é crescente, nem decrescente e nem convergente. Porém, ela é limitada.*

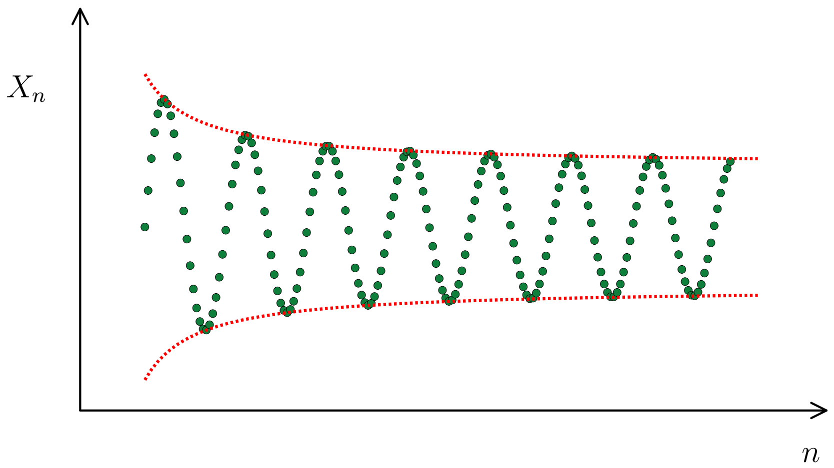


Figura 2.2: *Sequência infinita de números reais*

Proposição 2.1.3 (Teorema de Bolzano-Weierstrass)**.** *Se $\{x^k\}$ é limitada então $\{x^k\}$ possui uma subsequência convergente.*

Demonstração. Como $\{x^k\}$ é limitada, existe um número positivo M tal que, para todos os índices k, $-M < x^k < M$. Seja X o conjunto dos números x tais que existe uma infinidade de elementos da sequência à direita de x, isto é, $x < x^k$ para uma infinidade de índices k. É claro que $-M \in X$ e M é uma cota superior de X. Tratando-se. pois, de um conjunto não-vazio e limitado superiormente, X possui supremo, que designamos por A.

Vamos provar que existe uma subsequência convergindo para A. Começamos provando que, qualquer que seja $\epsilon > 0$, existem infinitos índices k tais que $A - \epsilon < x^k$ e somente um número finito satisfazendo $A + \epsilon < x^k$ De fato, sendo A o supremo de X, existe $x \in X$ à direita de $A - \epsilon$ e infinitos x^k à direita desse x, portanto à direita de $A - \epsilon$; ao mesmo tempo. só pode existir um número finito de elementos $x^k > A + \epsilon$; do contrário, qualquer número entre A e $A + \epsilon$ estaria em X.

Seja $\epsilon = 1$ e x^{k_1} um elemento da sequência no intervalo $(A-1, A+1)$. Em seguida, seja x^{k_2}, com $x^{k_2} > x^{k_1}$, um elemento da sequência no intervalo $(A - \frac{1}{2}, A + \frac{1}{2})$. Em seguida, seja x^{k_3}, $x^{k_3} > x^{k_2}$, um elemento da sequência no intervalo $(A - \frac{1}{3}, A + \frac{1}{3})$. Continuando com esse raciocínio, construímos uma subsequência $x^n = x^{k_j}$, que certamente converge para A, pois $|x^n - A| < \frac{1}{n}$. □

Proposição 2.1.4. *Se $\{x^k\}$ é monótona e limitada então $\{x^k\}$ é convergente.*

Demonstração. Seja $(x^1 \leq x^2 \leq x^1 \leq ... \leq x^k \leq ...)$ uma sequência não decrescente limitada. Tomemos $a = \sup\, \{x^k; k = 1, 2, ...\}$. Afirmamos que $a = \lim x^k$. Com efeito, dado qualquer $\epsilon > 0$, como $a - \epsilon < a$, o número $a - \epsilon$ não é cota superior do conjunto dos x^k. Logo existe algum $k^0 \in \mathbb{N}$ tal que $a - \epsilon < x^k$. Como a sequência é monótona, $k > k^0 \implies x^{k^0} \leq x^k$ e, portanto, $a - \epsilon < x^k$. Como $x^k \leq a$ para todo k, vemos que

$k > k^0 \implies a - \epsilon < x^k < a + \epsilon$. Assim, temos de fato $\lim x^k = a$, como queríamos demonstrar. □

Definição 2.1.4 (Fejér Convergência)**.** *Uma sequência $\{x^k\}$ em $\mathbb{R}$ é dita ser Fejér convergente para um conjunto $U \subset \mathbb{R}$ se*

$$|x^{k+1} - u| \leq |x^k - u|, \ \forall \ k \geq 0, \ \forall \ u \in U \tag{2.1}$$

Temos o seguinte resultado.

Proposição 2.1.5. *Se $\{x^k\}$ é Fejér convergente para $U \neq \emptyset$ então $\{x^k\}$ é limitada. Se um ponto de acumulação $\bar{x}$ de $\{x^k\}$ pertence a U então $\bar{x} = \lim\limits_{k\to\infty} x^k$*

Demonstração. (**??**) implica $|x^k - u| \leq |x^0 - u|$ para todos $k \geq 0$ e $u \in U$. Fixe $u_0 \in U$. Então, a sequência $\{x^k\}$ está contida no intervalo $[u_0 - |x^0 - u_0|, u_0 + |x^0 - u_0|]$ e, portanto, é limitada. Seja $\bar{x} \in U$ um ponto de acumulação de $\{x^k\}$. Então existe $\{x^{k_j}\}$ subsequência de $\{x^k\}$ tal que $\bar{x} = \lim\limits_{j\to\infty} x^{k_j}$. Desde que $\bar{x} \in U$, por (**??**), $\{|x^k - \bar{x}|\}$ é uma sequência decrescente e não negativa e, portanto, convergente. Como $|x^{k_j} - \bar{x}|$ converge para 0, toda a sequência também converge para 0, isto é, $\lim\limits_{k\to\infty} |x^k - \bar{x}| = 0$, implicando $\lim\limits_{k\to\infty} x^k = \bar{x}$. □

2.2 Mínimo de funções

Considere um conjunto $X \subset \mathbb{R}$ e uma função $f : X \to \mathbb{R}$. Um problema matemático muito estudado na literatura e que possui aplicações em diversas áreas de pesquisa, tais como Economia, Logística, Transporte, Engenharia e Biologia, consiste em encontrar um minimizador de f em X. Neste caso, escrevemos o problema como

$$\min \ \{f(x) \mid x \in X\}. \tag{2.2}$$

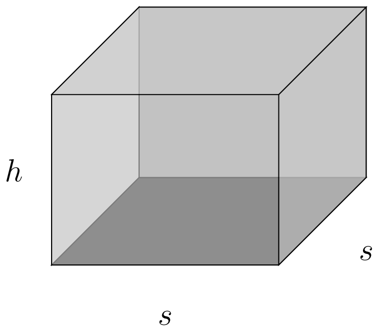
O conjunto X é chamado conjunto viável do problema, os pontos de X são chamados pontos viáveis, e f é chamada função objetivo.

*Um caso particular importante deste problema, e que será o objeto deste trabalho, é considerar $X = \mathbb{R}$. Neste caso, dizemos que o problema (**??**) é irrestrito e ele se torna*

$$\min \ \{f(x) \mid x \in \mathbb{R}\}. \tag{2.3}$$

Definição 2.2.1. *Dizemos que um ponto $x^* \in X$ é um minimizador local de f em X quando existe $\delta > 0$, tal que $f(x^*) \leq f(x)$ para todo $x \in (x^* - \delta, \ x^* + \delta) \cap X$. Caso $f(x^*) \leq f(x)$, para todo $x \in X$, x^* é dito minimizador global de f em X.*

Examplo 2.2.1. *Uma caixa sem tampa, de base quadrada, deve ser construída de forma que o seu voluma seja 2.500 m^3. O material da base vai custar R\$ 1.200,00 por m^2 e* *o material dos lados R\$ 980.00 por* m^2. *Encontre as dimensões da caixa de modo que o custo do material seja mínimo.*



$$f(h,s) = s^2 \cdot 1.200,00 + 4hs \cdot 980,00$$

Figura 2.3: *Caixa metálica de fundo quadrado*

A função que representa o custo do material é:

$$f(h,s) = s^2 \cdot 1.200,00 + 4hs \cdot 980,00 \tag{2.4}$$

Como $V = s^2h = 2.500\ cm^3$, *temos que a dimensão* h *pode ser escrita como* $h = 2.500/s^2$. *Substituindo esse resultado em* (**??**), *temos* $f(s) = 1.200,00 \cdot s^2 + 9.800.000,00/s$, *Com isso, a função fica dependente de apenas uma variável e, em seguida, iremos minimizar tal função.*

$$f'(s) = \frac{2.400,00s^3 - 9.800.000,00}{s^2}$$

Resolvendo essa equação $f'(s) = 0$, *obtemos* $s = 5\sqrt[3]{\frac{98}{3}} \approx 15,983$, *que é o ponto crítico que nos interessa.*

De fato, para $s \approx 15{,}983$ *teremos um ponto de mínimo, já que* $f''(15,983) > 0$. *Portanto, as dimensões da caixa de modo a obter o menor custo possível são* $s \approx 15{,}983$ *m e* $h \approx 9{,}785$ *m.*

Na figura **??**, *a função decresce à medida que* s *cresce de* 0 *até aproximadamente* $15,983m$. *A partir desse ponto, a função cresce à medida que* s *cresce. Logo, o mínimo global ocorre quando* $s \approx 15,983m$ *e* $h \approx 9,875m$.

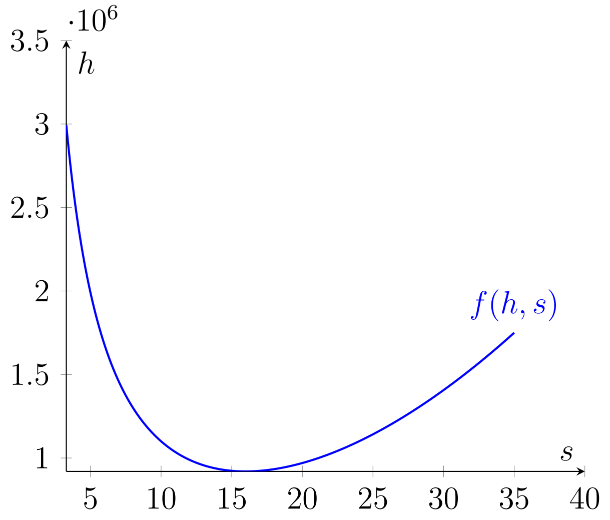


Figura 2.4: *Função de minimização da caixa metálica*

2.3 Funções contínuas e funções deriváveis

Definição 2.3.1 (Função Contínua)**.** *Uma função $f : X \to \mathbb{R}$, definida no conjunto $X \subset \mathbb{R}$, diz-se contínua no ponto $a \in X$ quando, para todo $\varepsilon > 0$ dado arbitrariamente, pode-se obter $\delta > 0$ tal que $x \in X$ e $|x - a| < \delta$ implicam $|f(x) - f(a)| < \varepsilon$. Diz-se que $f : X \to \mathbb{R}$ é uma função contínua quando f é contínua em todos os pontos $a \in X$.*

Examplo 2.3.1. *A função $f(x) = \dfrac{1}{x}$ é contínua embora seu gráfico tenha um salto em $x = 0$.*

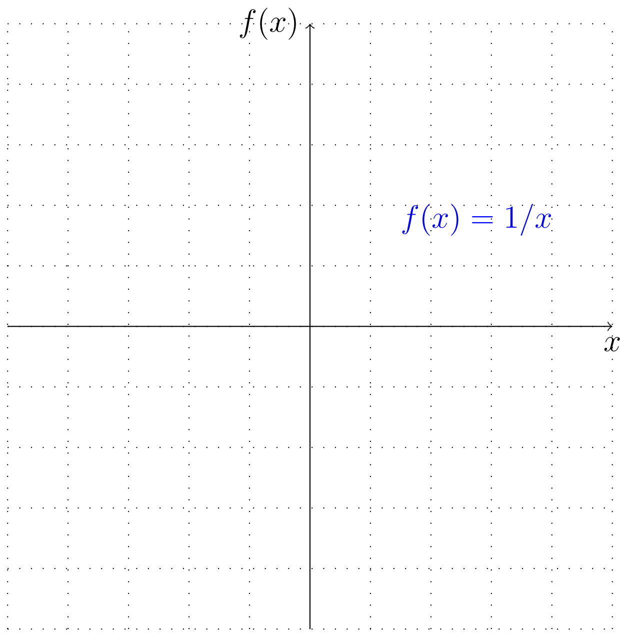


Figura 2.5: *Função recíproca*

Proposição 2.3.1. *Sejam $f : \mathbb{R} \to \mathbb{R}$ contínua e $X \subset \mathbb{R}$ um subconjunto compacto (fechado e limitado) não vazio. Então existe um minimizador global de f em X.*

Demonstração. XXXXX □

Proposição 2.3.2. *Uma função $f : X \to \mathbb{R}$ seja contínua no ponto $a \in X$ se, e somente se, para toda sequência $\{x^k\} \subset X$ com $\lim x^k = a$ se tenha $\lim f(x^k) = f(a)$.*

Demonstração. XXXXX □

Definição 2.3.2 (Função Derivável)**.** *Sejam $X \subset \mathbb{R}$ e $a \in X$ um ponto de acumulação de X. Diremos que $f : X \to \mathbb{R}$ é derivável no ponto a quando existir o limite $f'(a) = \lim\limits_{x \to a} \dfrac{f(x) - f(a)}{x - a}$.*

Quando existe, o limite $f'(a)$ chama-se a derivada de f no ponto a. Diremos que $f : X \to \mathbb{R}$ é derivável no conjunto X quando f é derivável em todos os pontos de acumulação de X que pertencem a X.

Quando uma função $f : I \to \mathbb{R}$ possui derivada em todos os pontos do intervalo I, considera-se a função derivada $f' : I \to \mathbb{R}$, que associa a cada $x \in I$ a derivada $f'(x)$. Quando a função f' é contínua, diz-se que f é uma função continuamente derivável no intervalo I, ou uma função de classe C^1.

Proposição 2.3.3. *Se $f : \mathbb{R} \to \mathbb{R}$ é derivável no ponto $a \in \mathbb{R}$ e possui um minimizador local nesse ponto, então $f'(a) = 0$.*

Demonstração. Como f derivável no ponto a, temos $f'_+(a) = f'_-(a) = f'(a)$. Como a é um minimizador local de f, para pontos x suficientemente próximo de a, $f(a) \leq f(x)$. Suponha que $x \to a^+$, então $x - a > 0$. Assim, $f'_+(a) = \lim\limits_{x \to a^+} \frac{f(x)-f(a)}{x-a} \geq 0$. De forma análoga, $f'_-(a) = \lim\limits_{x \to a^-} \frac{f(x)-f(a)}{x-a} \leq 0$. Portanto, como $f'_+(a) = f'_-(a) = f'(a)$ temos $f'(a) = 0$. □

2.4 Funções convexas

Definição 2.4.1 (Função convexa)**.** *Uma função $f : I \to \mathbb{R}$, definida num intervalo, chama-se convexa quando, para $a < x < b$ arbitrários em I, o ponto $(x, f(x))$ do gráfico de f está situado abaixo do segmento de reta que liga os pontos $(a, f(a))$ e $(b, f(b))$.*

Proposição 2.4.1. *$f : I \to \mathbb{R}$ (I intervalo) é convexa se, e somente se, para quaisquer a, b em I e $0 \leq t \leq 1$ vale $f((1-t)a + tb) \leq (1-t)f(a) + tf(b)$.*

Observação 2.4.1. *Tomando a desigualdade estrita ($<$) em vez de ($\leq$) na Proposição anterior, obtemos o conceito de função estritamente convexa.*

Geometricamente, esta inequalidade significa o segmento de reta $(x, f(x))$ e $(y, f(y))$, que é a corda de x para y representada na figura **??**.

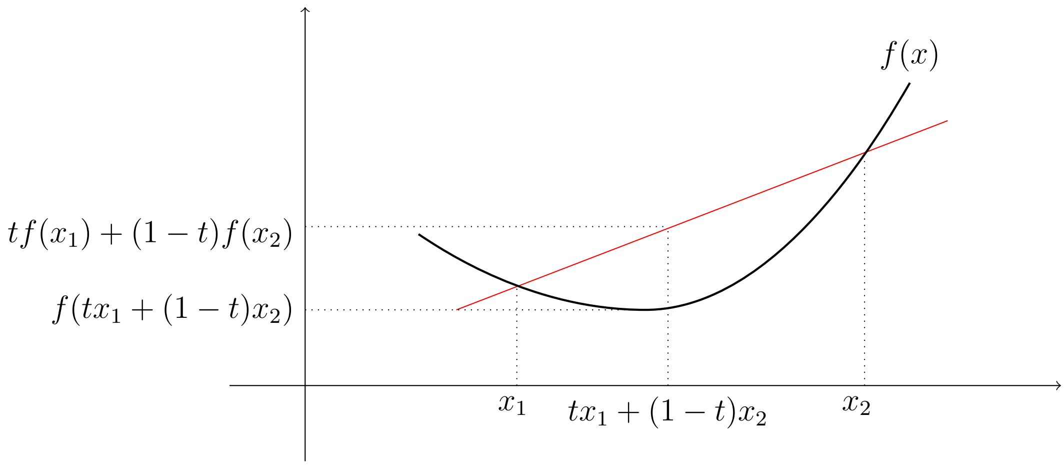


Figura 2.6: *Função convexa*

Examplo 2.4.1. *Abaixo temos alguns exemplos de funções convexas.*

- *A função $f(x) = x^2$ possui $f''(x) = 2 > 0$ em todos os pontos, logo f é uma função convexa. Esta função também é estritamente convexa.*
- *A função $f(x) = x^4$ possui $f''(x) = 12x^2 > 0$, logo f é uma função convexa. Também é estritamente convexa apesar que a segunda derivada não é estritamente positiva em todos os pontos. Portanto, esta função não é estritamente convexa.*
- *O valor absoluto da função $f(x) = |x|$ é convexa, apesar de não ter derivada no ponto $x = 0$. Esta função não é estritamente convexa.*
- *A função $f(x) = |x|^p$ para $p \geq 1$ é convexa.*

Proposição 2.4.2. *Seja I um intervalo. Se $f : I \to \mathbb{R}$ é uma função convexa e $g : I \to \mathbb{R}$ é uma função estritamente convexa, então $(f + g) : I \to \mathbb{R}$ é uma função estritamente convexa.*

Demonstração. Sejam a, b em I e $0 < t < 1$. Desde que f é convexa e g é estritamente convexa, temos:

$$\begin{aligned}(f+g)((1-t)a+tb) &= f((1-t)a+tb) + g((1-t)a+tb) \\ &< (1-t)f(a) + tf(b) + (1-t)g(a) + tg(b) \\ &= (1-t)(f(a)+g(a)) + t(f(b)+g(b)) \\ &= (1-t)((f+g)(a)) + t(f+g)(b)).\end{aligned}$$

Portanto $f + g$ é uma função estritamente convexa. $\square$

Proposição 2.4.3. *Seja $f : \mathbb{R} \to \mathbb{R}$ uma função convexa. Se x^* é um minimizador local de f, então x^* é um minimizador global de f.*

Demonstração. Como x^* é um minimizador local de f existe $\delta > 0$ tal que $f(x^*) \leq f(x)$ para todo $x \in (x^* - \delta;\ x^* + \delta)$. Dado $y \notin (x^* - \delta;\ x^* + \delta)$, considere $t > 0$ tal que $t|y - x^*| < \delta$. Defina $x = (1-t)x^* + ty$. Assim, $|x - x^*| = t|y - x^*| < \delta$ e então $x \in (x^* - \delta;\ x^* + \delta)$. Portanto, $f(x^*) \leq f(x) = f((1-t)x^* + ty)$. Logo, desde que f é convexa, $f(x^*) \leq (1-t)f(x^*) + tf(y)$ e assim, $f(x^*) \leq f(y)$. $\square$

Proposição 2.4.4. *Se $f : \mathbb{R} \to \mathbb{R}$ é uma função estritamente convexa então, não pode haver mais de um minimizador.*

Demonstração. Suponha que f possui dois minimizadores globais x_1 e x_2 com $x_1 \neq x_2$. Então, para todo x em $\mathbb{R}$, $\bar{x} = f(x_1) = f(x_2) \leq f(x)$. Considere $t \in (0,1)$ e defina $x = tx_1 + (1-t)x_2$. Desde que x_1 e x_2 são minimizadores de f, $\bar{x} = f(x_1) = f(x_2) \leq f(x) = f(tx_1 + (1-t)x_2)$. Então, pela convexidade estrita da f, $\bar{x} < tf(x_1) + (1-t)f(x_2) = t\bar{x} + (1-t)\bar{x} = \bar{x}$, o que é uma contradição. Portanto, não pode existir mais de um minimizador. $\square$

Proposição 2.4.5. *Seja I um intervalo aberto. Se $f : I \to \mathbb{R}$ é uma função convexa e derivável em I, então,*

$$f'(x)(x-y) \geq f(x) - f(y);\ \forall x, y \in I.$$

Demonstração. Fixe $x \in I$. É fácil ver que a desigualdade é satisfeita para todo $y = x$. Suponha, então, que $y \neq x$. Desde que f é convexa, $f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$ para todo $0 \leq t \leq 1$ e $y \in I$, e então, $f(x + t(y-x)) \leq f(x) + t(f(y) - f(x))$. Assim,

$$(y-x)\frac{f(x + t(y-x)) - f(x)}{t(y-x)} \leq f(y) - f(x),$$

ou ainda,

$$f(y) \geq f(x) + (y-x)\frac{f(x + t(y-x)) - f(x)}{t(y-x)}.$$

Portanto, fazendo $t \to 0^+$, obtemos: $f(y) \geq f(x) + (y-x)f'(x)$, donde segue o resultado desejado. $\square$

Proposição 2.4.6. *Seja $f : \mathbb{R} \to \mathbb{R}$ uma função convexa e derivável em $\mathbb{R}$ tal que $f'(x^*) = 0$. Então x^* é um minimizador de f em $\mathbb{R}$.*

Demonstração. Pela proposição anterior, $f'(x^*)(x^* - y) \geq f(x^*) - f(y);\ \forall y \in \mathbb{R}$. Então, como $f'(x^*) = 0$, temos: $0 \geq f(x^*) - f(y);\ \forall y \in \mathbb{R}$, isto é, $f(x^*) \leq f(y);\ \forall y \in \mathbb{R}$. Portanto x^* é um minimizador de f em $\mathbb{R}$. $\square$

2.5 Funções coercivas

Definição 2.5.1 (Função coerciva)**.** *Uma função contínua $f : \mathbb{R} \to \mathbb{R}$ é dita ser coerciva, se*

$$\lim_{|x|\to\infty} f(x) = \infty,$$

ou seja, quando para todo $M > 0$ existe $r > 0$ tal que $f(x) > M$ sempre que $|x| > r$.

Examplo 2.5.1. *A figura* **??**, *mostra a função $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^4 - x^2$ que é coerciva.*

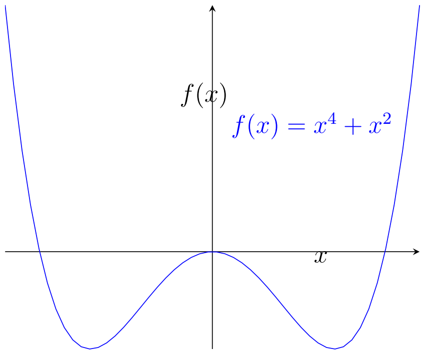


Figura 2.7: *Função coerciva*

Definição 2.5.2 (Conjunto de nível)**.** *O conjunto de nível da função $f : \mathbb{R} \to \mathbb{R}$ associado a $c \in \mathbb{R}$, é o conjunto dado por*

$$L_f(c) = \{x \in \mathbb{R} \mid f(x) \leq c\}.$$

Proposição 2.5.1. *Se $f : \mathbb{R} \to \mathbb{R}$ é uma função contínua e coerciva então, para todo $c \in \mathbb{R}$, o cojunto $L_f(c)$ é um conjunto compacto.*

Demonstração. Seja $c \in \mathbb{R}$. Se $L_f(c) = \emptyset$, então $L_f(c)$ é um conjunto compacto. Suponha agora que $L_f(c) \neq \emptyset$. Mostraremos que $L_f(c)$ é um conjunto fechado e limitado e portanto compacto. De fato, seja $\{x^k\} \subset L_f(c)$ tal que $\lim x^k = \bar{x}$. Então, desde que f é contínua e $f(x^k) \leq c$, concluímos que $f(\bar{x}) \leq c$ e assim $\bar{x} \in L_f(c)$. Logo $L_f(c)$ é um conjunto fechado. Considere $M > c$. Como f é coerciva existe $a > 0$ tal que se $|x| \geq a$ temos $f(x) \geq M > c$. Assim, o conjunto $L_f(c)$ está contido no intervalo $(-a, a)$ e portanto é limitado. $\square$

Corolário 2.5.1. *Se $f : \mathbb{R} \to \mathbb{R}$ é uma função contínua e coerciva então, f possui um minimizador global.*

Demonstração. Considere $x_0 \in \mathbb{R}$ fixado. Pela proposição anterior, o conjunto

$$L_f(f(x_0)) = \{x \mid f(x) \leq f(x_0)\}$$

é um conjunto compacto. Assim, desde que f é contínua, f possui ponto de mínimo em $L_f(f(x_0))$ que, por sua vez, coincide com o ponto de mínimo global. $\square$

Observação 2.5.1. *A Prova do corolário anterior nos mostra que quando a função* $f : \mathbb{R} \to \mathbb{R}$ *é uma função contínua e coerciva a minimização de* f *se reduz a um compacto.*

3 O algoritmo de ponto proximal

Neste capítulo, vamos enunciar e realizar a análise de convergência do algoritmo de ponto proximal (APP), denotado por APP real, que resolve o problema de otimização irrestrito

$$\min \{f(x) \mid x \in \mathbb{R}\}, \tag{3.1}$$

onde $f : \mathbb{R} \to \mathbb{R}$ *satisfaz as seguintes hipóteses:*

($\mathcal{H}$1) *f é convexa;*

($\mathcal{H}$2) *f é continuamente derivável.*

O APP gera uma sequência $\{x^k\} \subset \mathbb{R}$ *como segue:*

APP real

1. *Escolha* $x^0 \in \mathbb{R}$.

2. *Dado* x^k, *se* $x^k \in \operatorname{argmin}\{f(x) \mid x \in \mathbb{R}\}$, *então* $x^{k+p} = x^k, \forall p \geq 1$.

3. *Dado* x^k, *se* $x^k \notin \operatorname{argmin}\{f(x) \mid x \in \mathbb{R}\}$, *então, tome como próxima iterada, qualquer* x^{k+1} *tal que*

$$x^{k+1} = \operatorname{argmin}\left\{f(x) + \lambda_k (x - x^k)^2 \mid x \in \mathbb{R}\right\}, \tag{3.2}$$

onde $0 < \lambda_k < \bar{\lambda};\ \forall k \in \mathbb{N}$ *para algum* $\bar{\lambda} > 0$.

3.1 Existência e unicidade das iteradas

Nesta seção, provamos que a sequência $\{x^k\}$ *gerada pelo APP real é bem definida. Segue o resultado:*

Proposição 3.1.1. *Seja* $f : \mathbb{R} \to \mathbb{R}$ *uma aplicação convexa e continuamente derivável tal que o conjunto* U *dos minimizadores de* f *é não vazio. Então, o APP real sempre gera uma única sequência* $\{x^k\}$.

Demonstração. Por indução sobre k: x^0 é escolhido na etapa de inicialização. Supondo que o algoritmo atingiu a iteração k, vamos mostrar que um apropriado x^{k+1} existe. Pelo critério de parada, se $x^k \in \operatorname{argmin}\{f(x) \mid x \in \mathbb{R}\}$, então $x^{k+p} = x^k, \forall p \geq 1$. Por outro lado, defina

$$f_k(x) = f(x) + \lambda_k (x - x^k)^2.$$

Temos:

a) f é limitada inferiormente ($f(x) \geq M, \forall x \in \mathbb{R}$);

b) f_k é uma função contínua em $\mathbb{R}$;

c) f_k é uma função coerciva em $\mathbb{R}$;

d) f_k é uma função estritamente convexa em $\mathbb{R}$.

De fato:
a) Basta observar que $U \neq \emptyset$;
b) Como $f(x)$ e $\lambda_k(x - x^k)^2$ são funções contínuas em $\mathbb{R}$, a soma

$$f_k(x) = f(x) + \lambda_k(x - x^k)^2$$

também é uma função contínua em $\mathbb{R}$;
c) Pelo item a) $f(x) \geq M$ e assim, $f_k(x) \geq M + \lambda_k(x - x^k)^2$. Portanto, como $(x - x^k)^2$ é coerciva, isto é, $\lim\limits_{|x| \to \infty} (x - x^k)^2 = \infty$, temos $f_k(x) = f(x) + \lambda_k(x - x^k)^2$ também coerciva;
d) Desde que f é convexa e $\lambda_k(x - x^k)^2$ é estritamente convexa, a soma

$$f_k(x) = f(x) + \lambda_k(x - x^k)^2$$

é estritamente convexa.
O item c) implica que a minimização de $f_k(x) = f(x) + \lambda_k(x - x^k)^2$ em $\mathbb{R}$ se reduz a um compacto. Assim, desde que f_k é uma função contínua, f_k atinge seu mínimo. Além disto, como f_k é uma função estritamente convexa (item d)), f_k possui um único minimizador, isto é, x^{k+1} é unicamente determinada. $\square$

3.2 Critério de Parada

Nesta seção vamos estabelecer um critério de parada para o APP real. É importante observar que, devido aos erros de aproximação computacional, este critério de parada é quase impossível de ser atingido nos testes numéricos (conf. capítulo **??***). Neste sentido, para os testes numéricos utilizaremos o critério de parada ($|x^{k+1} - x^k| \leq tol$) que é justificado pela Proposição* **??** *[item a)]. Segue o resultado:*

Proposição 3.2.1. *Seja $\{x^k\}$ uma sequência gerada pelo APP real. Se $x^{k+1} = x^k$ então $x^k \in \operatorname{argmin}\{f(x) \mid x \in \mathbb{R}\}$.*

Demonstração. Suponha, por contradição, que $x^k \notin \operatorname{argmin}\{f(x) \mid x \in \mathbb{R}\}$. Então, existe $\bar{x} \in \mathbb{R}$ tal que $f(\bar{x}) < f(x^k)$. Assim, existe $\alpha > 0$ tal que $f(\bar{x}) = f(x^k) - \alpha$. Defina

$$x_\eta = \eta x^k + (1 - \eta)\bar{x}; \quad \eta \in (0; 1). \tag{3.3}$$

$x^{k+1} = x^k$, (**??**) e $x_\eta \in \mathbb{R}$ implicam:

$$f(x^k) \leq f(x_\eta) + \lambda_k (x_\eta - x^k)^2;\ \forall\, \eta \in (0,1). \tag{3.4}$$

Por (**??**),

$$\begin{aligned} x_\eta - x^k &= \eta x^k + (1-\eta)\bar{x} - x^k \\ &= (1-\eta)\bar{x} + (\eta - 1)x^k \\ &= (1-\eta)(\bar{x} - x^k). \end{aligned}$$

Então, por (**??**),

$$f(x^k) \leq f(x_\eta) + \lambda_k (1-\eta)^2 (\bar{x} - x^k)^2;\ \forall\, \eta \in (0,1). \tag{3.5}$$

Desde que f é convexa, $f(x_\eta) \leq \eta f(x^k) + (1-\eta) f(\bar{x})$. Logo, por (**??**),

$$\begin{aligned} f(x^k) &\leq \eta f(x^k) + (1-\eta) f(\bar{x}) + \lambda_k (1-\eta)^2 (\bar{x} - x^k)^2 \\ &\leq \eta f(x^k) + (1-\eta)(f(x^k) - \alpha) + \lambda_k (1-\eta)^2 (\bar{x} - x^k)^2 \\ &= \eta f(x^k) + f(x^k) - \eta f(x^k) - \alpha(1-\eta) + \lambda_k (1-\eta)^2 (\bar{x} - x^k)^2 \\ &= f(x^k) - \alpha(1-\eta) + \lambda_k (1-\eta)^2 (\bar{x} - x^k)^2. \end{aligned}$$

Daí, $\alpha(1-\eta) \leq \lambda_k (1-\eta)^2 (\bar{x} - x^k)^2$, e então,

$$\alpha \leq \lambda_k (1-\eta)(\bar{x} - x^k)^2.$$

Deste modo, fazendo $\eta \to 1^-$ concluímos que $\alpha \leq 0$, o que é uma contradição. Portanto $x^k \in \operatorname{argmin}\{f(x) \mid x \in \mathbb{R}\}$.

□

3.3 Fejér convergência da sequência gerada

Suponha que o conjunto U de minimizadores de f sobre $\mathbb{R}$ é não vazio. Nesta seção, provaremos que a sequência $\{x^k\}$ gerada pelo APP real é Fejér convergente para U.

Proposição 3.3.1. *Suponha $U \neq \emptyset$. Então a sequência $\{x^k\}$ gerada pelo APP real é Fejér convergente para U.*

Demonstração. Observe que:

$$\begin{aligned} (x^k - \bar{x})^2 &= (x^k - x^{k+1} + x^{k+1} - \bar{x})^2 \\ &= (x^k - x^{k+1})^2 + (x^{k+1} - \bar{x})^2 + 2(x^k - x^{k+1})(x^{k+1} - \bar{x}). \end{aligned} \tag{3.6}$$

Desde que x^{k+1} resolve (**??**) temos:

$$0 = f'(x^{k+1}) + 2\lambda_k(x^{k+1} - x^k). \tag{3.7}$$

De (**??**) e (**??**),

$$\begin{aligned} (x^k - \bar{x})^2 - (x^{k+1} - x^k)^2 - (x^{k+1} - \bar{x})^2 &= 2(x^k - x^{k+1})(x^{k+1} - \bar{x}) \\ &= 2\frac{1}{2\lambda_k} f'(x^{k+1})(x^{k+1} - \bar{x}) \\ &= \frac{1}{\lambda_k} f'(x^{k+1})(x^{k+1} - \bar{x}). \end{aligned} \tag{3.8}$$

Desde que f é uma função convexa, $f'(x)(x-y) \geq f(x) - f(y), \forall x, y \in \mathbb{R}$. Assim, de (**??**),

$$\begin{aligned} (x^k - \bar{x})^2 - (x^{k+1} - x^k)^2 - (x^{k+1} - \bar{x})^2 &= \frac{1}{\lambda_k} f'(x^{k+1})(x^{k+1} - \bar{x}) \\ &\geq \frac{1}{\lambda_k}[f(x^{k+1}) - f(\bar{x})] \geq 0. \end{aligned} \tag{3.9}$$

A última desigualdade decorre do fato de $\bar{x}$ ser um minimizador de f. Então,

$$(x^{k+1} - \bar{x})^2 \leq (x^k - \bar{x})^2 - (x^{k+1} - x^k)^2 \leq (x^k - \bar{x})^2. \tag{3.10}$$

Portanto,

$$|x^{k+1} - \bar{x}| \leq |x^k - \bar{x}|, \ \forall k \geq 0, \ \forall \bar{x} \in U. \tag{3.11}$$

□

3.4 Propriedades da sequência gerada

Nesta seção, estabelecemos duas propriedades da sequência gerada pelo APP real. Segue as propriedades:

Proposição 3.4.1. *Se $\{x^k\}$ for uma sequência gerada pelo APP real, então:*

a) $\lim_{k\to\infty}(x^{k+1} - x^k) = 0$ *e,*

b) $\{f(x^k)\}$ *é não crescente e convergente.*

Demonstração. a) Por (**??**), $\{|x^k - \bar{x}|\}$ é uma sequência decrescente e não negativa e, assim, convergente. Assim, desde que por (**??**),

$$0 \leq (x^{k+1} - x^k)^2 \leq (x^k - \bar{x})^2 - (x^{k+1} - \bar{x})^2,$$

temos: $\lim_{k\to\infty}(x^{k+1}-x^k)^2 = 0$. Portanto, $\lim_{k\to\infty}(x^{k+1}-x^k) = 0$.
b) Por (**??**),

$$f(x^{k+1}) + \lambda_k(x^{k+1}-x^k)^2 \leq f(x) + \lambda_k(x-x^k)^2,\ \forall x \in \mathbb{R},\ \forall k \geq 0.$$

Então, em particular, tomando $x = x^k$,

$$f(x^{k+1}) + \lambda_k(x^{k+1}-x^k)^2 \leq f(x^k),\ \forall k \geq 0.$$

Assim, desde que $\lambda_k > 0$, $\lambda_k(x^{k+1}-x^k)^2 \geq 0$ e então, $f(x^{k+1}) \leq f(x^k), \forall k \geq 0$. Logo, $\{f(x^k)\}$ é uma sequência não crescente. Como $U \neq \emptyset$, $\{f(x^k)\}$ é limitada inferiormente e, portanto, convergente. □

3.5 Convergência

Nesta seção vamos enunciar e provar o resultado principal deste trabalho. Segue o resultado.

Teorema 3.5.1. *Seja $f : \mathbb{R} \to \mathbb{R}$ convexa e continuamente derivável. Suponha que o conjunto U de minimizadores de f sobre $\mathbb{R}$ é não vazio. Então a sequência $\{x^k\}$ gerada pelo APP real converge para um ponto $x^* \in U$.*

Demonstração. Da Proposição **??**, $\{x^k\}$ é Féjer convergente para $U \neq \emptyset$. Assim, pela Proposição **??**, $\{x^k\}$ é limitada e, portanto, possui ponto de acumulação. Seja x^* um ponto de acumulação de $\{x^k\}$ e $\{x^{j_k}\}$ uma subsequência de $\{x^k\}$ tal que $\lim_{k\to\infty} x^{j_k} = x^*$. Por (**??**),

$$0 = f'(x^{j_k+1}) + 2\lambda_k(x^{j_k+1} - x^{j_k}). \tag{3.12}$$

Pela Proposição **??** [item a)], $\lim_{k\to\infty}(x^{j_k+1} - x^{j_k}) = 0$. Logo $\lim_{k\to\infty} x^{j_k+1} = \lim_{k\to\infty} x^{j_k} = x^*$. Como por (**??**), $\{\lambda_k\}$ é limitada, tomando o limite em (**??**) quando $k \to \infty$ e observando que f é continuamente derivável em $\mathbb{R}$, concluímos que $f'(\bar{x}) = 0$. Assim, desde que f é convexa, $x^* \in U$. Portanto, as hipóteses da proposição **??** são verdadeiras para a sequência $\{x^k\}$ e, portanto, existe $x^* \in U$ tal que $\lim_{k\to\infty} x^k = x^*$.

□

3.6 Convergência Finita

Nesta seção, vamos mostrar que se a sequência gerada pelo APP real converge para um ponto de mínimo que seja acentuado (sharp minimum) então, a convergência se realiza em um número finito de iterações.

Definição 3.6.1 (Mínimo acentuado)**.** *Uma função* $f : \mathbb{R} \to \mathbb{R}$ *é dita ter um mínimo acentuado ("sharp minimum") em* $\bar{x} \in \mathbb{R}$ *se, existem duas constantes positivas* η *e* δ *tal que*

$$\eta|x - \bar{x}| \leq f(x) - f(\bar{x}), \ \forall \ x \in (\bar{x} - \delta; \ \bar{x} + \delta). \tag{3.13}$$

Proposição 3.6.1. *Suponha, que uma sequência* $\{x^k\}$ *gerada pelo APP real, convirja para um minimizador de* f*, que seja mínimo acentuado. Então, a convergência se realiza em um número finito de iterações.*

Demonstração. Seja $\{x^k\}$ uma sequência gerada pelo APP real tal que $\lim_{k\to\infty} x^k = x^*$ onde x^* é um mínimo acentuado de f. Por (**??**),

$$(x^* - x^k)^2 - (x^{k+1} - x^k)^2 = (x^{k+1} - x^*)^2 + 2(x^k - x^{k+1})(x^{k+1} - x^*); \ \forall \ k \in \mathbb{N}\,. \tag{3.14}$$

Como x^{k+1} resolve (**??**), $f(x^{k+1}) + \lambda_k(x^{k+1} - x^k)^2 \leq f(x^*) + \lambda_k(x^* - x^k)^2; \ \forall \ k \in \mathbb{N}$, e então,

$$f(x^{k+1}) - f(x^*) \leq \lambda_k \left[(x^* - x^k)^2 - (x^{k+1} - x^k)^2\right]; \ \forall \ k \in \mathbb{N}\,. \tag{3.15}$$

Combinando (**??**) e (**??**) obtemos:

$$f(x^{k+1}) - f(x^*) \leq \lambda_k \left[(x^{k+1} - x^*)^2 + 2(x^k - x^{k+1})(x^{k+1} - x^*)\right]; \ \forall \ k \in \mathbb{N}\,. \tag{3.16}$$

Desde que $x^k \to x^*$, pela Proposição (**??**) item a), temos também $x^{k+1} \to x^*$. Logo, sem perda de generalidades, considere k tal que $x^{k+1} \in (x^* - \delta; \ x^* + \delta)$. □

4 Experimentos Numéricos

Este capítulo apresenta experimentos utilizando o APP para encontrar mínimos globais de funções de uma variável real, ilustrando os diferentes tipos de comportamento qualitativo dos métodos numéricos utilizados. São considerados exemplos simples e avançados para enfatizar os aspectos qualitativos dos experimentos. Os experimentos realizados servem de motivação para investigação teórica e prática dos conceitos apresentados nos capítulos anteriores.

4.1 Pseudocódigo do APP

Para evitar que o trabalho fique extenso ao entrar em detalhes técnicos de programação, o código implementado em GNU Octave foi transcrito na forma de pseudocódigo para ilustrar a estrutura geral do algoritmo e facilitar sua implementação em outras linguagens de programação.

O algoritmo em si é bastante simples e consiste de valores iniciais aleatórios ou fixados que serão processados em um loop *com um condicional indicando o critério de parada (ver seção* **??***). No* loop *das iterações temos a função* minimize *que, em GNU Octave, é representada pelo comando* fminunc *do GNU Octave.*

Para realizar os testes, dois algoritmos foram codificados: um para funções com apenas um único mínimo global e um para funções com vários mínimos globais, respectivamente, PPALG Single e PPALG Multiple.

Algorithm 4.1.1 PPALG Single

procedure PPALG–SINGLE()
    for $t \in T$ **do**
        $tol \in \mathbb{R}$
        $x_0 \in \mathbb{R}$
        $M \leftarrow \{0..m\}$
        for $k \in M$ **do**
            $x_{k+1} \leftarrow \text{minimize}(f(x_k) + \lambda_k ||x - x_k||^2)$

            if $|x_k - x_{k+1}| \leq tol$ **then**
                stop
            end if
        end for
    end for
end procedure

Algorithm 4.1.2 PPALG Multiple

procedure PPALG–MULTIPLE()
    $tol \leftarrow 10^{-3}$
    $x_0 \in \mathbb{R}$
    $M \leftarrow \{0..m\}$
    for $k \in M$ **do**
        $x_{k+1} \leftarrow \text{minimize}(f(x_k) + \lambda_k ||x - x_k||^2)$

        if $|x_k - x_{k+1}| <= tol$ **then**
            stop
        end if
    end for
end procedure

- x_0*: valor inicial para* x*, ou seja, um número qualquer* $x_0 \in \mathbb{R}$
- M*: número de iterações. Um número qualquer* $M \in \mathbb{N}$ *onde* $0 \leq m_k < M$
- λ_k*: número real que satisfaz a condição* $0 < \lambda_k < \bar{\lambda}$
- tol*: tolerância*

A principal diferença entre as duas versões do APP é a forma como os resultados serão obtidos e analisados. No algoritmo **??**, *são realizados vários testes de valor máximo T com um número máximo M de iterações e. em seguida, obtemos a média desses resultados. Nessa versão do APP. as funções possuem apenas um mínimo global e, portanto, o ponto inicial x_0 pode ser qualquer um dentro do conjunto $\mathbb{R}$.*

Já no algoritmo **??**, *a função a ser analisada pode possuir vários mínimos globais. Assim, o ponto inicial x_0 e o mínimo global a ser atingido são fornecidos como parâmetros iniciais do algoritmo. Note que nesta versão do algoritmo, os testes são executados uma única vez com um número máximo M de iterações.*

4.2 O método de Newton

4.3 O algoritmo BFGS

4.4 Solução passo-a-passo

Antes de detalharmos os experimentos numéricos realizados com ajdua de um computador, iremos realizar a resolução passo-a-passo das seguintes funções:

- $f(x) = x^2 + 1$
- $f(x) = (x-2)^2(x+2)^2 + 2$
- $f(x) = \dfrac{x}{1+x^2}$

Dada a função $f(x) = x^2 + 1$, onde $x^* = 0$, escolhendo $x^0 = 3$ e $\lambda = 1$, temos

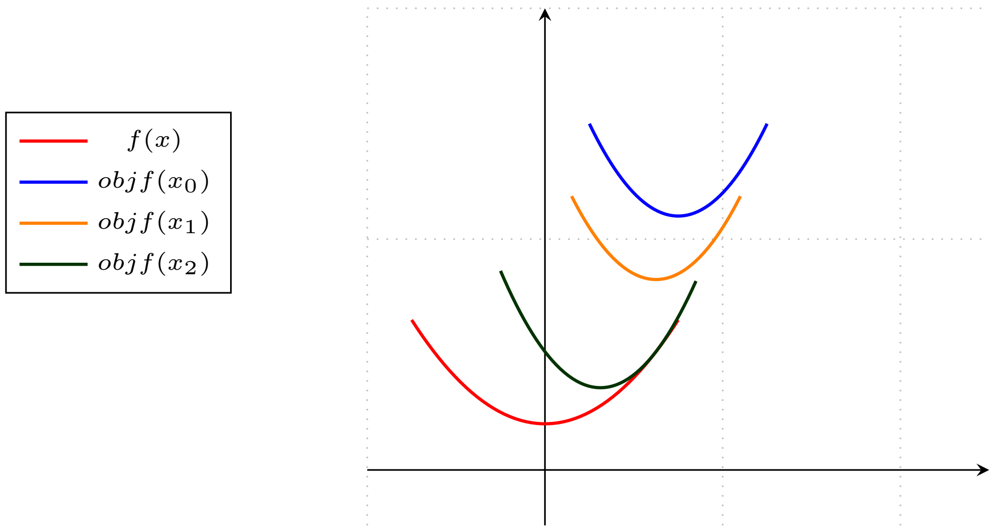
$$
\begin{aligned}
&x^1 = \arg\min\{f(x) + \lambda_0|x - x^0| \ ;\ x \in \mathbb{R}\}\\
&x^1 = \arg\min\{x^2 + 1 + (x-3)^2 \ ;\ x \in \mathbb{R}\}\\
&x^1 = \arg\min\{x^2 + 1 + x^2 - 6x + 9 \ ;\ x \in \mathbb{R}\}\\
&x^1 = \arg\min\{2x^2 - 6x + 10 \ ;\ x \in \mathbb{R}\}\\
&x^1 = 2.50000
\end{aligned}
$$

$$
\begin{aligned}
&x^2 = \arg\min\{f(x) + \lambda_1|x - x^1| \ ;\ x \in \mathbb{R}\}\\
&x^2 = \arg\min\{x^2 + 1 + (x-2.50000)^2 \ ;\ x \in \mathbb{R}\}\\
&x^2 = \arg\min\{x^2 + 1 + x^2 - 5x + 6.25000 \ ;\ x \in \mathbb{R}\}\\
&x^2 = \arg\min\{2x^2 - 5x + 7.25000 \ ;\ x \in \mathbb{R}\}\\
&x^2 = 1.25000
\end{aligned}
$$

$$
\begin{aligned}
&x^3 = \arg\min\{f(x) + \lambda_2|x - x^2| \ ;\ x \in \mathbb{R}\}\\
&x^3 = \arg\min\{x^2 + 1 + (x-1.25000)^2 \ ;\ x \in \mathbb{R}\}
\end{aligned}
$$

$x^3 = \arg\min\{x^2 + 1 + x^2 - 2.50000x + 1.56250 \text{ ; } x \in \mathbb{R}\}$
$x^3 = \arg\min\{2x^2 - 2.50000x + 2.56250 \text{ ; } x \in \mathbb{R}\}$
$x^3 = 0.62500$

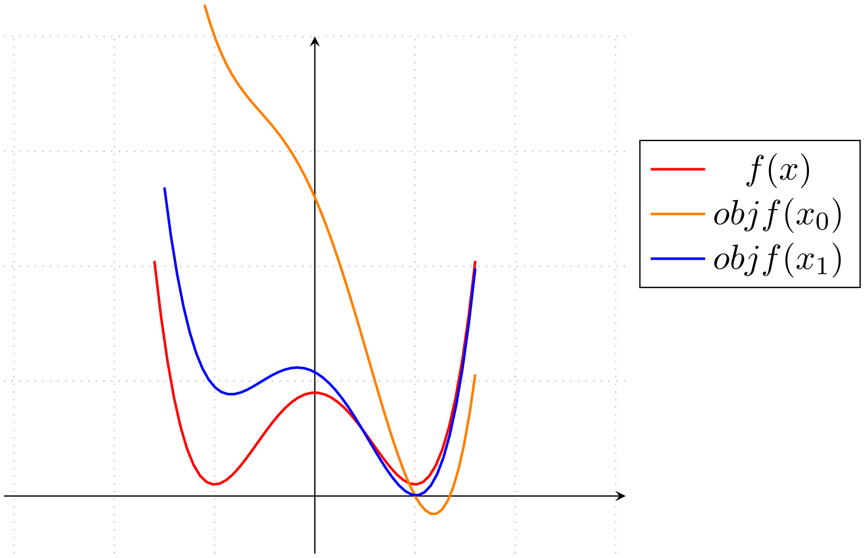
$x^4 = \arg\min\{f(x) + \lambda_3|x - x^3| \text{ ; } x \in \mathbb{R}\}$
$x^4 = \arg\min\{x^2 + 1 + (x - 0.62500)^2 \text{ ; } x \in \mathbb{R}\}$
$x^4 = \arg\min\{x^2 + 1 + x^2 - 1.25000x + 0.39100 \text{ ; } x \in \mathbb{R}\}$
$x^4 = \arg\min\{2x^2 - 1.25000x + 1.39100 \text{ ; } x \in \mathbb{R}\}$
$x^4 = 0.31250$



Dada a função $f(x) = (x-2)^2(x+2)^2 + 2$*, onde* x^* *possui dois pontos, -2 e 2, escolhendo* $x^0 = 6$ *e* $\lambda = 1$*, temos*

$x^1 = \arg\min\{f(x) + \lambda_0|x - x^0| \text{ ; } x \in \mathbb{R}\}$
$x^1 = \arg\min\{(x-2)^2(x+2)^2 + 2 + (x - 6)^2 \text{ ; } x \in \mathbb{R}\}$
$x^1 = \arg\min\{x^4 - 8x^2 + 16 + x^2 - 12x + 36 \text{ ; } x \in \mathbb{R}\}$
$x^1 = \arg\min\{x^4 - 7x^2 - 20x + 52 \text{ ; } x \in \mathbb{R}\}$
$x^1 = 2.36000$

$x^2 = \arg\min\{f(x) + \lambda_1|x - x^1| \text{ ; } x \in \mathbb{R}\}$
$x^2 = \arg\min\{(x-2)^2(x+2)^2 + 2 + (x - 2.36000)^2 \text{ ; } x \in \mathbb{R}\}$
$x^2 = \arg\min\{x^4 - 8x^2 + 16 + x^2 - 4.72000x + 5.56960 \text{ ; } x \in \mathbb{R}\}$
$x^2 = \arg\min\{x^4 - 7x^2 - 4.72000x + 21.56000 \text{ ; } x \in \mathbb{R}\}$
$x^2 = 2.02087$



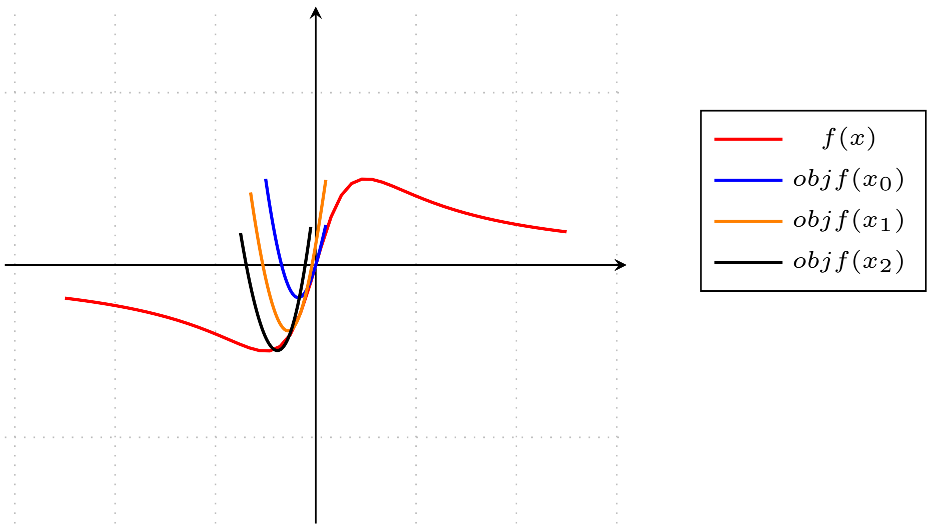
Dada a função $f(x) = \dfrac{x}{1+x^2}$, *onde* $x^* = -1$, *escolhendo* $x^0 = 0$ *e* $\lambda = 1$, *temos*

$$x^1 = \arg\min\left\{\left(\frac{x}{1+x^2}\right) + \lambda_0|x - x^0|^2\right\}$$
$$x^1 = \arg\min\left\{\left(\frac{x}{1+x^2}\right) + |x - 0|^2\right\}$$
$$x^1 = \arg\min\left\{\left(\frac{x}{1+x^2}\right) + x^2\right\}$$
$$x^1 = \arg\min\left\{\frac{x}{1+x^2} + x^2\right\}$$
$$x^1 = -0.34897$$

$$x^2 = \arg\min\left\{\left(\frac{x}{1+x^2}\right) + \lambda_1|x - x^1|^2\right\}$$
$$x^2 = \arg\min\left\{\left(\frac{x}{1+x^2}\right) + |x + 0.34897|^2\right\}$$
$$x^2 = \arg\min\left\{\left(\frac{x}{1+x^2}\right) + x^2 + 0.69800x + 0.12170\right\}$$
$$x^2 = -0.66830$$

$$x^3 = \arg\min\left\{\left(\frac{x}{1+x^2}\right) + \lambda_2|x - x^2|^2\right\}$$
$$x^3 = \arg\min\left\{\left(\frac{x}{1+x^2}\right) + |x + 0.66830|^2\right\}$$
$$x^3 = \arg\min\left\{\left(\frac{x}{1+x^2}\right) + x^2 + 1.36660x + 0.44660\right\}$$
$$x^3 = -0.76550$$

$$x^4 = \arg\min\left\{\left(\frac{x}{1+x^2}\right) + \lambda_2|x - x^3|^2\right\}$$
$$x^4 = \arg\min\left\{\left(\frac{x}{1+x^2}\right) + |x + 0.76550|^2\right\}$$
$$x^4 = \arg\min\left\{\left(\frac{x}{1+x^2}\right) + x^2 + 1.53100x + 0.58600\right\}$$
$$x^4 = -0.82290$$



4.5 Testes computacionais

Os testes a seguir foram realizados utilizando uma máquina com processador Intel Core i7 5500U, 8GB de RAM e sistema operacional Windows 10 64 bits.

EXPLICAR METODOLOGIA DOS TESTES

$x^* = -1$, $tol = 10^{-5}$, $k_{max} = 300$

Tabela 4.1: *Testes numéricos para* $f(x) = x/(1+x^2)$

$No.$	x^0	λ_k	k	T	x_k	$\lvert x_k - x^*\rvert$
1	0.5	$1+1/k$	36	0.2043240070	-0.9997671368	0.0002328632
2	0	$1+1/k$	34	0.1958851814	-0.9997059174	0.0002940826
3	-2	$1/k$	13	0.0789399147	-1.0000694690	0.0000694690
4	-3	1	64	0.3279368877	-1.0002452090	0.0002452090
5	-4	$2-1/k$	177	1.0065820217	-1.0009525885	0.0009525885
6	-4.5	$3+(-1)^k$	294	1.6411769390	-1.0009134014	0.0009134014
7	-6	$1/k$	24	0.1571619511	-1.0000520350	0.0000520350

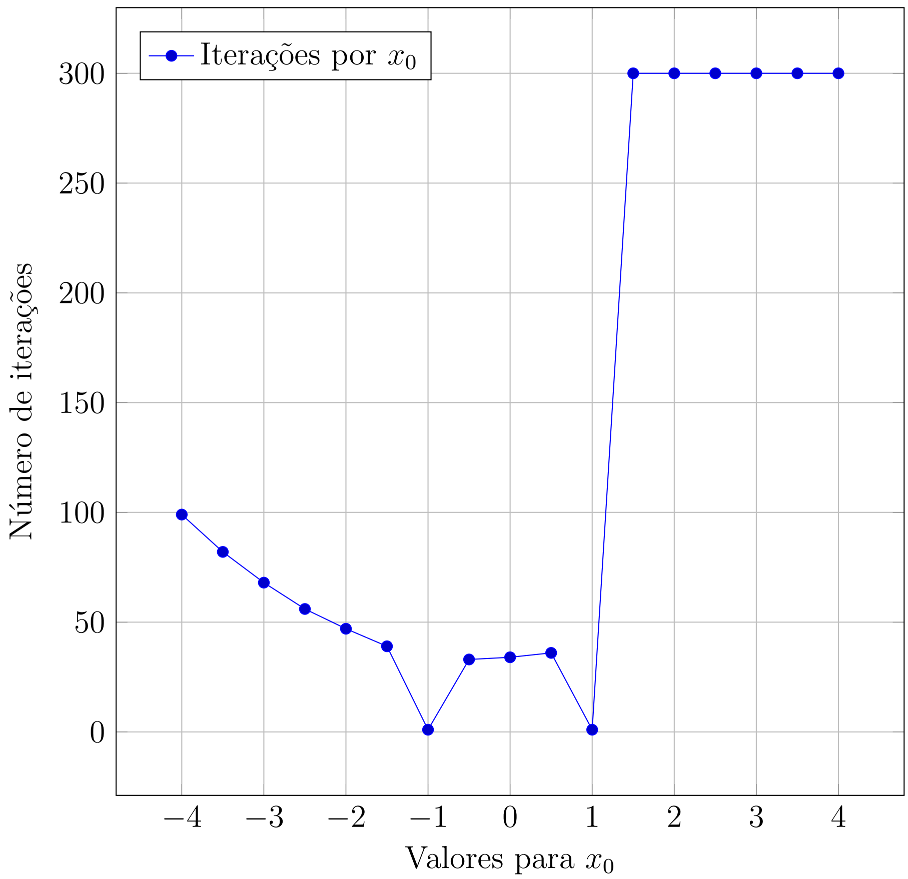


Figura 4.1: *Iterações por* x_0 *para* $f(x) = x/(1+x^2)$ *com* $\lambda_k = 1+1/k$

$x^* = 1,\ k_{max} = 300$

Tabela 4.2: *Testes numéricos para $f(x) = -x/e^x$*

$No.$	tol	λ_k	$\bar{k}$	$\bar{T}$	$\bar{x_k}$	$\lvert\bar{x_k} - x^*\rvert$
1	10^{-3}	$1 + 1/k$	19.00	0.1232542562	0.9507484160	0.0492515840
2	10^{-4}	$1 + 1/k$	32.94	0.2012740707	0.9951057732	0.0048942268
3	10^{-5}	$1 + 1/k$	46.75	0.2588648272	0.9996138464	0.0003861536
1	10^{-3}	$1/k$	9.00	0.0689284921	0.9956346880	0.0043653120
2	10^{-4}	$1/k$	11.73	0.0786435270	0.9996319656	0.0003680344
3	10^{-5}	$1/k$	14.00	0.0877796340	0.9998715698	0.0001284302
1	10^{-3}	1	17.08	0.1102390718	0.9504805198	0.0495194802
2	10^{-4}	1	30.63	0.1817432737	0.9951211131	0.0048788869
3	10^{-5}	1	44.00	0.2387099981	0.9995929016	0.0004070984
1	10^{-3}	$2 - 1/k$	22.92	0.1445117092	0.9092888780	0.0907111220
2	10^{-4}	$2 - 1/k$	46.75	0.2780978322	0.9899133078	0.0100866922
3	10^{-5}	$2 - 1/k$	71.00	0.3933139539	0.9986916709	0.0013083291
1	10^{-3}	$3 + (-1)^k$	24.00	0.1502644038	0.8372268509	0.1627731491
2	10^{-4}	$3 + (-1)^k$	54.00	0.3196566200	0.9804455581	0.0195544419
3	10^{-5}	$3 + (-1)^k$	92.44	0.5123488069	0.9986989529	0.0013010471

$x^* = 2,\ k_{max} = 100$

Tabela 4.3: *Testes numéricos para* $f(x) = (x-2)^4 - (x-2)^3 + (x-2)^2 + 1$

$No.$	tol	λ_k	$\bar{k}$	$\bar{T}$	$\bar{x_k}$	$\lvert\bar{x_k} - x^*\rvert$
1	10^{-3}	$1+1/k$	8.97	0.0536327481	1.9959030584	0.0040969416
2	10^{-4}	$1+1/k$	12.70	0.0730510926	1.9996377060	0.0003622940
3	10^{-5}	$1+1/k$	15.94	0.0849111366	1.9999468091	0.0000531909
1	10^{-3}	$1/k$	5.54	0.0350753999	1.9996027336	0.0003972664
2	10^{-4}	$1/k$	6.66	0.0380098605	1.9999675003	0.0000324997
3	10^{-5}	$1/k$	8.07	0.0442206478	1.9999999800	0.0000000200
1	10^{-3}	1	7.81	0.0482297301	1.9961353842	0.0038646158
2	10^{-4}	1	11.02	0.0622688532	2.0002121291	0.0002121291
3	10^{-5}	1	12.00	0.0649589300	2.0000416622	0.0000416622
1	10^{-3}	$2-1/k$	10.14	0.0607877707	1.9927548751	0.0072451249
2	10^{-4}	$2-1/k$	15.63	0.0826803327	1.9994193172	0.0005806828
3	10^{-5}	$2-1/k$	20.99	0.1053015375	1.9999593442	0.0000406558
1	10^{-3}	$3+(-1)^k$	12.55	0.0711864877	1.9875900588	0.0124099412
2	10^{-4}	$3+(-1)^k$	19.28	0.1093654108	1.9987759633	0.0012240367
3	10^{-5}	$3+(-1)^k$	26.79	0.1376921368	1.9999738393	0.0000261607

$tol = 10^{-5}$, $k_{max} = 200$

Tabela 4.4: *Testes numéricos para* $f(x) = (x-1)^2(x-3)^2(x-5)^2 + 1$

f	x^*	x^0	λ_k	k	T	x_k	$\|x_k - x^*\|$
1	1	0	$1+1/k$	4	0.0395190716	0.9999979428	0.0000020572
2	1	0	$1/k$	4	0.0392930508	0.9999999068	0.0000000932
3	1	0	1	4	0.0382940769	0.9999994526	0.0000005474
4	1	0	$2-1/k$	4	0.0383739471	0.9999986573	0.0000013427
5	1	0	$3+(-1)^k$	4	0.0330469608	0.9999921664	0.0000078336
1	3	2	$1+1/k$	5	0.0493650436	3.0000268953	0.0000268953
2	3	2	$1/k$	4	0.0338649750	3.0000182829	0.0000182829
3	3	2	1	5	0.0343091488	3.0000060172	0.0000060172
4	3	2	$2-1/k$	5	0.0395460129	3.0000235686	0.0000235686
5	3	2	$3+(-1)^k$	6	0.0471849442	3.0000278075	0.0000278075
1	5	6	$1+1/k$	4	0.0476438999	5.0000020572	0.0000020572
2	5	6	$1/k$	4	0.0291290283	5.0000000932	0.0000000932
3	5	6	1	4	0.0361688137	5.0000005473	0.0000005473
4	5	6	$2-1/k$	4	0.0354418755	5.0000013427	0.0000013427
5	5	6	$3+(-1)^k$	4	0.0356090069	5.0000078255	0.0000078255

$tol = 10^{-5}$, $k_{max} = 200$

Tabela 4.5: *Testes numéricos para* $f(x) = ((x-1)*(x+1)*(x-2)*(x+2))^2 - 2$

f	x^*	x^0	λ_k	k	T	x_k	$\lvert x_k - x^* \rvert$
1	-1	-0.5	$1 + 1/k$	4	0.0341498852	-0.9999805041	0.0000194959
2	-1	-0.5	$1/k$	4	0.0353391171	-0.9999991097	0.0000008903
3	-1	-0.5	1	4	0.0322761536	-0.9999948531	0.0000051469
4	-1	-0.5	$2 - 1/k$	4	0.0272729397	-0.9999875313	0.0000124687
5	-1	-0.5	$3 + (-1)^k$	5	0.0387141705	-0.9999928186	0.0000071814
1	1	0.5	$1 + 1/k$	4	0.0365440845	0.9999805041	0.0000194959
2	1	0.5	$1/k$	4	0.0320730209	0.9999991097	0.0000008903
3	1	0.5	1	4	0.0322299004	0.9999948531	0.0000051469
4	1	0.5	$2 - 1/k$	4	0.0347738266	0.9999875313	0.0000124687
5	1	0.5	$3 + (-1)^k$	5	0.0422041416	0.9999928186	0.0000071814
1	-2	-4	$1 + 1/k$	4	0.0539360046	-1.9999979091	0.0000020909
2	-2	-4	$1/k$	3	0.0389671326	-1.9999594772	0.0000405228
3	-2	-4	1	9	0.0693631172	-1.9999565401	0.0000434599
4	-2	-4	$2 - 1/k$	13	0.0829291344	-1.9999570205	0.0000429795
5	-2	-4	$3 + (-1)^k$	4	0.0439000130	-1.9999918094	0.0000081906
1	2	5	$1 + 1/k$	4	0.0464608669	1.9999970071	0.0000029929
2	2	5	$1/k$	6	0.0527169704	2.0000432298	0.0000432298
3	2	5	1	12	0.0872209072	2.0000464313	0.0000464313
4	2	5	$2 - 1/k$	4	0.0455379486	1.9999979852	0.0000020148
5	2	5	$3 + (-1)^k$	4	0.0474781990	1.9999883551	0.0000116449