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

Um Método Interior Para Minimização Sob o Ortante Não Negativo.

Leonardo Araújo de Sousa

Teresina - 2013

Leonardo Araújo de Sousa

Dissertação de Mestrado:

Um Método Interior Para Minimização Sob o Ortante Não Negativo.

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

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

Teresina - 2013

Sousa, L. A. de

Um Método Interior Para Minimização Sob o Ortante Não Negativo.

Leonardo Araújo de Sousa – Teresina: 2013.

Orientadora: Profª. Drª. Sissy da Silva Souza.

1. Otimização.
2. Matemática Aplicada.

CDD 516.36

Aos meus pais Manoel Eduardo e Ana Leozina.

Aos meus irmãos Dudu e Marina.

Ao meu bisavô Luiz Piripiri (In memoriam).

Ao meu avô João Vieira de Matos (In memoriam).

À minha família e aos amigos.

Agradecimentos

Agradeço ao meu pai guerreiro e trabalhador, Manoel Eduardo de Sousa Filho, e à minha supermãe coruja, abelhuda e superprotetora, Ana Leozina Araújo de Sousa, pelo amor exagerado e o apoio em todos os momentos da minha vida. Agradeço ao meu irmão, Manoel Eduardo de Sousa Neto (Dudu), meu braço esquerdo (porque ele é canhoto!), que sempre esteve ao meu lado, pela presença, as conversas e a companhia em todos esses anos, e à minha irmã chata, Marina Araújo de Sousa (Nina), pelas brigas desnecessárias e por sua alegria.

Agradeço à vovó Francisca, à Dagraça (Maria Moiáda!), à dona Maria (outra Moiáda!), ao Manoel Eduardo de Sousa (Manelão), ao tio Manel (in memorian), à Mãe-Maria (in memorian) e ao Luiz Gonzaga de Sousa (Seu Piripiri) (in memorian).

Agradeço à vovó Dometília, professora de matemática; ao João Vieira de Matos (Seu Matos) (in memorian); à tia Luana e ao Kim; aos tios Disney, Davis e Dewey; aos primos Erick, Lauro, Lucas e João Vitor; às queridas primas Huana Sarah, Lara, Ana Liz e Geovana.

Agradeço aos amigos de infância Augusto, Tácito, Tháila e Roque pelas partidas de Age of Mythology, Counter Strike e World of Warcraft, pelas peladas na quadra, pelos domingos na pizzaria e pelas conversas. Agradeço aos amigos do WoW, em particular, Ruan, Phylipe, George, Alvaro, Yago e Willian.

Agradeço também à Ana Paula, ao Fernando (Português) e à Elizângela, pelas comidas portuguesas, pelo natal que passamos em sua casa, e a pela companhia em casa e no aniversário.

Agradeço aos colegas do Colégio Certo: Darlison (Dadá), Hellan, Mariellany, Guilherme, Adelmárcio (Dedé) e Samille.

Agradeço aos colegas do curso de Ciência da Computação: Darlison (de novo o Dadá!), Evandro, Pedro, Thiago, João, Walyson, Rudeíza, Diego, Rafael, Fabbio, Tito, Batista Junior, Diego

Cordeiro e Ferreirinha; aos colegas da UFPI, pela presença e pelos estudos, no mestrado: Gilson, Samara, Valdir, Ailton, Edvalter, Félipe, Rosa, Israel, Joel, Renata, Jefferson, Yuri, Valdinês, Sandoel, Weslay, Thiago Esteves e Regina, Rui, Carlos Adriano, Lucas Vidal, Alex, Diego, Mykael e Franciane; na graduação: Giovana Santos, Ellen, Elilenice, Ramon, Liana Mara, Aquiles, Rafael, Ricardo, Paulo Érison, Whanderson Bruno, Mariane, Jordana, Thiago Henrique, José Carlos, Andrelino, Lucas (Quaresma, Cassiano e Viana!), Roberto, Raimundo, Alexson, Kleiton, Éder, Gideône e Milena, que organizou uma excelente aula da saudade.

Agradeço aos professores de matemática: Zaigla, Renan, Marcos Vinicius, Newton, Barnabé, Paulo Alexandre, Alexandro Marinho e principalmente Mário, Sissy e Paulinho, pelo grande apoio no começo do curso. Agradeço também aos grandes professores que já tive: a professora de história Fabíola Lemos, a professora de algoritmos e programação Rosianni, o professor de física Renato Germano e os professores de inglês, Donald e Gabriela.

Agradeço aos colegas do IMPA: Mariela Penton, Gisele Teixeira, Ítalo Dowell, Gleison, Edileno, Rafael, Victor Nolasco, José Vitor, Davi, Hudson, Felippe, Maurício, Mateus e Alan. Os colegas do IELTS: Hugo, Diego, Thais, Yago e Fernanda.

Agradeço aos colegas da 7º série do Colégio Lourenço Filho: Daniel, Willame, Mário Sérgio, Athaysi, Iaponira, Luíza, Iuri, Carlos Roger, Diego, Ricardo e Rebeca Gaspar (a desenhista!). Foi incrível encontrá-los novamente.

Agradeço a minha professora orientadora Sissy da Silva Souza e ao professor Paulo Sérgio Marques dos Santos (Paulinho), à ela pela orientação desta dissertação, à ambos pelas conversas, pelos incentivos, pela paciência e pelo apoio em Parnaíba.

Agradeço aos professores João Xavier da Cruz Neto e Susana Sheimberg de Makler pelas críticas e surgestões de aperfeiçoamento deste trabalho.

Agradeço a CAPES pelo apoio financeiro.

"You grew up good and wonderful. Then the time come for you to be your own man and take on the world and you did, but somewhere along the line, you changed. You stopped being you. You let people stick a finger in your face and tell you you're no good, and when things got hard, you started looking for something to blame... like a big shadow. Let me tell you something you already know:

The world ain't all sunshine and rainbows.

It's a very mean and nasty place and I don't care how tough you are... it will beat you to your knees and keep you there permanently if you let it... You, me or nobody is gonna hit as hard as life. But it ain't about how hard you hit, it's about how hard you can get hit and keep moving forward. How much you can take and keep moving forward. That's how winning is done!

If you know what you're worth, go out and get what you're worth. But you gotta be willing to take the hits, and not pointing fingers, saying you ain't where you wanna be because of him or her or anybody! Cowards do that and that ain't you. You're better than that!

I'm aways gonna love you no matter what, no matter what happens, but until you start believing in yourself, you ain't gonna have a life".

Trecho do filme Rocky Balboa.

Resumo

Analisamos dois algoritmos para resolver o problema de minimizar uma função sob a restrição de não-negatividade. O algoritmo geral estudado é do tipo interior-proximal, cujo núcleo é dado por uma métrica variável que depende de um parâmetro r e do último ponto gerado pelo algoritmo. No primeiro método, supomos que o gradiente da função objetivo é L-lipschitziano e o parâmetro de regularização é definido dependendo de r, aqui escolhido com valor igual ou maior que 1, e da constante L. No segundo algoritmo, requeremos que o parâmetro r seja escolhido com valor igual ou maior que 2. Em ambos os casos, supomos que a função objetivo seja convexa e mostramos que os algoritmos geram sequências bem definidas, que convergem subsequencialmente para o conjunto solução do problema. Também estudamos a taxa de convergência de ambos os algoritmos. Apresentamos, finalmente, alguns exemplos numéricos ilustrativos da aplicação dos algoritmos para funções quadráticas convexas.

Palavras-chave: Método Proximal, Algoritmo de Ponto Interior, Problema Convexo.

Abstract

We analyse two algorithms for solving the problem of minimizing a function under nonnegative constraints. The general algorithm studied is a Interior-proximal, whose kernel is a variable metric that depends on a parameter r and the last point generated by the algorithm. In the first method, we suppose that the gradient of the objective function be L-lipschitzian and the regularization parameter is defined depending on r, here choosed with initial value greater or equal to one, and the L constant. In the second algorithm, we require that the parameter r be choosed with initial value greater or equal to 2. In both cases, we suppose that the objective function be convex and we show that the algorithms generate well defined sequences, which converge subsequentialy to feasible set of the problem. We also study the convergence rate of both algorithms. We show, finally, some numerical and illustratives examples of aplication of algorithms to convex quadratic functions.

Keywords: Proximal Methods, Interior-point Algorithms, Convex Problem.

Sumário

Introdução

Em alguns de seus primeiros trabalhos de análise convexa, em 1962 e 1965, Moreau iniciou a noção de aproximação proximal de uma função [27, 28]. Baseado nestes trabalhos, em 1970, Martinet [26] introduziu o algoritmo do ponto proximal para minimizar uma função convexa $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. Este algoritmo foi desenvolvido em 1976 por Rockafellar [32], para encontrar zeros de operadores monótonos maximais. Eggermont [11] trabalhou com algoritmos multiplicativos iterativos para a minimização de funções convexas diferenciáveis definidas no ortante positivo, ele mostrou que se a função objetivo tiver conjuntos de nível compactos e tiver um gradiente localmente Lipschitz contínuo, então os algoritmos convergem para uma solução do problema de minimização. Casos especiais desses algoritmos foram aplicados na tomografia por emissão de pósitrons.

Censor e Zenios [5], substituíram no algoritmo do ponto proximal, os termos de regularização quadrática por D-funções, que se assemelham a funções distância. Caracterizaram as propriedades de tais D-funções que, quando usadas no algoritmo do ponto proximal, preserva sua convergência global. Esse algoritmo foi profundamente estudado por Chen e Teboulle [6, 36, 37], que fizeram uma prova alternativa da convergência do algoritmo usando funções de Bregman. A análise permite o estabelecimento de uma taxa de convergência global do algoritmo expressa em termos dos valores da função.

Iusem, Svaiter e Teboulle [18] estudaram uma extensão do método do ponto proximal para programação convexa, e substituíram o núcleo de regularização quadrática por uma classe de distâncias estatísticas convexas, chamadas de φ-divergência. Iusem [15] estudou um método de ponto interior multiplicativo para problemas com restrições, onde o passo iterativo depende de um parâmetro encontrado através de busca linear. Bonnans et al. [3] combinaram a regularização de Moreau-Yoshida na análise convexa e aproximação quase Newton de funções suaves.

Kiwiell [19], fez um trabalho em que considera métodos para minimizar funções convexas, usando o algoritmo para gerar uma sequência usando distâncias de Bregman. São feitas extensões para B-funções que generalizam funções de Bregman e cobrem mais aplicações. São feitas aplicações para métodos de multiplicadores não quadráticos para programas não lineares. Em 1997, Eckstein [10] estabelece a convergência dos algoritmos de ponto proximal baseados em funções de Bregman generaliadas quando a iteração é calculada apenas aproximadamente.

Burachik e Iusem [4], em 1998, estudaram método do ponto proximal generalizado para resolver problemas de desigualdades variacionais com operadores monótonos num espaço de Hilbert. A diferença entre o método do ponto proximal clássico também é pelo uso de distâncias de Bregman, essa distância força a sequência gerada pelo método a permanecer no interior do conjunto viável, assim o método se torna de ponto interior. Eles estenderam resultados similares para o método do ponto proximal com distâncias de Bregman que lidava apenas com o caso de dimensão finita e que se aplicava apenas a problemas de otimização convexa ou para encontrar zeros de operadores monótonos, que são casos particulares de problemas de desigualdade variacional.

Auslender, Teboulle e Ben-Tiba [1, 2], estudaram uma classe de algoritmos de ponto interior e métodos de multiplicadores não quadráticos para resolver problemas de programação convexo, o termo quadrático usual foi substituído por um funcional homogêneo de ordem dois, definido em termos da função convexa.

Outros estudos sobre o método do ponto proximal pode ser encontrado em [17, 29, 31, 34, 35].

O método interior para minimização sobre o ortante não negativo consiste em um problema de minimização de uma função $f : \mathbb{R}^n \to \mathbb{R}$, onde o conjunto viável é o conjunto dos $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ tais que $x_i \geq 0$, $\forall i = 1, \cdots, n$. Isto é,

$$(P) \quad \min f(x) \quad \text{sujeito} \quad \text{a} \quad x \geq 0. \tag{1}$$

Existem vários métodos de solução para o problema (P).

Esta dissertação está baseada nas referências [29] e [34] e focamos em algoritmos do ponto proximal. Este trabalho está dividido da seguinte maneira: no Capítulo 1 comentamos sobre as notações utilizadas, em seguida introduzimos alguns conceitos básicos de Álgebra Linear, Análise no $\mathbb{R}^n$ e Otimização, com destaque para a parte de condições de otimalidade e tipos de convergência. No capítulo 2 comentamos sobre o método do ponto proximal clássico, onde enunciamos e demonstramos a convergência do método. No capítulo 3 destacamos os algoritmos que dão título a esta dissertação, método de ponto interior para um ortante não negativo, onde apresentamos a definição dos algoritmos, provamos a existência, analisamos a convergência e estudamos taxa de convergência e, finalmente, a relação entre os valores de aderência da sequência e o problema (P). Encerramos esta dissertação no capítulo 5, com a implementação deste algoritmo, para diferentes tipos de funções e diferentes valores de r, onde usamos exemplos de funções quadráticas e o software *Scilab*.

Capítulo 1

Resultados Preliminares

Nesta seção, apresentamos as definições e resultados básicos que serão úteis no desenvolvimento desta dissertação. Iniciamos com tópicos de Álgebra Linear e Análise, onde exibimos algumas definições e propriedades de matrizes utilizadas. Em seguida, fazemos uma breve seção de introdução à Otimização, com destaque para convexidade de funções, condições de otimalidade e tipos de convergência.

1.1 Tópicos de Álgebra Linear e Análise no Espaço $\mathbb{R}^n$

Recordamos algumas definições e teoremas de Álgebra Linear a fim de tornar essa dissertação autocontida. Mais detalhes podem ser encontrados em [20], nossa referência para esta seção. Para a seção de Análise no espaço $\mathbb{R}^n$, podemos encontrar mais detalhes em [21], [22] ou [23].

1.1.1 Matrizes, Espaços e Subespaços Vetoriais.

Um *espaço vetorial* E é um conjunto, onde os elementos são chamados vetores, no qual estão definidas duas operações: a adição, que faz corresponder cada par de vetores $x, y \in E$ em um novo vetor $x + y \in E$, chamado soma de x e y, e a multiplicação por um número real, que a cada número real $\alpha \in \mathbb{R}$ e a cada vetor $y \in E$ faz corresponder um vetor αy, chamado de produto de α por y. Essas operações satisfazem, para quaisquer $\alpha, \beta \in \mathbb{R}$ e $x, y, z \in E$, as seguintes propriedades:

1. $x + y = y + x$;

2. $(x+y)+z = x+(y+z)$ e $(\alpha\beta)y = \alpha(\beta y)$;

3. existe um vetor $0 \in E$, tal que $y+0 = 0+y = y$;

4. para todo vetor $y \in E$, existe um vetor $-y \in E$ tal que $y+(-y) = (-y)+y = 0$;

5. $(\alpha+\beta)y = \alpha y + \beta y$ e $\alpha(x+y) = \alpha x + \alpha y$;

6. $1 \cdot y = y$;

Consideramos o espaço euclideano $\mathbb{R}^n$ com as operações usuais de soma e multiplicação por escalar definidas por

$$x+y = (x_1+y_1, \cdots, x_n+y_n) \quad \text{e} \quad \alpha x = (\alpha x_1, \cdots, \alpha x_n).$$

Por definição, a igualdade vetorial $x = y$ significa que são iguais cada coordenada, isto é $x_1 = y_1, \cdots, x_n = y_n$. Definimos também o vetor nulo, o vetor onde suas coordenadas são todas iguais a zero, isto é $0 = (0, \cdots, 0) \in \mathbb{R}^n$.
O inverso aditivo de $x = (x_1, \cdots, x_n)$ é $-x = (-x_1, \cdots, -x_n)$.
Estas definições tornam o $\mathbb{R}^n$ um espaço vetorial.

Uma *matriz* real de ordem $m \times n$ $A = [a_{ij}]$ é uma lista de números reais a_{ij} com índices duplos, onde $1 \leq i \leq m$ e $1 \leq j \leq n$. Representamos a matriz A como um quadro numérico com m linhas e n colunas, no qual o elemento a_{ij} situa-se no cruzamento da i-ésima linha com a j-ésima coluna:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}_{m \times n}.$$

O vetor $(a_{i1}, \cdots, a_{in}) \in \mathbb{R}^n$ é o i-ésimo vetor-linha da matriz A e o vetor $(a_{1j}, \cdots, a_{nj}) \in \mathbb{R}^n$ é o j-ésimo vetor coluna de A. O conjunto de todas as matrizes $m \times n$ torna-se um espaço vetorial quando nele se define a soma das matrizes $A = [a_{ij}]$ e $B = [b_{ij}]$ como $A + B = [a_{ij} + b_{ij}]$ e o produto da matriz A pelo número real α como $\alpha A = [\alpha a_{ij}]$. A matriz nula é aquela formada por zeros e o inverso aditivo da matriz $A = [a_{ij}]$ é $-A = [-a_{ij}]$. Denotaremos este conjunto por $\mathbb{R}(m \times n)$. Quando $m = n$ diz-se que a matriz é *quadrada*. A lista composta pelos elementos a_{ii}, $1 \leq i \leq n$ de uma matriz quadrada $A \in \mathbb{R}(n \times n)$ é chamada de diagonal principal da matriz.

A matriz que possui zeros em todos os termos que não estão na diagonal principal é chamada de *matriz diagonal*. Uma matriz quadrada $A = [a_{ij}]$ é chamada *simétrica* quando $a_{ij} = a_{ji}$. A *transposta* de A, denotada por A^t é a matriz que tem como linhas as colunas de A e como colunas as linhas de A. Observe que $A^t \in \mathbb{R}(n \times m)$.

Exemplo 1.1.1 *Para uma matriz diagonal D, temos que se*

$$D = \begin{pmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{mn} \end{pmatrix}$$

A matriz transposta de D será

$$D^t = \begin{pmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{mn} \end{pmatrix}.$$

1.1.2 Bases, Transformações Lineares e Produto Interno.

Uma *base* de um espaço vetorial E é um conjunto $U \subset E$ linearmente independente que gera E. Em outras palavras, todo vetor $x \in E$ pode ser escrito de modo único como combinação linear $x = \alpha_1 u_1 + \cdots + \alpha_n u_n$ de elementos $u_1, \cdots, u_n$ da base U.

Uma *transformação linear* $T : E \to F$ é uma correspondência que associa a cada vetor $x \in E$ em um vetor $Tx = T(x) = T \cdot x \in F$ de modo que valham as seguintes propriedades, para quaisquer $x, y \in E$ e $\alpha \in \mathbb{R}$:

$$T(x + y) = T(x) + T(y) \quad \text{e} \quad T(\alpha x) = \alpha T(x).$$

Quando a transformação linear vai de um espaço nele mesmo, i. e., $T : E \to E$ chamamos de *operador linear* e quando $T : E \to \mathbb{R}$ chamamos de *funcional linear*.

Uma transformação linear $A : \mathbb{R}^n \to \mathbb{R}^m$ fica inteiramente determinada por uma matriz $A = [a_{ij}] \in \mathbb{R}^{(m \times n)}$, os vetores-coluna dessa matriz são imagens $A(e_j)$ dos vetores da base canônica de $\mathbb{R}^n$.

A seguinte definição é apresentada para recordarmos matrizes invertíveis, que serão usadas na determinação da Taxa de Convergência da última seção.

Uma transformação linear $T : E \to F$ é *invertível* quando existe $S : F \to E$ linear tal que $TS = I_F$ e $ST = I_E$, onde I_E é a identidade em E. Dizemos que S é a inversa de T e escrevemos $S = T^{-1}$. Vimos que toda transformação linear de $\mathbb{R}^n$ em $\mathbb{R}^m$ pode ser representada por uma matriz de ordem $m \times n$ e vice-versa, assim também definimos uma matriz invertível, só que neste caso I_E será a matriz identidade referente ao espaço.

Seja E um espaço vetorial. Um *produto interno* é uma função $\langle \cdot, \cdot \rangle : E \times E \to \mathbb{R}$, que associa a cada par de vetores $x, y \in E$ um número real $\langle x, y \rangle$, chamado de produto interno de x por y, de modo que valem as seguintes propriedades, para quaisquer $x, \tilde{x}, y, \tilde{y} \in E$ e $\alpha \in \mathbb{R}$:
Bilinearidade:

$$\langle x + \alpha\tilde{x}, y \rangle = \langle x, y \rangle + \alpha \langle \tilde{x}, y \rangle ,$$

$$\langle x, y + \alpha\tilde{y} \rangle = \langle x, y \rangle + \alpha \langle x, \tilde{y} \rangle ;$$

Simetria:

$$\langle x, y \rangle = \langle y, x \rangle$$

Positividade:

$$\langle x, x \rangle > 0, \quad \text{se} \quad x \neq 0.$$

Nesta dissertação consideramos apenas o espaço vetorial $\mathbb{R}^n$ e o corpo $\mathbb{R}$. Assim, se $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ e $y = (y_1, \cdots, y_n) \in \mathbb{R}^n$, o produto interno que usamos é definido por $\langle x, y \rangle = x_1 y_1 + \cdots + x_n y_n$.

Observe que se representarmos os vetores $x, y \in \mathbb{R}^n$ por uma matriz, teremos:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{e} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Assim, o produto interno euclidiano será equivalente a:

$$\langle x, y \rangle = y^t x = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix} \times \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 y_1 + \cdots + x_n y_n,$$

onde $y^t x$ é o produto de matrizes usual.

1.1.3 Norma e a Adjunta.

Seja E um espaço vetorial. Uma *norma* é uma função $\|\cdot\| : E \to \mathbb{R}$ que associa o vetor $x \in E$ a um número real $\|x\|$, chamado de norma de x, de modo que valem as seguintes propriedades, $\forall x, y \in E$ e $\forall \alpha \in \mathbb{R}$:

1. $\|x\| \geq 0$ e $\|x\| = 0$ se, e só se, $x = 0$;

2. $\|\alpha x\| = |\alpha| \, \|x\|$;

3. $\|x + y\| \leq \|x\| + \|y\|$ (desigualdade triangular).

Se $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$, então $\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$. Por definição, temos que $\langle x, x \rangle = \|x\|^2$.

Quando $\|x\| = 1$, dizemos que x é um vetor unitário. Observe que para todo $x \neq 0$, o vetor $x / \|x\|$ é unitário.

Seja E um espaço vetorial com produto interno. Dois vetores $x, y \in E$ são chamados de *ortogonais* quando $\langle x, y \rangle = 0$. Denotamos dois vetores ortogonais por $x \perp y$. Note que 0 é ortogonal a qualquer vetor de E. Um conjunto $Y \subset E$ é chamado de ortogonal, quando dois vetores distintos quaisquer em Y são ortogonais. Mais ainda, se todos os vetores de Y são unitários, então este conjunto é chamado de conjunto *ortonormal*. Uma base ortonormal é uma base de E que é um conjunto ortonormal.

Teorema 1.1.2 (Teorema de Pitágoras) *Se $x \perp y$, então $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.*

Demonstração: $\|x + y\|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + 2 \langle x, y \rangle + \langle y, y \rangle = \langle x, x \rangle + \langle y, y \rangle = \|x\|^2 + \|y\|^2$. ■

Teorema 1.1.3 (Desigualdade de Cauchy-Schwarz) *Para quaisquer vetores $x, y \in \mathbb{R}^n$, temos $|\langle x, y \rangle| \leq \|x\| \, \|y\|$. A igualdade ocorre quando, um dos vetores x, y é um múltiplo do outro.*

Demonstração: É trivial o caso $y = 0$. Supomos que $y \neq 0$, tome $\alpha = \langle x, y \rangle / \|y\|^2$ e observe que o vetor $z = x - \alpha y$ é ortogonal a y, $\forall x, y \in \mathbb{R}^n$, com $y \neq 0$. De fato, $\langle z, y \rangle = \langle x - \alpha y, y \rangle = \langle x, y \rangle - \alpha \langle y, y \rangle = \langle x, y \rangle - \alpha \|y\|^2 = \langle x, y \rangle - \langle x, y \rangle = 0$. Usando a definição de norma e o Teorema de Pitágoras 1.1.2, temos

$$\|x\|^2 = \langle z + \alpha y, z + \alpha y \rangle = \|z\|^2 + \alpha^2 \|y\|^2$$

implicando que

$$\|x\|^2 \geq \alpha^2 \|y\|^2 = \langle x, y\rangle^2 / \|y\|^2, \quad \text{ou seja,}$$

$\|x\|^2 \|y\|^2 \geq \langle x, y\rangle^2$. A igualdade acontece se, e somente se, $z = 0$, isto é, $x = \alpha y$. ■

A *adjunta* de A é uma transformação linear $A^* : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ tal que para todo $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^m$ quaisquer, satisfaz

$$\langle Ax, y\rangle = \langle x, A^* y\rangle .$$

Mostraremos agora que a adjunta preserva a soma e a multiplicação por um escalar. Sejam $y, z \in \mathbb{R}^m$ e $\alpha \in \mathbb{R}$, daí

$$\begin{aligned} \langle x, A^*(y + \alpha z)\rangle &= \langle Ax, y + \alpha z\rangle \\ &= \langle Ax, y\rangle + \alpha \langle Ax, z\rangle \\ &= \langle x, A^* y\rangle + \alpha \langle x, A^* z\rangle \\ &= \langle x, A^* y + \alpha A^* z\rangle . \end{aligned} \tag{1.1}$$

Como isso ocorre para todo $x \in \mathbb{R}^n$, podemos concluir que

$$A^*(y + z) = A^*(y) + A^*(z).$$

As seguintes propriedades de matrizes adjuntas podem ser provadas de maneira análoga. Sejam A e B transformações lineares e α um escalar, temos

1. $I^* = I$;

2. $(\alpha A)^* = \alpha A^*$;

3. $(A + B)^* = A^* + B^*$;

4. $(AB)^* = B^* A^*$;

5. $A^{**} = A$.

Quando a transformação linear A satisfaz $A = A^*$, dizemos que ela é *autoadjunta*, quando ela satisfaz $AA^* = A^* A$, dizemos que a transformação linear é *normal* e se $AA^* = A^* A = I$, ela é chamada de *ortonormal*.

Usaremos matrizes autoadjuntas no desenvolvimento do trabalho e para tanto o próximo resultado será bastante últil.

As propriedades da adjunta de uma transformação linear se traduzem para propriedades da transposta de uma matriz, isto é, se $A, B \in \mathbb{R}(m \times n)$ e $\alpha \in \mathbb{R}$, temos

1. $I^t = I$;
2. $(\alpha A)^t = \alpha A^t$;
3. $(A + B)^t = A^t + B^t$;
4. $(AB)^t = B^t A^t$;
5. $(A^t)^t = A$.

1.1.4 Autovetores e Autovalores.

Um vetor $x \neq 0$ em E é chamado de *autovetor* do operador $A : E \to E$ quando existe um $\lambda \in \mathbb{R}$ tal que

$$Ax = \lambda x.$$

O número $\lambda \in \mathbb{R}$, por sua vez, é chamado de *autovalor* do operador A quando existe um vetor não nulo $x \in E$ tal que $Ax = \lambda x$. Diz-se então que o autovalor λ corresponde ao autovetor x e vice-versa, que o autovetor x também corresponde ao autovalor λ.

Analogamente, diz-se que o número $\lambda \in \mathbb{R}$ é um autovalor da matriz $\mathbf{A} \in \mathbb{R}(n \times n)$ quando λ é um autovalor do operador $A : \mathbb{R}^n \to \mathbb{R}^n$, cuja matriz na base canônica é $\mathbf{A}$.

Teorema 1.1.4 (Teorema Espectral) *Para todo operador autoadjunto $T : E \to E$, num espaço vetorial de dimensão finita munido de um produto interno, existe uma base ortonormal $\{e_1, \cdots, e_n\} \subset E$ gerada por autovetores de T.*

Demonstração: Veja a referência [20]. ■

Observe que se $A : \mathbb{R}^n \to \mathbb{R}^n$ é um operador autoadjunto, então o Teorema Espectral nos garante que $\mathbb{R}^n$ possui uma base $\{x_1, \cdots, x_n\}$ formada por autovetores, assim um vetor $x \in \mathbb{R}^n$ pode ser escrito como combinação linear dos autovetores dessa base, isto é, $x = \alpha_1 x_1 + \cdots + \alpha_n x_n$.

A seguir definimos a norma matricial que será utilizada neste trabalho.

Sejam $x \in \mathbb{R}^n$ e $A \in \mathbb{R}(n \times n)$ quaisquer. Definimos a norma de x em relação a matriz A por

$$\|x\|_A = \langle Ax, x \rangle^{1/2},$$

ou equivalentemente

$$\|x\|_A = x^t Ax.$$

Uma matriz $M \in \mathbb{R}(n \times n)$ é positiva definida quando $x^t Mx > 0$, para todo $x \neq 0$ em $\mathbb{R}^n$.

Exemplo 1.1.5 *Um exemplo trivial de matriz positiva definida é a seguinte matriz.* $M = [m_{ij}] \in \mathbb{R}(2 \times 2)$ *é definida por*

$$M = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}, \quad com \quad m_1, m_2 > 0.$$

Dado $x = (x_1, x_2) \in \mathbb{R}^2 - \{(0,0)\}$, *temos que*

$$x^t Mx = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \times \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix} \times \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = m_1 x_1^2 + m_2 x_2^2 > 0.$$

Como x *é arbitrário, segue que* M *é positiva definida.*

Exemplo 1.1.6 *Seja* $A \in \mathbb{R}(n \times n)$ *uma matriz simétrica. O Teorema Espectral garante que existe uma base formada por autovetores de* A*. Seja* $x_1, \cdots, x_n$ *uma base de autovetores de* A*. Definindo a matriz* $Q = [x_1, \cdots, x_n]$*, isto é, cada linha de* Q *é um autovetor de* A*. Observe que* $Q^t Q = T$ *e portanto* $Q^t = Q^{-1}$*. Observe também que*

$$Q^{-1}AQ = Q^t AQ = Q^t[Ax_1, \cdots, Ax_n] = Q^t[\lambda_1 x_1, \cdots, \lambda_n x_n].$$

Portanto

$$Q^{-1}AQ = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}.$$

e portanto A *é semelhante a uma matriz diagonal.*

Uma conexão entre a definição da matriz positiva definida com os autovalores de A é feita da seguinte maneira: Para cada $x \in \mathbb{R}^n$ fazemos $y = Q^{-1}x$, onde Q é definida acima. Então

$$x^t Ax = y^t Q^t AQy = \sum_{i=1}^{n} \lambda_i y_i^2.$$

Como os y_i's são arbitrários, observamos que A é positiva definida se, e só se, todos os autovalores de A são positivos.

Mais detalhes podem ser encontrados em [24].

Finalizamos essa seção apresentando dois resultados que serão úteis em nosso estudo de taxa de convergência.

Uma matriz quadrada é singular quando não admite uma inversa, ou seja, quando o seu determinante é nulo. Uma matriz quadrada é não singular quando o seu determinante é não nulo (portanto admite inversa).

Proposição 1.1.7 *Se $A, B \in \mathbb{R}(n \times n)$ são simétricas positivas definidas, então existem $P, D \in \mathbb{R}(n \times n)$ (P não singular e D diagonal positiva) tais que $AB = PDP^{-1}$.*

Demonstração: Veja a referência [13]. $\blacksquare$

Aqui a norma euclidiana será denotada por $\|\cdot\|_2$.

Proposição 1.1.8 *Seja $M \in \mathbb{R}(n \times n)$ uma matriz não singular e considere a norma $\|x\| = \|M^{-1}x\|_2$. Se A, B são simétricas positivas definidas, $AB = PDP^{-1}$, com D diagonal positiva, e β um número real positivo, então*
(i) $\left\|(I + \frac{1}{\beta}AB)^{-1}\right\| \leq \|M^{-1}P\|_2 \, \|P^{-1}M\|_2 \, (1 + \frac{\lambda\xi}{\beta})^{-1}$
(ii) $\|(\beta A^{-1} + B)^{-1}\| \leq \|M^{-1}\|_2 \, \|M\|_2 \, \xi^{-1}$,
onde λ e ξ são os menores autovalores de A e B respectivamente.

Demonstração: Veja a referência [16]. $\blacksquare$

1.1.5 Sequências em $\mathbb{R}^n$

As seguintes definições podem ser encontradas em [21], [22] ou [23].

Uma *sequência* em $\mathbb{R}^n$ é uma função $x : \mathbb{N} \longrightarrow \mathbb{R}^n$ que associa a cada número natural k um ponto $x^k \in \mathbb{R}^n$. A notação para sequência que usaremos neste trabalho será $\{x^k\}$, onde $x^k = (x_1^k, x_2^k, \cdots, x_n^k) \in \mathbb{R}^n$.

Uma *subsequência* de $\{x^k\}$ é a restrição desta sequência a um subconjunto infinito $\mathbb{N}' = \{k_1 < \cdots < k_m < \cdots\} \subset \mathbb{N}$. A notação $\{x^{k_j}\}$ é usada pra indicar uma subsequência de $\{x^k\}$.

Dizemos que um ponto $a \in \mathbb{R}^n$ é um *valor de aderência* da sequência $\{x^k\}$ quando existe uma subsequência de $\{x^k\}$ que converge para a, isto é, quando existe uma subsequência $\{x^{k_j}\}$ tal que $\lim_{j \to +\infty} x^{k_j} = a$.

Dizemos que um conjunto $K \subset \mathbb{R}^n$ é compacto quando é limitado e fechado.

O seguinte teorema será útil na parte de existência e resultados preliminares.

Teorema 1.1.9 (Weierstrass) *Seja $K \subset \mathbb{R}^n$ compacto. Se $f : K \to \mathbb{R}$ é uma função contínua, então existem $a, b \in K$ tais que $f(a) \leq f(x) \leq f(b)$, para todo $x \in K$.*

Demonstração: Veja a referência [21]. ■

1.2 Introdução à Otimização.

Nesta seção apresentamos algumas definições e resultados básicos em otimização. As principais referências para esta seção foram [14] e [24].

Sejam $C \subset \mathbb{R}^n$ e $D \subset \mathbb{R}^n$ tais que $D \subset C$ e uma função $f : C \longrightarrow \mathbb{R}$. O problema de otimização a ser considerado é o de encontrarmos um minimizador de f no conjunto D. Isto é,

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

onde D é chamado de conjunto viável, os pontos de D são chamados de pontos viáveis e f de função objetivo. Também podemos falar em problema de maximização

$$\max f(x) \quad \text{sujeito a} \quad x \in D,$$

mas este problema pode ser transformado em um problema de minimização equivalente:

$$\min -f(x) \quad \text{sujeito a} \quad x \in D.$$

As soluções locais e globais desses dois problemas são as mesmas, com os sinais opostos para os valores ótimos. Assim, do ponto de vista matemático, não existe diferença entre os problemas de minimização e maximização. Podemos estender os resultados obtidos de uma classe de problemas para outra sem muita dificuldade. Nesta dissertação consideramos apenas os problemas de minimização.

Definição 1.2.1 *Dizemos que um ponto $\bar{x} \in D$ é um minimizador local de* (1.2)*, se existe uma vizinhança V de $\bar{x}$ tal que*

$$f(\bar{x}) \leq f(x), \quad \forall x \in D \cap V.$$

O ponto $\bar{x} \in D$ será um minimizador global de (1.2) *se*

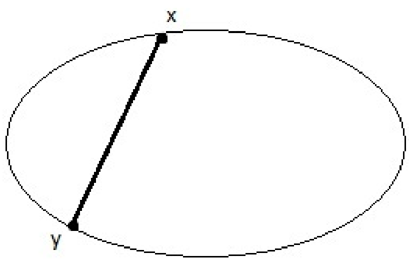
$$f(\bar{x}) \leq f(x), \quad \forall x \in D.$$

Se para todo $x \neq \bar{x}$ as desigualdades acimas são estritas, então $\bar{x}$ será chamado de minimizador local estrito ou minimizador global estrito.

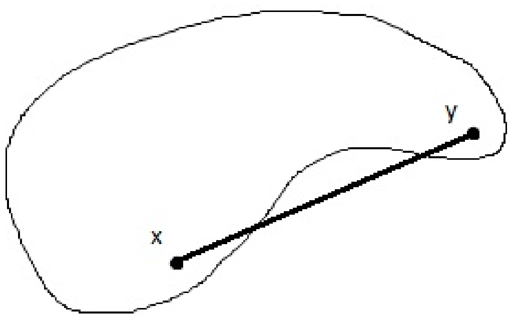
Agora comentaremos alguns conceitos básicos de análise convexa necessários para esta dissertação.

1.2.1 Funções Convexas.

Definição 1.2.2 *Dizemos que um conjunto $C \subset \mathbb{R}^n$ é convexo quando o segmento de reta que une quaisquer dois pontos de C ainda pertence a este conjunto. Em outras palavras, quando dados $x, y \in C$ e $t \in [0, 1]$ tivermos que $(1-t)x + ty \in C$.*



Conjunto convexo.



Conjunto não convexo.

Definição 1.2.3 *Seja $C \subset \mathbb{R}^n$ um conjunto convexo. Dizemos que uma função $f : C \longrightarrow \mathbb{R}$ é convexa quando para todo $x, y \in C$ e para todo $t \in [0,1]$, tivermos que*

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y).$$

Quando a desigualdade acima é estrita para todo $x \neq y$ e para todo $t \in (0,1)$ dizemos que a função é estritamente convexa.

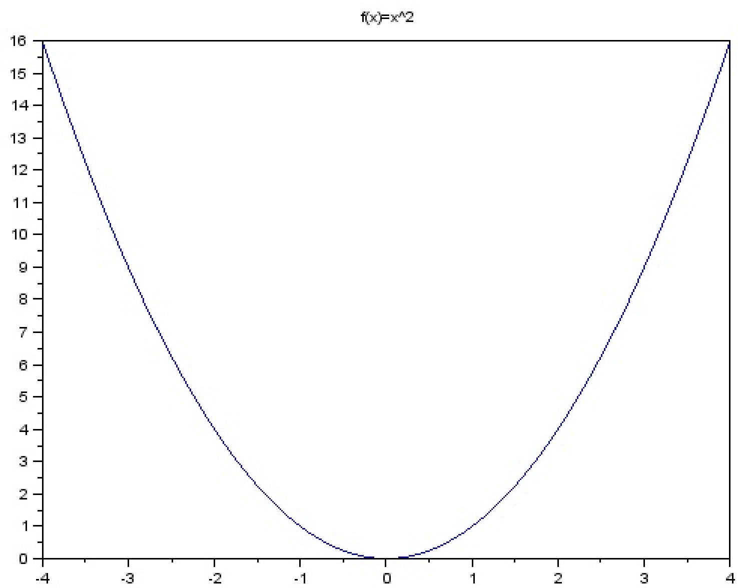
Definição 1.2.4 *Seja $C \subset \mathbb{R}^n$ um conjunto convexo. Dizemos que uma função $f : C \longrightarrow \mathbb{R}$ é fortemente convexa de módulo $\gamma > 0$, quando para quaisquer $x, y \in C$ e $t \in [0,1]$, tivermos que*

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \gamma t(1-t)\left\|x - y\right\|^2.$$

Exemplo 1.2.5 $f : \mathbb{R} \longrightarrow \mathbb{R}, x \mapsto f(x) = x^2$. *De fato $\forall x, y \in \mathbb{R}^n$ e $\forall t \in [0,1]$, temos que*

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

logo f é convexa.

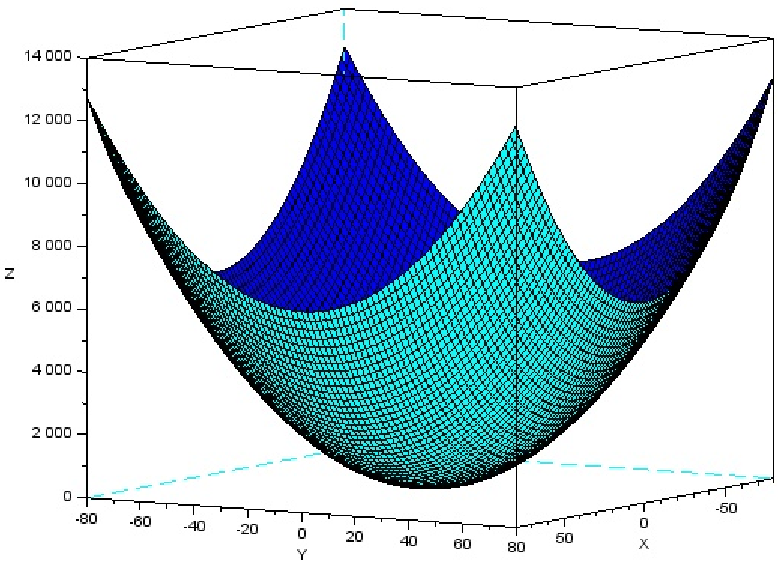


Exemplo 1.2.6 *$f : \mathbb{R}^2 \longrightarrow \mathbb{R}, x = (x_1, x_2) \mapsto f(x_1, x_2) = x_1^2 + x_2^2 = \|x\|^2$. Dados $x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R}^2$ e $t \in [0, 1]$, obtemos*

$$\begin{aligned}
f((1-t)(x_1, x_2) + t(y_1, y_2)) &= f((1-t)x_1 + ty_1, (1-t)x_2 + ty_2) \\
&= (x_1 + t(y_1 - x_1))^2 + (x_2 + t(y_2 - x_2))^2 \\
&= x_1^2 + 2tx_1(y_1 - x_1) + t^2(y_1 - x_1)^2 + x_2^2 + 2tx_2(y_2 - x_2) + t^2(y_2 - x_2)^2 \\
&\leq x_1^2 + 2tx_1(y_1 - x_1) + t(y_1 - x_1)^2 + x_2^2 + 2tx_2(y_2 - x_2) + t(y_2 - x_2)^2 \\
&= (1-t)(x_1^2 + x_2^2) + t(y_1^2 + y_2^2) \\
&= (1-t)f(x_1, x_2) + tf(y_1, y_2).
\end{aligned}$$

Como os pontos são arbitrários, segue que f é convexa. Mais ainda, f é fortemente convexa de módulo 1, pois

$$\begin{aligned}
f(tx + (1-t)y) &= \|tx + (1-t)y\|^2 = \\
&= t^2\|x\|^2 + 2t(1-t)\langle x, y\rangle + (1-t)^2\|y\|^2 = \\
&= t^2\|x\|^2 - 2(t^2 - t)\langle x, y\rangle + \|y\|^2 - 2t\|y\|^2 + t^2\|y\|^2 = \\
&= t^2\|x\|^2 - 2(t^2 - t)\langle x, y\rangle + \|y\|^2 - t\|y\|^2 - t\|y\|^2 + t^2\|y\|^2 + t\|x\|^2 - t\|x\|^2 = \\
&= t\|x\|^2 + \|y\|^2 - t\|y\|^2 + t^2\|x\|^2 - t\|x\|^2 - 2(t^2 - t)\langle x, y\rangle + t^2\|y\|^2 - t\|y\|^2 = \\
&= t\|x\|^2 + (1-t)\|y\|^2 + (t^2 - t)\|x\|^2 - 2(t^2 - t)\langle x, y\rangle + (t^2 - t)\|y\|^2 = \\
&= t\|x\|^2 + (1-t)\|y\|^2 + (t^2 - t)\left(\|x\|^2 - 2\langle x, y\rangle + \|y\|^2\right) = \\
&= t\|x\|^2 + (1-t)\|y\|^2 + (t^2 - t)\|x - y\|^2 = \\
&= tf(x) + (1-t)f(y) - t(1-t)\|x - y\|^2.
\end{aligned}$$



Teorema 1.2.7 (Convexidade do conjunto de nível de funções convexas.) *Suponhamos que o conjunto $C \subset \mathbb{R}^n$ seja convexo e a função $f : C \to \mathbb{R}$ seja convexa em C. Então o conjunto de nível*

$$l(f, \alpha) = \{x \in C; f(x) \leq \alpha\}$$

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

Demonstração: Tomamos $\alpha \in \mathbb{R}$ arbitrário. Se $l(f, \alpha) = \emptyset$, a conclusão segue, pois o conjunto vazio é convexo trivialmente. Sejam $x, y \in l(f, \alpha)$, i. e., $x \in C$, $f(x) \leq \alpha$ e $y \in C$, $f(y) \leq \alpha$. Já que C é convexo, dado $t \in [0, 1]$, $tx + (1 - t)y \in C$. Pela convexidade de f em C, temos

$$\begin{aligned} f(tx + (1-t)y) &\leq tf(x) + (1-t)f(y) \\ &\leq t\alpha + (1-t)\alpha = \alpha, \end{aligned} \tag{1.3}$$

mostrando que $(tx + (1 - t)y) \in l(f, \alpha)$. ■

Teorema 1.2.8 *Sejam $C \in \mathbb{R}^n$ um conjunto aberto e convexo e $f : C \longrightarrow \mathbb{R}$ uma função diferenciável em C. Então as seguintes propriedades são equivalentes:*

(a) A função é convexa em C.

(b) Para todos $x, y \in C$,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle .$$

(c) Para todo $x \in C$,

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0.$$

Quando f é duas vezes diferenciável em C, as propriedades acima também são equivalentes a

(d) A matriz Hessiana de f é semidefinida positiva em todo ponto de C:

$$\left\langle \nabla^2 f(x)d, d \right\rangle \geq 0 \quad \forall x \in C, \, \forall d \in \mathbb{R}^n.$$

Demonstração: Primeiramente, mostraremos que (a)$\Rightarrow$(b).

Seja f convexa. Para $x, y \in C$ e $t \in (0, 1]$ quaisquer, definindo $d = y - x$, temos que

$$\begin{aligned} f(x + td) &= f(ty + (1 - t)x) \\ &\leq tf(y) + (1 - t)f(x), \end{aligned} \tag{1.4}$$

que implica que

$$t(f(y) - f(x)) \geq f(x + td) - f(x).$$

Dividindo os dois lados da desigualdade acima por $t > 0$ e fazendo $t \to 0+$, obtemos

$$\begin{aligned} f(y) - f(x) &= \lim_{t \to 0+} \frac{f(x + td) - f(x)}{t} \\ &= \langle \nabla f(x), d \rangle = \langle \nabla f(x), y - x \rangle \,. \end{aligned} \tag{1.5}$$

(b) $\Rightarrow$(c): Usando o item (b), temos:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \quad \text{e}$$

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \,.$$

Somando as duas desigualdades acima obtemos (c).

(c) $\Rightarrow$ (b): Sejam $x, y \in C$. Pelo Teorema do Valor Médio, existe $t \in (0, 1)$ tal que

$$f(y) - f(x) = \langle \nabla f(x + t(y - x)), y - x \rangle \,. \tag{1.6}$$

Usando o item (c) para os pontos $(x + t(y - x))$ e x, temos

$$\begin{aligned} \langle \nabla f(x + t(y - x)), y - x \rangle &= t^{-1} \langle \nabla f(x + t(y - x)), t(y - x) \rangle \\ &\geq t^{-1} \langle \nabla f(x), t(y - x) \rangle \\ &= \langle \nabla f(x), y - x \rangle \,. \end{aligned} \tag{1.7}$$

Assim

$$\langle \nabla f(x + t(y - x)), y - x\rangle \geq \langle \nabla f(x), y - x\rangle .$$

Usando a equação (1.6) e a desigualdade acima, obtemos

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle .$$

(b) $\Rightarrow$ (a): Novamente, definindo $d = y - x$ e usando (b) para os pontos x e $x + td$; y e $x + td$, respectivamente, temos

$$f(x) \geq f(x + td) - t \langle \nabla f(x + td), d\rangle ,$$

$$f(y) \geq f(x + td) + (1 - t) \langle \nabla f(x + td), d\rangle .$$

Multiplicando a primeira desigualdade por $1 - t \geq 0$, a segunda por $t \geq 0$ e somando, obtemos

$$\begin{aligned}(1 - t)f(x) + tf(y) &\geq (1 - t)(f(x + td) - t \langle \nabla f(x + td), d\rangle) \\ &\quad + t(f(x + td) + (1 - t) \langle \nabla f(x + td), d\rangle) \\ &= f(x + td) \\ &= f((1 - t)x + ty), \end{aligned} \tag{1.8}$$

provando o item (a).

Suponhamos agora que f seja duas vezes diferenciável em C. Mostraremos que (b) $\Leftrightarrow$(d). Fixemos $x \in C$ e $d \in \mathbb{R}^n$ quaisquer. Como C é aberto, $x+td \in C$ para todo $t > 0$ suficientemente pequeno. Por (b)

$$f(x + td) - f(x) \geq t \langle \nabla f(x), d\rangle .$$

Usando ainda o fato de que f é duas vezes diferenciável,

$$\begin{aligned}0 &\leq f(x + td) - f(x) - t \langle \nabla f(x), d\rangle \\ &= \frac{t^2}{2} \left\langle \nabla^2 f(x)d, d\right\rangle + r(t),\end{aligned} \tag{1.9}$$

onde $\lim_{t\to 0} \frac{r(t)}{t^2} = 0$. Dividindo por $t^2 > 0$ e fazendo $t \to 0+$, obtemos (d). Sejam $x, y \in C$ quaisquer. Pelo Teorema do Valor Médio, existe $t \in (0, 1)$ tal que

$$\begin{aligned}&f(y) - f(x) - \langle \nabla f(x), y - x\rangle \\ &= \frac{1}{2} \left\langle \nabla^2 f(x + t(y - x))(y - x), y - x\right\rangle \geq 0,\end{aligned} \tag{1.10}$$

onde a desigualdade segue de (d). ■

Observe que o item (c) diz que o gradiente de uma função convexa é monótono.

1.2.2 Condições de Otimalidade.

Consideramos o problema de minimização irrestrita

$$\min f(x), \quad x \in \mathbb{R}^n, \tag{1.11}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$.

Teorema 1.2.9 *Suponhamos que a função $f : \mathbb{R}^n \to \mathbb{R}$ seja diferenciável no ponto $\bar{x} \in \mathbb{R}^n$. Se $\bar{x}$ é um minimizador local irrestrito de f, então*

$$\nabla f(\bar{x}) = 0. \tag{1.12}$$

Demonstração: Seja $d \in \mathbb{R}^n$ arbitrário, porém fixo. Pela definição de minimizador local, existe $\epsilon > 0$ tal que

$$f(\bar{x}) \leq f(\bar{x} + td), \quad \forall t \in [0, \epsilon].$$

Pela diferenciabilidade de f em $\bar{x}$,

$$f(\bar{x} + td) = f(\bar{x}) + t \langle \nabla f(\bar{x}), d \rangle + o(t).$$

Logo,

$$0 \leq t \langle \nabla f(\bar{x}), d \rangle + o(t).$$

Dividindo por $t > 0$, temos que

$$0 \leq \langle \nabla f(\bar{x}), d \rangle + o(t)/t,$$

e tomando o limite quando $t \to 0+$, obtemos

$$0 \leq \langle \nabla f(\bar{x}), d \rangle .$$

Como $d \in \mathbb{R}^n$ é arbitrário, podemos escolher $d = -\nabla f(\bar{x})$, o que resulta na condição $0 \leq \langle \nabla f(\bar{x}), d \rangle = - \|\nabla f(\bar{x})\|^2$. Donde segue que $\nabla f(\bar{x}) = 0$. ∎

Teorema 1.2.10 *Suponhamos que $f : \mathbb{R}^n \to \mathbb{R}$ seja duas vezes diferenciável no ponto $\bar{x} \in \mathbb{R}^n$. Se $\bar{x}$ é um minimizador local irrestrito de f, então vale* (1.12) *e a matriz Hessiana de f no ponto $\bar{x}$ é semidefinida positiva, i. e.,*

$$\left\langle \nabla^2 f(\bar{x}) d, d \right\rangle \geq 0, \quad \forall d \in \mathbb{R}^n. \tag{1.13}$$

Demonstração: A condição (1.12) já foi obtida acima.

Seja $d \in \mathbb{R}^n$ arbitrário, porém fixo. Se $\bar{x}$ é minimizador local do problema (1.11), então para todo $t > 0$ suficientemente pequeno

$$\begin{aligned} 0 &\leq f(\bar{x} + td) - f(\bar{x}) \\ &= \langle \nabla f(\bar{x}), td \rangle + \frac{1}{2} \left\langle \nabla^2 f(\bar{x}) td, td \right\rangle + o(t^2) \\ &= \frac{t^2}{2} \left\langle \nabla^2 f(\bar{x}) d, d \right\rangle + o(t^2), \end{aligned} \tag{1.14}$$

onde usamos o Teorema 1.2.9. Dividindo os dois lados da desigualdade acima por $t^2 > 0$, temos que

$$0 \leq \frac{1}{2} \left\langle \nabla^2 f(\bar{x}) d, d \right\rangle + o(t^2)/t^2.$$

Passando ao limite quando $t \to 0+$, obtemos (1.13). ■

Dizemos que

$$\min f(x) \quad \text{sujeito a} \quad x \in C \tag{1.15}$$

é um problema de minimização convexo quando $C \subset \mathbb{R}^n$ é um conjunto convexo e $f : C \to \mathbb{R}$ é uma função convexa em C.

Teorema 1.2.11 *Seja $C \subset \mathbb{R}^n$ um conjunto convexo e $f : C \to \mathbb{R}$ uma função convexa em C. Então o minimizador local do problema* (1.15) *é global. Além disso, o conjunto de minimizadores é convexo. Se f é estritamente convexa, o minimizador é único.*

Demonstração: Suponhamos que $\bar{x} \in C$ seja um minimizador local que não é global. Então existe $y \in C$ tal que $f(y) < f(\bar{x})$. Definimos $\alpha(t) = ty + (1-t)\bar{x}$. Como C é convexo, segue que $\alpha(t) \in C$, para todo $t \in [0,1]$. Usando a convexidade de f, para todo $t \in [0,1]$, temos que

$$f(\alpha(t)) \leq tf(y) + (1-t)f(\bar{x}) = f(\bar{x}) + t(f(y) - f(\bar{x})) < f(\bar{x}).$$

Tomando t suficientemente pequeno podemos garantir que o ponto $\alpha(t)$ é arbitrariamente próximo do ponto $\bar{x}$, temos também que $f(\alpha(t)) < f(\bar{x})$ e $\alpha(t) \in C$. Contradição, pois $\bar{x}$ é minimizador local do problema (1.15). Portanto qualquer solução deve ser global.

Seja $M \subset C$ o conjunto dos minimizadores globais e v^* o valor ótimo do problema, i. e., $f(x) = v^*$, $\forall x \in M$. Para quaisquer $x, \bar{x} \in M$ e $t \in [0,1]$, a convexidade de f nos garante que

$$f(tx + (1-t)\bar{x}) \leq tf(x) + (1-t)f(\bar{x}) = tv^* + (1-t)v^* = v^*,$$

implicando que $f(tx + (1-t)\bar{x}) = v^*$ e portanto $tx + (1-t)\bar{x} \in M$. Portanto M é convexo.

Resta mostrarmos a última afirmação. Supondo que f seja estritamente convexa e que existam $x, \bar{x} \in M$ tais que $x \neq \bar{x}$. Dado $t \in [0,1]$, já que x e $\bar{x}$ são minimizadores globais e a combinação convexa desses pontos pertence a C, usando a convexidade de C, temos que

$$f(tx + (1-t)\bar{x}) \geq f(x) = f(\bar{x}) = v^*.$$

Mas observe que, pela convexidade estrita de f,

$$f(tx + (1-t)\bar{x}) < tf(x) + (1-t)f(\bar{x}) = tv^* + (1-t)v^* = v^*.$$

Absurdo. Portanto o minimizador é único. ∎

1.2.3 Problema com Restrições e Condições de KKT

Sejam $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$ e $h : \mathbb{R}^n \to \mathbb{R}^m$ funções contínuas e suponhamos que elas sejam de classe C^1 e $D \subset \mathbb{R}^n$. Consideremos problemas de programação não linear da forma:

$$\min f(x) \quad \text{sujeito a} \quad h(x) = 0, \quad g(x) \leq 0 \quad \text{e} \quad x \in D.$$

As restrições $h(x) = 0$ e $g(x) \leq 0$ são chamadas de *restrições funcionais*. Um ponto $x \in \mathbb{R}^n$ que satisfaz todas as restrições funcionais é chamado de ponto viável.

A igualdade $h(x) = 0$ significa que $h_i(x) = 0, \forall i = 1, \cdots, m$. Analogamente $g(x) \leq 0$ significa que cada $g_i(x) \leq 0, \forall i = 1, \cdots, p$.

Definição 1.2.12 *Um ponto x^* satisfazendo a restrição $h(x^*) = 0$ é chamado de **valor regular para a restrição** se os vetores gradientes $\nabla h_1(x^*)$, $\nabla h_2(x^*)$, $\cdots$, $\nabla h_m(x^*)$ são linearmente independentes.*

Teorema 1.2.13 (Condições de Karush-Kuhn-Tucker) *Seja x^* um ponto de mínimo relativo para o problema*

$$\begin{aligned} &\min f(x) \\ &\textit{sujeito a} \quad h(x) = 0, \quad g(x) \leq 0, \end{aligned}$$

e suponha que x^* seja um valor regular para as restrições. Então existe um vetor $\lambda \in \mathbb{R}^m$ e um vetor $\mu \in \mathbb{R}^p$ com $\mu \geq 0$, tais que

$$\nabla f(x^*) + \lambda \nabla h(x^*) + \mu \nabla g(x^*) = 0 \tag{1.16}$$

$$\mu g(x^*) = 0. \tag{1.17}$$

Demonstração: Veja a referência [24]. $\blacksquare$

Exemplo 1.2.14 *As condições de KKT para o problema (P) são:*

$$\bar{x}_i \geq 0, \quad (\nabla f(\bar{x}))_i \geq 0 \quad e \quad \bar{x}_i(\nabla f(\bar{x}))_i = 0, \quad i = 1, \cdots, n.$$

Mais sobre condições de KKT pode ser encontrado em [14].

1.2.4 Tipos de Convergência.

Nesta seção apresentaremos algumas definições básicas de convergência que serão necessárias para a compreensão de algumas demonstrações posteriores.

Seja $\{x^k\}$ uma sequência em $\mathbb{R}^n$ que converge para x^*. A sequência $\{x^k\}$ converge linearmente para x^* se existe um $\lambda \in (0,1)$ tal que

$$\lim_{k\to+\infty} \frac{\left\|x^{k+1} - x^*\right\|}{\|x^k - x^*\|} = \lambda.$$

Se $\lambda = 1$, dizemos que a sequência converge sublinearmente e se $\lambda = 0$, dizemos que a sequência converge superlinearmente.

Exemplo 1.2.15 *A sequência $\{\frac{e^k}{k!}\}$ converge superlinearmente para 0, pois $\{\frac{e^k}{k!}\}$ converge para 0 e*

$$\lim_{k\to+\infty} \frac{|\frac{e^{k+1}}{(k+1)!} - 0|}{|\frac{e^k}{k!} - 0|} = \lim_{k\to+\infty} \frac{e^k \cdot e}{(k+1)\cdot k!} \cdot \frac{k!}{e^k} = \lim_{k\to+\infty} \frac{e}{k+1} = 0.$$

Capítulo 2

Método do Ponto Proximal: Caso Clássico

Neste capítulo abordaremos definição e análise de convergência do método do ponto proximal clássico. Mais detalhes podem ser encontrados em [17].

2.1 Algoritmo

Como apresentado em [25] e [30], um algoritmo é uma sequência de instruções ou passos finitos, especificados em uma determinada linguagem que mostram como resolver determinado problema. Podemos dizer também que é um conjunto de regras formais para a obtenção de um resultado ou da solução de um problema, englobando fórmulas de expressões aritméticas. Esta definição não está presa à utilização ou não do computador. Qualquer sequência de instruções que mostram como resolver um problema constitui-se um algoritmo. Uma receita de bolo é um bom exemplo de algoritmo, as instruções da receita são uma sequência de passos finitos e lógicos, os quais mostram o que um cozinheiro(a) deve fazer, e em que ordem deve fazer. Uma receita especifica todos os passos em uma sequência onde a ordem das instruções ou passos precisam ser seguidas para que possa obter o resultado final (nesse caso o bolo).

Um algoritmo deve especificar quais são as instruções que devem ser executadas, e em que ordem as mesmas devem ser executadas.

2.2 Método do Ponto Proximal

O método do ponto proximal clássico resolve o problema de minimização convexo irrestrito, isto é:

$$\min_{x\in\mathbb{R}^n} f(x)$$

Sendo assim, o método busca encontrar um ponto crítico de f, isto é, um ponto $\bar{x}$ que satisfaça $\nabla f(\bar{x}) = 0$, quando a função f é diferenciável.

O algoritmo do ponto proximal para otimização, em $\mathbb{R}^n$, gera uma sequência $\{x^k\} \subset \mathbb{R}^n$ da seguinte maneira:

$$x^0 \in \mathbb{R}^n \tag{2.1}$$

$$x^{k+1} = \text{argmin}_{x\in\mathbb{R}^n}\{f(x) + \lambda_k \left\|x - x^k\right\|^2\}, \tag{2.2}$$

onde λ_k é um número real satisfazendo $0 < \lambda_k \leq \tilde{\lambda}$ para algum $\tilde{\lambda} > 0$ (que inclui o caso de λ constante) e $\|\cdot\|$ é a norma euclidiana. Chamamos λ_k de parâmetro de regularização.

Sob alguma hipótese, a sequência gerada por (2.1)-(2.2) converge para o minimizador de f. Para uma visualização mais clara do ponto de vista computacional, escrevemos o algoritmo em portugol.

Tome $x^0 \in \mathbb{R}^n$ *e um* $\epsilon > 0$ *pequeno*;

$k = 0$;

Enquanto $\left\|\nabla f(x^k)\right\| \geq \epsilon$ {

$$x^{k+1} = \text{argmin}\{f(x) + \lambda_k \left\|x - x^k\right\|\};$$

$$k = k + 1;$$

}

A seguir definimos uma outra noção de convergência e provamos a validade do método do ponto proximal.

Uma sequência $\{y^k\} \subset \mathbb{R}^n$ é chamada de *Féjer convergente* para um conjunto $U \subset \mathbb{R}^n$ em relação à distância euclidiana se

$$\left\|y^{k+1}-u\right\| \leq \left\|y^k-u\right\|, \quad \forall k \geq 0, \quad \forall u \in U. \tag{2.3}$$

Temos o seguinte resultado.

Proposição 2.2.1 *Se $\{y^k\}$ é Féjer convergente para $U \neq \emptyset$, então $\{y^k\}$ é limitada. Se um valor de aderência y de $\{y^k\}$ pertence a U, então $y = \lim_{k\to\infty} y^k$.*

Demonstração: A desigualdade (2.3) implica que $\left\|y^k-u\right\| \leq \|y^0-u\|$ para qualquer $u \in U$, assim a sequência $\{y^k\}$ está contida na bola de centro u e raio $\|y^0-u\|$, portanto, é limitada. Para a segunda sentença, seja $\{y^{k_j}\}$ uma subsequência de $\{y^k\}$ tal que $\lim_{k\to\infty} y^{k_j} = y$. Como $y \in U$, por (2.3) a sequência $\{\left\|y^k-y\right\|\}$ é decrescente e não negativa, e ela tem uma subsequência $(\{\left\|y^{k_j}-y\right\|\})$ que converge a zero. Então toda a sequência converge para 0, isto é, $0 = \lim_{k\to\infty} \left\|y^k-y\right\|$, implicando que $\lim_{k\to\infty} y^k = y$. ■

Agora podemos provar a convergência do algoritmo do ponto proximal.

Teorema 2.2.2 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa e continuamente diferenciável. Suponha que o conjunto U de minimizadores de f em $\mathbb{R}^n$ seja não vazio. Então a sequência $\{x^k\}$ gerada por* (2.1)-(2.2) *converge para um ponto $x^* \in U$.*

Demonstração: Dividimos a demonstração em 4 passos. No Passo 1 provamos que $\{x^k\}$ está bem definida. No Passo 2 provamos que a sequência é Féjer convergente para U. No Passo 3 garantimos o resultado que $\lim_{k\to\infty}(x^{k+1}-x^k) = 0$, para ser usado no Passo 4, onde provamos que qualquer valor de aderência de $\{x^k\}$ pertence a U. Os resultados dos Passos 2 e 4, junto com a Proposição 2.2.1, garantem o Teorema.

Passo 1: A sequência $\{x^k\}$ está bem definida.

Por indução. Seja $f_k(x) = f(x) + \lambda_k \left\|x-x^k\right\|^2$. Como f atinge seu mínimo, ela é limitada inferiormente, logo $\lim_{\|x\|\to\infty} f_k(x) = +\infty$. Já que f_k é contínua e coerciva, a minimização em (2.2) se reduz a conjuntos compactos, pois o comportamento de $f_k(x)$ para x suficientemente grande não nos interessa. Assim f_k atinge seu mínimo. Como f é convexa e $\lambda_k \left\|x-x^k\right\|^2$ é estritamente convexa, segue que f_k é estritamente convexa e portanto possui um único minimizador, isto é, x^{k+1} é unicamente determinado.

Passo 2: $\left\|x^{k+1}-\bar{x}\right\| \leq \left\|x^k-\bar{x}\right\|$ para todo $k \geq 0$ e todo $\bar{x} \in U$.

De fato,

$$\begin{aligned}\left\|x^k-\bar{x}\right\|^2 &= \left\|x^k-x^{k+1}+x^{k+1}-\bar{x}\right\|^2 \\ &= \left\|x^k-x^{k+1}\right\|^2+\left\|x^{k+1}-\bar{x}\right\|^2+2\left\langle x^k-x^{k+1}, x^{k+1}-\bar{x}\right\rangle. \end{aligned} \tag{2.4}$$

Como x^{k+1} resolve (2.2), temos

$$0=\nabla f_k(x^{k+1})=\nabla f(x^{k+1})+2\lambda_k(x^{k+1}-x^k) \tag{2.5}$$

De (2.4), (2.5) e da convexidade de f

$$\begin{aligned}&\left\|x^k-\bar{x}\right\|^2-\left\|x^{k+1}-x^k\right\|^2-\left\|x^{k+1}-\bar{x}\right\|^2=2\left\langle x^k-x^{k+1}, x^{k+1}-\bar{x}\right\rangle \\ &=\frac{1}{\lambda_k}\left\langle \nabla f(x^{k+1}), x^{k+1}-\bar{x}\right\rangle \geq \frac{1}{\lambda_k}[f(x^{k+1})-f(\bar{x})] \geq 0\end{aligned} \tag{2.6}$$

usando o fato que $\bar{x}$ é um minimizador de f. O resultado segue de (2.6). ■

Passo 3: $\lim_{k\to\infty}(x^{k+1}-x^k)=0$.

Do Passo 2

$$0 \leq \left\|x^{k+1}-x^k\right\|^2 \leq \left\|x^k-\bar{x}\right\|^2-\left\|x^{k+1}-\bar{x}\right\|^2. \tag{2.7}$$

Como $\{\left\|x^k-\bar{x}\right\|\}$ é uma sequência decrescente e não negativa, ela é convergente. Logo o lado direito de (2.7) converge a zero. O resultado segue.

Passo 4: $\{x^k\}$ possui valores de aderência e todos eles pertencem a U.

A existência dos valores de aderência segue diretamente do Passo 2 e da primeira sentença da Proposição 2.2.1. Seja $\bar{x}$ um valor de aderência de $\{x^k\}$ e $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$ tal que $\lim_{k\to\infty} x^{k_j}=\bar{x}$. Por (2.5)

$$\nabla f(x^{k_j+1})=2\lambda_{k_j}(x^{k_j}-x^{k_j+1}). \tag{2.8}$$

Pelo Passo 3, $\lim_{k\to\infty} x^{k_j+1}=\lim_{k\to\infty} x^{k_j}=\bar{x}$. Tomando limites em (2.8) quando $k\to+\infty$, usando $\lambda_k \leq \tilde{\lambda}$ e a diferenciabilidade contínua de f, obtemos $\nabla f(\bar{x})=0$. Pela convexidade da f, $\bar{x}\in U$. Os Passos 2 e 4 indicam que ambas as sentenças da Proposição 2.2.1 se mantém e portanto existe um $\bar{x}\in U$ tal que $\bar{x}=\lim_{k\to\infty} x^k$. ■

Capítulo 3

Método Interior para Minimização Sob o Ortante Não Negativo.

Neste capítulo apresentaremos dois algoritmos para resolver o problema (P), que reescrevemos abaixo

$$(P) \quad \min f(x) \quad \text{sujeito} \quad \text{a} \quad x \geq 0, \tag{3.1}$$

onde $x \geq 0$ significa que cada coordenada do vetor $x \in \mathbb{R}^n$ é não negativa.
Assumimos que
(A1) $f : \mathbb{R}^n \to \mathbb{R}$ é uma função continuamente diferenciável em $\mathbb{R}^n$.
(A2) Existe um $\alpha \in \mathbb{R}$ tal que $f(x) \geq \alpha$, para todo $x \in \mathbb{R}^n$.
(A3) A função objetivo f é convexa.
A forma geral do algoritmo analisado nesta dissertação é apresentado abaixo:

$$x^0 > 0 \tag{3.2}$$

$$x^{k+1} = \text{argmin}_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{\beta_k}{2} \left\| x - x^k \right\|^2_{(X^k)^{-r}} \right\}, \tag{3.3}$$

onde $(X^k)^{-r}$ é definida por

$$(X^k)^{-r} := \begin{pmatrix} (x_1^k)^{-r} & 0 & \cdots & 0 \\ 0 & (x_2^k)^{-r} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (x_n^k)^{-r} \end{pmatrix},$$

$\left\|x-x^k\right\|^2_{(X^k)^{-r}}$ é a norma na matriz definida anteriormente e os parâmetros de regularização β_k são calculados de acordo com uma das seguintes definições.

Algoritmo I. Para propormos o Algoritmo I, supomos que f satisfaz (A1)-(A3) e a hipótese:

(A4) O gradiente de f é L-Lipschitz, isto é:

$$\exists L > 0; \quad \|\nabla f(x) - \nabla f(y)\| \leq L\,\|x-y\| \quad \forall x, y \in \mathbb{R}^n.$$

Também consideramos a seguinte condição:

(R1) Um número real $r \geq 1$.

Neste algoritmo, dado x^k, calculamos β_k como segue

$$\beta_k = (max(x^k))^{r-1}\left\|\nabla f(x^k)\right\| + (L+1)(max(x^k))^r. \tag{3.4}$$

Algoritmo II. Para propormos o Algoritmo II, supomos que valem as hipóteses (A1)-(A3) e consideramos a seguinte condição:

(R2) Um número real $r \geq 2$.

Neste algoritmo, dado x^k, calculamos β_k como segue

$$\beta_k = (max(x^k))^{r-1}\left\|\nabla f(x^k)\right\| + (max(x^k))^r. \tag{3.5}$$

Note que o Algoritmo II pode ser aplicado para uma classe maior de funções, por não exigirmos que a função objetivo tenha gradiente lipschitziano.

Observe que a definição de x^{k+1} em (3.3) é equivalente a x^{k+1} satisfazer $\nabla f_k(x^{k+1}) = 0$, isto é,

$$\nabla f(x^{k+1}) + \beta_k (X^k)^{-r}(x^{k+1} - x^k) = 0, \tag{3.6}$$

onde $f_k(x) := f(x) + \frac{\beta_k}{2}\left\|x-x^k\right\|^2_{(X^k)^{-r}}$. Além disso, é interessante mencionar que a motivação para a definição dos parâmetros β_k, apresentados nos Algoritmos I e II, é garantir que a sequência $\{x^k\}$ permaneça no interior do conjunto viável, i. e., $x^k \in \mathbb{R}^n_{++}$ para todo $k \in \mathbb{N}$. Este resultado será provado posteriormente.

3.1 Existência e Resultados Preliminares.

Nesta seção nosso principal objetivo é mostrar a existência da sequência $\{x^k\}$ gerada por (3.2)-(3.3) usando os Algoritmos I e II. Além disso, apresentamos resultados preliminares necessários para nossa análise de convergência.

Proposição 3.1.1 *Existe uma sequência* $\{x^k\}$ *gerada por* (3.2)-(3.3), *usando os Algoritmos I ou II.*

Demonstração: Pela hipótese (A2), seja α a cota inferior para f em $\mathbb{R}^n$, ela garante que

$$\begin{aligned} f_k(x) &= f(x) + \frac{\beta_k}{2}\left\|x - x^k\right\|^2_{(X^k)^{-r}} \\ &\geq \alpha + \frac{\beta_k}{2}\left\|x - x^k\right\|^2_{(X^k)^{-r}}. \end{aligned} \tag{3.7}$$

Já que $h(x) = \frac{\beta_k}{2}\left\|x - x^k\right\|^2_{(X^k)^{-r}}$ é uma função nível-limitada, temos que a minimização em (3.3) se reduz a conjuntos compactos e, portanto, o mínimo de f_k é atingido. Logo x^{k+1} existe.

■

3.2 Análise de Convergência.

Iniciaremos esta seção mostrando que (3.2)-(3.3) é um método de ponto interior com parâmetro de regularização β_k definido de acordo com os Algoritmos I ou II. A partir de agora, $\{x^k\}$ refere-se à sequência gerada por (3.2)-(3.3) usando o Algoritmo I ou II.

Proposição 3.2.1 *A sequência* $\{x^k\}$ *satisfaz* $x^k > 0$, *para todo* $k \in \mathbb{N}$.

Demonstração: Dividiremos a demonstração em dois casos: primeiro quando β_k é dado por (3.4) e segundo quando β_k é calculado por (3.5).
Primeiro caso. O parâmetro de regularização β_k é definido por (3.4).
Faremos esta prova por indução. É imediato o caso $k = 0$, pois segue diretamente de (3.2). Suponha que $x^k > 0$. De (3.6), temos

$$\nabla f(x^{k+1}) + \beta_k (X^k)^{-r}(x^{k+1} - x^k) = 0.$$

Logo

$$\beta_k (X^k)^{-r}(x^{k+1} - x^k) = -\nabla f(x^{k+1}) \Leftrightarrow (X^k)^{-r}(x^{k+1} - x^k) = -\frac{1}{\beta_k}\nabla f(x^{k+1}).$$

Portanto

$$(X^k)^{-1}(x^{k+1} - x^k) = -\frac{1}{\beta_k}(X^k)^{r-1}\nabla f(x^{k+1}). \tag{3.8}$$

Agora, basta mostrarmos que $\left\|(X^k)^{-1}(x^{k+1}-x^k)\right\| < 1$. De fato, pela equação (3.8) e pela desigualdade triangular, obtemos

$$\begin{aligned}
\left\|(X^k)^{-1}(x^{k+1}-x^k)\right\| &= \frac{1}{\beta_k}\left\|(X^k)^{r-1}\nabla f(x^{k+1})\right\| \\
&\leq \frac{1}{\beta_k}\left\|(X^k)^{r-1}\nabla (f(x^k)+f(x^{k+1})-f(x^k))\right\| \\
&\leq \frac{1}{\beta_k}\left\|(X^k)^{r-1}\nabla f(x^k)\right\| + \frac{1}{\beta_k}\left\|(X^k)^{r-1}(\nabla f(x^{k+1})-\nabla f(x^k))\right\|. \qquad (3.9)
\end{aligned}$$

Portanto,

$$\left\|(X^k)^{-1}(x^{k+1}-x^k)\right\| \leq \frac{1}{\beta_k}\left\|(X^k)^{r-1}\nabla f(x^k)\right\| + \frac{1}{\beta_k}(max(x^k))^{r-1}\left\|\nabla f(x^{k+1})-\nabla f(x^k)\right\|$$

e, pela hipótese (A3), ∇f é L-Lipschitz, logo

$$\left\|(X^k)^{-1}(x^{k+1}-x^k)\right\| \leq \frac{1}{\beta_k}\left\|(X^k)^{r-1}\nabla f(x^k)\right\| + \frac{L}{\beta_k}(max(x^k))^{r-1}\left\|x^{k+1}-x^k\right\|. \qquad (3.10)$$

A desigualdade (3.10) pode ser escrita como

$$\left\|(X^k)^{-1}(x^{k+1}-x^k)\right\| \leq \frac{1}{\beta_k}\left\|(X^k)^{r-1}\nabla f(x^k)\right\| + \frac{L}{\beta_k}(max(x^k))^{r}(max(x^k))^{-1}\left\|x^{k+1}-x^k\right\|. \qquad (3.11)$$

Observe que $\left\|(X^k)^{-1}(x^{k+1}-x^k)\right\| \geq (max(x^k))^{-1}\left\|x^{k+1}-x^k\right\|$, pois

$$\begin{aligned}
\left\|(X^k)^{-1}(x^{k+1}-x^k)\right\| &= \left\|\begin{pmatrix} (x_1^k)^{-1} & 0 & \cdots & 0 \\ 0 & (x_2^k)^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (x_n^k)^{-1} \end{pmatrix} \cdot \begin{pmatrix} x_1^{k+1}-x_1^k \\ x_2^{k+1}-x_2^k \\ \vdots \\ x_n^{k+1}-x_n^k \end{pmatrix}\right\| \\
&= \left\|\begin{pmatrix} (x_1^k)^{-1}(x_1^{k+1}-x_1^k) & (x_2^k)^{-1}(x_2^{k+1}-x_2^k) & \cdots & (x_n^k)^{-1}(x_n^{k+1}-x_n^k) \end{pmatrix}\right\| \\
&= \left(\sum_{i=1}^{n}\left[(x_i^k)^{-1}(x_i^{k+1}-x_i^k)\right]^2\right)^{\frac{1}{2}} \\
&\geq \left(\sum_{i=1}^{n}\left[(\max(x^k))^{-1}(x_i^{k+1}-x_i^k)\right]^2\right)^{\frac{1}{2}} \\
&= (\max(x^k))^{-1}\left(\sum_{i=1}^{n}(x_i^{k+1}-x_i^k)^2\right)^{\frac{1}{2}} \\
&= (\max(x^k))^{-1}\left\|x^{k+1}-x^k\right\|.
\end{aligned}$$

Portanto, por esta desigualdade e por (3.11), obtemos

$$\left\|(X^k)^{-1}(x^{k+1}-x^k)\right\| \leq \frac{1}{\beta_k}\left\|(X^k)^{r-1}\nabla f(x^k)\right\| + \frac{L}{\beta_k}(max(x^k))^r \left\|(X^k)^{-1}(x^{k+1}-x^k)\right\|, \tag{3.12}$$

o qual implica que

$$(\beta_k - L(max(x^k))^r)\left\|(X^k)^{-1}(x^{k+1}-x^k)\right\| \leq \left\|(X^k)^{r-1}\nabla f(x^k)\right\|. \tag{3.13}$$

Da definição de β_k dada em (3.4), temos que $\beta_k - L(max(x^k))^r > 0$. Assim

$$\begin{aligned}\left\|(X^k)^{-1}(x^{k+1}-x^k)\right\| &\leq \frac{\left\|(X^k)^{r-1}\nabla f(x^k)\right\|}{(\beta_k - L(max(x^k))^r)}\\ &\leq \frac{(max(x^k))^{r-1}\left\|\nabla f(x^k)\right\|}{(\beta_k - L(max(x^k))^r)}\\ &< 1. \end{aligned} \tag{3.14}$$

Observe que a desigualdade estrita acima ocorre devido a $x^k > 0$, pela hipótese de indução. Finalmente, a desigualdade (3.14) implica que $|\frac{x_i^{k+1}}{x_i^k} - 1| < 1$, para todo $i = 1, \cdots, n$, e consequentemente

$$-1 < \frac{x_i^{k+1}}{x_i^k} - 1 < 1 \Rightarrow 0 < \frac{x_i^{k+1}}{x_i^k} < 2 \Rightarrow 0 < x_i^{k+1} < 2x_i^k, \quad \forall i = 1, \cdots, n.$$

Portanto, $x^{k+1} > 0$.

Segundo caso. O parâmetro de regularização β_k é definido por (3.5).

Mostraremos a positividade de x^k por indução. Veja que $x^0 > 0$, por (3.2). Suponhamos que $x^k > 0$. Da monotonicidade do gradiente de f, temos que

$$\left\langle \nabla f(x^{k+1}) - \nabla f(x^k), x^k - x^{k+1} \right\rangle \leq 0. \tag{3.15}$$

Por (3.6) e (3.15), ganhamos que

$$\left\langle -\beta_k (X^k)^{-r}(x^{k+1}-x^k) - \nabla f(x^k), x^k - x^{k+1} \right\rangle \leq 0,$$

e portanto

$$\left\langle \beta_k (X^k)^{-r}(x^{k+1}-x^k), x^{k+1}-x^k \right\rangle \leq \left\langle \nabla f(x^k), x^k - x^{k+1} \right\rangle. \tag{3.16}$$

Usando o fato que $(X^k)^{-r} = (X^k)^{-1}(X^k)^{-r+2}(X^k)^{-1}$ e que $(X^k)^{-1}$ é autoadjunta em (3.16), obtemos

$$\begin{aligned}\left\langle \nabla f(x^k), x^k - x^{k+1} \right\rangle &\geq \left\langle \beta_k (X^k)^{-1}(X^k)^{-r+2}(X^k)^{-1}(x^{k+1} - x^k), x^{k+1} - x^k \right\rangle \\ &= \beta_k \left\langle (X^k)^{-r+2}(X^k)^{-1}(x^{k+1} - x^k), (X^k)^{-1}(x^{k+1} - x^k) \right\rangle \\ &\geq \beta_k (max(x^k))^{-r+2} \left\langle (X^k)^{-1}(x^{k+1} - x^k), (X^k)^{-1}(x^{k+1} - x^k) \right\rangle \\ &= \beta_k (max(x^k))^{-r+2} \left\| (X^k)^{-1}(x^{k+1} - x^k) \right\|^2,\end{aligned} \tag{3.17}$$

onde $r \geq 2$, por (R2). Observe que

$$\left\| (X^k)^{-1}(x^{k+1} - x^k) \right\| \geq (max(x^k))^{-1} \left\| x^{k+1} - x^k \right\|.$$

Multiplicando a desigualdade acima por $\beta_k (max(x^k))^{-r+2} \left\| (X^k)^{-1}(x^{k+1} - x^k) \right\|$, obtemos

$$\left\| (X^k)^{-1}(x^{k+1} - x^k) \right\|^2 \beta_k (max(x^k))^{-r+2} \geq \beta_k (max(x^k))^{-r+1} \left\| x^{k+1} - x^k \right\| \left\| (X^k)^{-1}(x^{k+1} - x^k) \right\|. \tag{3.18}$$

Usando as desigualdades (3.17) e (3.18), segue que

$$\begin{aligned}\beta_k (max(x^k))^{-r+1} \left\| x^{k+1} - x^k \right\| \left\| (X^k)^{-1}(x^{k+1} - x^k) \right\| &\leq \left\langle \nabla f(x^k), x^k - x^{k+1} \right\rangle \\ &\leq \left\| \nabla f(x^k) \right\| \left\| x^k - x^{k+1} \right\|,\end{aligned} \tag{3.19}$$

onde a última desigualdade segue diretamente da desigualdade de Cauchy-Schwartz. Logo

$$\beta_k (max(x^k))^{-r+1} \left\| (X^k)^{-1}(x^{k+1} - x^k) \right\| \leq \left\| \nabla f(x^k) \right\|,$$

e, assim

$$\left\| (X^k)^{-1}(x^{k+1} - x^k) \right\| \leq \frac{\left\| \nabla f(x^k) \right\| (max(x^k))^{r-1}}{\beta_k}. \tag{3.20}$$

Da definição de β_k em (3.5) e da desigualdade (3.20), concluimos que

$$\left\| (X^k)^{-1}(x^{k+1} - x^k) \right\| < 1. \tag{3.21}$$

Um argumento análogo ao primeiro caso nos garante que $x^{k+1} > 0$. ■

Proposição 3.2.2 *Para os Algoritmos I e II, temos que*

(i) $\{f(x^k)\}$ *é uma sequência decrescente e convergente.*

(ii) $\sum_{k=0}^{+\infty} \beta_k \left\| x^{k+1} - x^k \right\|^2_{(X^k)^{-r}} < +\infty$.

(iii) $\sum_{k=0}^{+\infty} \left\| x^{k+1} - x^k \right\|^2 < +\infty$.

Demonstração: Parte (i). Da definição de x^{k+1} em (3.3), temos

$$f(x^{k+1}) + \frac{\beta_k}{2} \left\|x^{k+1} - x^k\right\|^2_{(X^k)^{-r}} \leq f(x) + \frac{\beta_k}{2} \left\|x - x^k\right\|^2_{(X^k)^{-r}},$$

para todo $x \in \mathbb{R}^n$. Logo, tomando $x = x^k$ na última desigualdade, obtemos

$$f(x^{k+1}) + \frac{\beta_k}{2} \left\|x^{k+1} - x^k\right\|^2_{(X^k)^{-r}} \leq f(x^k), \tag{3.22}$$

e portanto

$$f(x^{k+1}) \leq f(x^k).$$

Temos que f é decrescente, assim f é limitada superiormente por $f(x^0)$. Além disso, pela hipótese (A2), f é limitada inferiormente. A conclusão segue.

Parte (ii). Veja que se α é uma cota inferior para f, então $-\alpha$ é uma conta superior para $-f$. Este argumento, junto com a desigualdade (3.22), implica em:

$$\begin{aligned}
\sum_{k=0}^{l} \frac{\beta_k}{2} \left\|x^{k+1} - x^k\right\|^2_{(X^k)^{-r}} &\leq \sum_{k=0}^{l} \left(f(x^k) - f(x^{k+1})\right) \\
&= f(x^0) - f(x^1) + f(x^1) - f(x^2) + \cdots + f(x^l) - f(x^{l+1}) \\
&= f(x^0) - f(x^{l+1}) \\
&\leq f(x^0) - \alpha.
\end{aligned} \tag{3.23}$$

Multiplicando a desigualdade acima por 2 e fazendo $l \longrightarrow +\infty$, obtemos

$$\sum_{k=0}^{\infty} \beta_k \left\|x^{k+1} - x^k\right\|^2_{(X^k)^{-r}} \leq 2(f(x^0) - \alpha) < +\infty. \tag{3.24}$$

Assim segue (ii).

Parte (iii). Nesta prova, consideramos dois casos para β_k, de acordo com as definições dadas pelos Algoritmos I e II.

Primeiro caso. Quando β_k é gerado pelo Algoritmo I, isto é, β_k é definido por (3.4), ele garante que

$$\beta_k \left\|x^{k+1} - x^k\right\|^2_{(X^k)^{-r}} \geq L(max(x^k))^r \left\|x^{k+1} - x^k\right\|^2_{(X^k)^{-r}} \geq L \left\|x^{k+1} - x^k\right\|^2. \tag{3.25}$$

Somando de 0 a l a equação (3.25), fazendo $l \longrightarrow +\infty$ e usando a Parte (ii), concluímos a prova.

Segundo caso. Quando β_k é dado por (3.5) no Algoritmo II, obtemos

$$\beta_k \left\|x^{k+1} - x^k\right\|^2_{(X^k)^{-r}} \geq (max(x^k))^r \left\|x^{k+1} - x^k\right\|^2_{(X^k)^{-r}} \geq \left\|x^{k+1} - x^k\right\|^2. \tag{3.26}$$

Prosseguindo como no primeiro caso, o resultado também segue da Parte (ii). ■

Corolário 3.2.3 *A sequência* $\{x^k\}$ *satisfaz*

$$\lim_{k\to+\infty}(x^{k+1}-x^k)=0.$$

Demonstração: Pelo item (iii) da Proposição (3.2.2), temos que $\sum_{k=0}^{+\infty}\left\|x^{k+1}-x^k\right\|^2<+\infty$, e sabemos que, se uma série converge, seu termo geral vai a zero. Portanto

$$\lim_{k\to+\infty}(x^{k+1}-x^k)=0. \qquad \blacksquare$$

3.3 Relação entre os Valores de Aderência da Sequência e o Problema (P).

Nesta seção mostraremos que, se $\{x^k\}$ possui valores de aderência, então todos eles são soluções para o problema (P). Mostraremos também que se o conjunto solução S^* de (P) é vazio, então a sequência $\{x^k\}$ é ilimitada.

Teorema 3.3.1 *Se* $\{x^k\}$ *é limitada então todos valores de aderência de* $\{x^k\}$ *pertencem a* S^**. A condição que garante a existência dos valores de aderência de* $\{x^k\}$ *é a limitação do conjunto solução ou de qualquer outro conjunto de nível de* f*.*

Demonstração: Faremos a demonstração em quatro passos. Nos passos 1-3 verificaremos que qualquer valor de aderência satisfaz as condições de Karush-Kuhn-Tucker para o problema (P) em (3.1). No quarto passo apresentaremos condições para a limitação de $\{x^k\}$.

Passo 1. Seja $\bar{x}$ um valor de aderência de $\{x^k\}$. Então $\bar{x}\geq 0$.

Se $\bar{x}$ é um valor de aderência de $\{x^k\}$, então existe uma subsequência $\{x^{k_j}\}$ de $\{x^k\}$ tal que $x^{k_j}\longrightarrow\bar{x}$ quando $j\to+\infty$. Pela Proposição 3.2.1 temos que $x^k>0$, para todo $k\in\mathbb{N}$, assim $x^{k_j}>0$, $\forall j\in\mathbb{N}$. Passando ao limite na última desigualdade, obtemos

$$\bar{x}=\lim_{j\to+\infty}x^{k_j}\geq 0.$$

Passo 2. Qualquer valor de aderência $\bar{x}$ de $\{x^k\}$ satisfaz $\bar{x}_i(\nabla f(\bar{x}))_i=0$, para todo $i=1,\cdots,n$.

Por (3.6), temos

$$\frac{1}{\beta_k}(X^k)^r\nabla f(x^{k+1})=x^{k+1}-x^k,$$

passando ao limite e usando a Proposição 3.2.3, resulta

$$\lim_{k\to+\infty} \frac{1}{\beta_k}(X^k)^r \nabla f(x^{k+1}) = \lim_{k\to+\infty}(x^{k+1} - x^k) = 0. \tag{3.27}$$

Já que β_k é limitada (por causa da limitação de x_k), o limite (3.27) implica que

$$\lim_{k\to+\infty}(X^k)^r \nabla f(x^{k+1}) = 0. \tag{3.28}$$

Seja $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$ tal que $\lim_{j\to+\infty} x^{k_j} = \bar{x}$. Por (3.28), temos

$$\lim_{j\to+\infty}(X^{k_j})^r \nabla f(x^{k_j+1}) = 0,$$

e como $\lim_{j\to+\infty} x^{k_j+1} = \lim_{j\to+\infty} x^{k_j} = \bar{x}$ (Proposição 3.2.3), obtemos

$$\bar{X}^r \nabla f(\bar{x}) = 0.$$

Portanto $\bar{x}_i(\nabla f(\bar{x}))_i = 0$, para todo $i = 1, \cdots, n$.

Passo 3. Qualquer valor de aderência $\bar{x}$ de $\{x^k\}$ satisfaz $(\nabla f(\bar{x}))_i \geq 0$, $\forall i = 1, \cdots, n$. A menos de algumas pequenas modificações, a prova que segue é dada por Iusem em [16]. Para a conveniência do leitor, vamos escrevê-la em detalhes.

Seja $\{x^{k_j}\}$ uma sequência de $\{x^k\}$ tal que $\lim_{j\to+\infty} x^{k_j} = \bar{x}$. Suponha por absurdo que exista um i tal que $(\nabla f(\bar{x}))_i < 0$. Consideramos os três seguintes conjuntos:

$$I = \{i;\ (\nabla f(\bar{x}))_i < 0\}, \quad J = \{1, 2, \cdots, n\} - I \quad \text{e} \quad V = \{x \in \mathbb{R}^n_+;\ x_i = 0,\ i \in I\}.$$

Observe que V é um conjunto convexo e pela nossa hipótese, $I \neq \emptyset$. Do Passo 2, temos que $\bar{x}_i(\nabla f(\bar{x}))_i = 0$, para todo $i = 1, \cdots, n$. Logo, $\bar{x}_i = 0$, $\forall i \in I$ e portanto $\bar{x} \in V$. Além disso, pelo Passo 2 e pela Proposição 3.2.1, temos:

$$\bar{x}_i \geq 0, \quad \bar{x}_i(\nabla f(\bar{x}))_i = 0 \quad \text{e} \quad (\nabla f(\bar{x}))_i \geq 0,\ \forall i \in J,$$

que são precisamente as condições de KKT para o seguinte problema:

$$(PV) \quad \min f(x) \quad \text{sujeito a} \quad x \in V. \tag{3.29}$$

Como f é convexa, $\bar{x}$ é uma solução para o problema (3.29). Agora, consideremos o seguinte conjunto aberto de $\mathbb{R}^n_+$:

$$B = \{x \in \mathbb{R}^n_+;\ (\nabla f(x))_i < 0,\ i \in I\},$$

que é um conjunto não vazio, pois $\bar{x} \in B$. Como B é um conjunto aberto, podemos supor que $\{x^{k_j}\} \subset B$.

• Afirmamos que $\{x^k\}$ possui infinitos termos $\{x^{k_l}\}$ tais que $x^{k_l} \notin B$. De fato, se isto não se mantém, existiria um $k_0 \in \mathbb{N}$ tal que $\{x^k\} \in B$, $\forall k \geq k_0$. Por (3.6), temos que

$$x^{k+1} = x^k - \frac{1}{\beta_k}(X^k)^r \nabla f(x^{k+1}).$$

Portanto $x_i^{k+1} \geq x_i^k$, $\forall i \in I$, $\forall k \geq k_0$ e consequentemente $x_i^{k+1} \geq x_i^0 > 0$, $\forall i \in I$ e $\forall k \geq k_0$. Para os termos da subsequência $\{x^{k_j}\}$, que converge a $\bar{x}$, temos que

$$x_i^{k+1} \geq x_i^k > 0. \tag{3.30}$$

Passando ao limite em (3.30) e usando a Proposição 3.2.3, segue que

$$0 = \bar{x}_i = \lim_{j \to +\infty} x_i^{k_j+1} = \lim_{j \to +\infty} x_i^{k_j} \geq x_i^0 > 0,$$

o que é uma contradição. Agora, considere a subsequência definida por

$$p_j = \max\{q < k_j + 1;\ x^q \notin B\}.$$

Faça $p_j = 0$, se $x^q \in B$ para todo $q \leq k_j + 1$. Portanto, para todo q tal que $p_j + 1 \leq q \leq k_j + 1$, temos que $x^q \in B$. Assim, para todo q satisfazendo $p_j + 1 \leq q \leq q + 1 \leq k_j + 1$, temos

$$x_i^{k_j+1} \geq x_i^{q+1} \geq x_i^q \geq x_i^{p_j+1} \quad \forall i \in I. \tag{3.31}$$

• Afirmamos que $\lim_{j \to +\infty} p_j = +\infty$. Com efeito, se existisse algum $j_0 \in \mathbb{N}$ suficientemente grande, tal que $p_j = p_{j_0}$, $\forall j \geq j_0$, então (3.31) implicaria que

$$x_i^{k_j+1} \geq x_i^{q+1} \geq x_i^q \geq x_i^{p_j+1} \geq x^{p_{j_0}+1} \quad \forall i \in I. \tag{3.32}$$

Fazendo $j \to +\infty$ em (3.32), obtemos

$$0 = \bar{x}_i = \lim_{j \to +\infty} x_i^{k_j+1} \geq x_i^{p_{j_0}+1} > 0,$$

que é uma contradição. Da desigualdade (3.31), temos que $\lim_{j \to +\infty} x^{p_j+1} = 0$, para cada $i \in I$. Já que $\lim_{j \to +\infty}(x^{k+1} - x^k) = 0$, segue que

$$\lim_{j \to +\infty} x^{p_j} = 0, \quad \forall i \in I. \tag{3.33}$$

Da definição de p_j, $x^{p_j} \in B$ para todo $j \in \mathbb{N}$. Considere $\{x^{p_l}\} \subset \{x^{p_j}\}$ tal que $x^{p_l} \to \widetilde{x}$. Como B é aberto, temos que $\widetilde{x} \notin B$. Sabemos que $f(x^k)$ é convergente. Seja f^* o limite de $f(x^k)$. Assim,

$$f(\bar{x}) = f(\widetilde{x}) = \lim_{k \to +\infty} f(x^k) = f^*.$$

De (3.33), resulta que

$$\bar{x}_i = \lim_{l \to +\infty} x_i^{p_l} = \lim_{j \to +\infty} x_i^{p_j} = 0, \quad \forall i \in I.$$

Portanto, $\widetilde{x} \in V$ e $f(\widetilde{x}) = f(\bar{x})$. Isto significa que $\widetilde{x}$ resolve (PV) em (3.29). Como $\widetilde{x} \notin B$, temos que $(\nabla f(\widetilde{x}))_s \geq 0$, para algum $s \in I$. Consideremos o seguinte problema

$$(PW) \quad \min f(x) \quad \text{sujeito a} \quad x \in W, \tag{3.34}$$

onde $W = \{x \in \mathbb{R}^n_+;\ x_i = 0,\ i \in I - \{s\}\}$. As condições de KKT para o problema (PW) são

$$x_j \geq 0, \quad (\nabla f(x))_j \geq 0, \quad x_j(\nabla f(x))_j = 0, \quad j \in J \cup \{s\}. \tag{3.35}$$

É óbvio que $\widetilde{x}$ satisfaz estas condições. Por causa da convexidade de f (e o fato que $\widetilde{x} \in W$), $\widetilde{x}$ é solução de (PW). Como $V \subset W$, $\bar{x} \in V$ e

$$f(\widetilde{x}) = f(\bar{x}) = \lim_{k \to +\infty} f(x^k) = f^*.$$

Segue que $\bar{x}$ também satisfaz as condições de otimalidade para (PW), isto é, $\bar{x}$ é solução do problema (PW). Em particular, $\nabla f(\bar{x})_s \geq 0$, que contradiz nossa hipótese, já que $s \in I$. Finalmente, $I = \emptyset$ e portanto $\nabla f(\bar{x}) \geq 0$.

Passo 4. A existência dos valores de aderência de $\{x^k\}$ é garantida pela limitação do conjunto solução ou qualquer outro conjunto de nível de f. Finalmente, pelas hipóteses (A1)-(A3), a limitação de qualquer conjunto de nível de f implica a limitação de todos os conjuntos de nível. Em particular, $\{x^k;\ f(x^k) \leq f(x^0)\}$ é limitado. Da Proposição 3.2.2, $\{f(x^k)\}$ é convergente, consequentemente é limitada. Portanto $\{x^k\}$ é limitada. ∎

3.4 Taxa de Convergência.

Nesta seção apresentaremos duas proposições relacionadas à taxa de convergência dos Algoritmos I e II e mostraremos a influência do r. De fato, provaremos que sob certas hipóteses, $\{x^k\}$ converge linearmente no interior de $\mathbb{R}^n_+$, com um erro assintótico limitado por uma expressão que depende de r.

Proposição 3.4.1 *Suponha a convergência da sequência* $\{x^k\}$ *gerada por* (3.2)-(3.3)*, usando o Algoritmo I. Seja* x^* *o seu limite. Suponha que (i)* f *é duas vezes continuamente diferenciável e sua Hessiana é positiva definida em* x^**. (ii)* $x^* > 0$*. Então existe uma norma* $\|\cdot\|$ *tal que*

$$\limsup_{k\to+\infty} \frac{\left\|x^{k+1} - x^*\right\|}{\left\|^k - x^*\right\|} \leq \left(1 + \frac{\xi}{(L+1)}\left(\frac{\lambda}{\gamma}\right)^r\right)^{-1},$$

onde $\lambda = min(x^*)$*,* $\gamma = max(x^*)$ *e* ξ *é o menor autovalor de* $\nabla^2 f(x^*)$*.*

Demonstração: Pelo Passo 2 da Proposição 3.2.1, como $x^* > 0$, temos que $\nabla f(x^*) = 0$. Expandindo $\nabla f(x)$ em x^*, resulta

$$\nabla f(x) = \nabla^2 f(x^*)(x - x^*) + v, \tag{3.36}$$

onde $\nabla^2 f(x^*)$ é simétrica positiva definida por (i), e $\|v\| \leq \mu \|x - x^*\|^2$ para algum $\mu > 0$ e todo x em alguma vizinhança de x^*. Pode ser facilmente mostrado que $\mu \geq L/2$, onde L é a constante de Lipschitz de ∇f. Por (3.6) e (3.36), temos que

$$\nabla^2 f(x^*)(x^{k+1} - x^*) = -\beta_k (X^k)^{-r}(x^{k+1} - x^k) - v^k, \tag{3.37}$$

onde v^k satisfaz

$$\left\|v^k\right\| \leq \mu \left\|x^{k+1} - x^*\right\|^2. \tag{3.38}$$

Somando $\beta_k (X^k)^{-r}(x^{k+1} - x^*)$ dos dois lados da equação (3.37), obtemos

$$(\nabla^2 f(x^*) + \beta_k (X^k)^{-r})(x^{k+1} - x^*) = \beta_k (X^k)^{-r}(x^k - x^*) - v^k. \tag{3.39}$$

Como $(\nabla^2 f(x^*) + \beta_k (X^k)^{-r})$ é invertível (pois é simétrica positiva definida), por (3.39) obtemos

$$x^{k+1} - x^* = (\nabla^2 f(x^*) + \beta_k (X^k)^{-r})^{-1}(\beta_k (X^k)^{-r}(x^k - x^*) - v^k). \tag{3.40}$$

A última igualdade implica que

$$x^{k+1} - x^* = (I + \frac{1}{\beta_k}(X^k)^r \nabla^2 f(x^*))^{-1}(x^k - x^*) - (\nabla^2 f(x^*) + \beta_k (X^k)^{-r})^{-1} v^k. \tag{3.41}$$

Pela Proposição 1.1.7, existem matrizes P, D, P_k e D_k tais que

$$(X^*)^{-r} \nabla^2 f(x^*) = PDP^{-1} \quad \text{e} \quad (X^k)^r \nabla^2 f(x^*) = P_k D_k P_k^{-1},$$

onde D e D_k são diagonais positivas. Já que $\lim_{k\to+\infty} (X^k)^{-r} = (X^*)^{-r}$, então P_k e D_k podem ser escolhidos tais que $\lim_{k\to+\infty} P_k = P$ e $\lim_{k\to+\infty} D_k = D$. Consideramos a norma definida por $\|x\| = \|P^{-1}x\|_2$. Aplicando esta norma em (3.41), obtemos

$$\left\|x^{k+1} - x^*\right\| \leq \left\|(I + \frac{1}{\beta_k}(X^k)^r \nabla^2 f(x^*))^{-1}\right\| \left\|x^k - x^*\right\| + \\ + \left\|(\nabla^2 f(x^*) + \beta_k (X^k)^{-r})^{-1}\right\| \left\|v^k\right\|. \tag{3.42}$$

Por (3.42), supondo que μ seja escolhido de tal maneira que (3.38) se mantenha para esta nova norma, resulta

$$\left\|x^{k+1} - x^*\right\| \leq \left\|(I + \frac{1}{\beta_k}(X^k)^r \nabla^2 f(x^*))^{-1}\right\| \left\|x^k - x^*\right\| + \\ + \mu \left\|(\nabla^2 f(x^*) + \beta_k (X^k)^{-r})^{-1}\right\| \left\|x^{k+1} - x^*\right\|. \tag{3.43}$$

Tomando $A = (X^k)^r$, $B = \nabla^2 f(x^*)$, $M = P$, a desigualdade acima e a Proposição 1.1.8 implicam que

$$\left\|x^{k+1} - x^*\right\| \leq \left\|P^{-1}P_k\right\|_2 \left\|P_k^{-1}P\right\|_2 \left(1 + \frac{(min(x^k))^r \xi}{\beta_k}\right)^{-1} \left\|x^k - x^*\right\| + \\ + \mu \left\|P^{-1}\right\|_2 \|P\|_2 \, \xi^{-1} \left\|x^{k+1} - x^*\right\|^2. \tag{3.44}$$

Subtraindo o último termo de ambos os lados da desigualdade, temos

$$\left\|x^{k+1} - x^*\right\| - \mu \left\|P^{-1}\right\|_2 \|P\|_2 \, \xi^{-1} \left\|x^{k+1} - x^*\right\|^2 \leq \left\|P^{-1}P_k\right\|_2 \left\|P_k^{-1}P\right\|_2 \times \\ \times \left(1 + \frac{(min(x^k))^r \xi}{\beta_k}\right)^{-1} \left\|x^k - x^*\right\| \tag{3.45}$$

Assim

$$\left\|x^{k+1} - x^*\right\| (1 - \mu \left\|P^{-1}\right\|_2 \|P\|_2 \, \xi^{-1} \left\|x^{k+1} - x^*\right\|) \leq \left\|P^{-1}P_k\right\|_2 \left\|P_k^{-1}P\right\|_2 \times \\ \times \left(1 + \frac{(min(x^k))^r \xi}{\beta_k}\right)^{-1} \left\|x^k - x^*\right\| \tag{3.46}$$

Como $\lim_{k\to+\infty} x^k = x^*$, então para k suficientemente grande,

$$1 - \mu \left\|P^{-1}\right\|_2 \|P\|_2 \, \xi^{-1} \left\|x^{k+1} - x^*\right\| > 0.$$

Dividindo (3.46) por $(1 - \mu \|P^{-1}\|_2 \|P\|_2 \, \xi^{-1} \left\|x^{k+1} - x^*\right\|) \left\|x^k - x^*\right\| > 0$, obtemos

$$\frac{\left\|x^{k+1} - x^*\right\|}{\|x^k - x^*\|} \leq \frac{\|P^{-1}P_k\|_2 \left\|P_k^{-1}P\right\|_2}{(1 - \mu \|P^{-1}\|_2 \|P\|_2 \, \xi^{-1} \|x^{k+1} - x^*\|)} \left(1 + \frac{(min(x^k))^r \xi}{\beta_k}\right)^{-1} \tag{3.47}$$

Passando ao limite em (3.47) e observando que $\lim_{k\to+\infty} P_k = P$, $\lim_{k\to+\infty} \|P^{-1}P_k\| = \lim_{k\to+\infty} \|P_k^{-1}P\| = 1$, $\lim_{k\to+\infty}(min(x^k))^r = (min(x^*))^r$, e

$$\lim_{k\to+\infty} \beta_k = (L+1)(max(x^*))^r, \tag{3.48}$$

obtemos

$$\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \leq \left(1 + \frac{\xi}{(L+1)}\left(\frac{\lambda}{\gamma}\right)^r\right)^{-1} < 1. \qquad \blacksquare$$

Observamos que, quando β_k é gerado pelo Algoritmo II, então, sob todas as condições da Proposição 3.4.1, a taxa de convergência da sequência $\{x^k\}$ é sublinear, com um erro assintótico limitado por $\left(1 + \frac{\xi}{2}\left(\frac{\lambda}{\gamma}\right)^r\right)^{-1}$. De fato, a prova é análoga a esta apresentada na Proposição 3.4.1, fazendo $\lim_{k\to+\infty} \beta_k = (max(x^*))^r$ no lugar de (3.48).

Agora mostraremos que $\{x^k\}$ converge sublinearmente na fronteira de $\mathbb{R}^n_+$. Para isso, consideramos os conjuntos $I^* := \{l \in \{1, 2, \cdots, n\};\ x_l^* > 0\}$ e $J^* := \{1, 2, \cdots, n\} - I^*$.

Proposição 3.4.2 *Suponha que a sequência* $\{x^k\}$ *gerada por* (3.2)-(3.3)*, usando os Algoritmos I e II, converge para* x^* *e os conjuntos* I^* *e* J^* *são ambos não vazios, i. e., existe algum* l *tal que* $x_l^* = 0$ *e* $x^* \neq 0$*. Então* $\{x^k\}$ *converge sublinearmente.*

Demonstração: Da definição de x^{k+1} e (3.6), temos

$$x^{k+1} = x^k - \frac{1}{\beta_k}(X^k)^r \nabla f(x^{k+1})$$

e portanto, para todo $l \in \mathbb{N}$, segue que

$$x_l^{k+1} = x_l^k \left(1 - \frac{1}{\beta_k}(x^k)_l^{r-1}(\nabla f(x^{k+1})_l\right). \tag{3.49}$$

Seja $l \in J^*$. Assim, $x_l^* = 0$. Por (3.49), obtemos

$$\lim_{k\to+\infty} \frac{|x_l^{k+1} - x_l^*|}{|x_l^k - x_l^*|} = \lim_{k\to+\infty} \left(1 - \frac{1}{\beta_k}(x^k)_l^{r-1}(\nabla f(x^{k+1}))_l\right) \leq 1,$$

como f é continuamente diferenciável e $\lim_{k\to+\infty} \beta_k = \beta^* > 0$. Logo, a l-ésima componente converge sublinearmente. Portanto, o mesmo acontece para toda a sequência $\{x^k\}$. $\blacksquare$

Observação 1 *Observe que, na Proposição 3.4.2, quando tomamos* β_k *gerado pelo Algoritmo I,* $r = 1$*, e supomos a complementaridade estrita de* x^**, obtemos convergência linear para coordenadas nulas da solução (isto é,* $x_l^* = 0$ *implica que* $(\nabla f(x^*))_l > 0$*, e se* $x_l^* > 0$*, então* $(\nabla f(x^*))_l = 0$*). De fato, usando* $r = 1$ *na equação* (3.49)*, ela se torna*

$$x_l^{k+1} = x_l^k \left(1 - \frac{1}{\beta_k}(\nabla f(x^{k+1}))_l\right). \tag{3.50}$$

Por causa da complementaridade estrita de x^**, para* $l \in J^*$*, temos que* $(\nabla f(x^*))_l > 0$ *e portanto* $(\nabla f(x^{k+1}))_l > 0$ *para* k *suficientemente grande. Logo, passando ao limite em* (3.50)*, obtemos*

$$\frac{\left\|x^{k+1} - x^*\right\|}{\left\|x^k - x^*\right\|} \leq 1 - \frac{(\nabla f(x^*))_l}{\beta^*} < 1.$$

Capítulo 4

Implementações.

Neste capítulo implementamos o algoritmo no software Scilab e analisamos os dados obtidos em exemplos de funções quadráticas.

Definição 4.0.3 *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *definida por*

$$f(x) = \langle Qx, x \rangle + \langle c, x \rangle + \gamma,$$

onde $Q \in \mathbb{R}(n \times n)$, $\gamma \in \mathbb{R}$ *e* $c \in \mathbb{R}^n$. *Esta função é chamada função quadrática.*

Consideramos o caso particular em que a função objetivo é dada por

$$f(x) = \frac{1}{2} \langle Qx, x \rangle + \langle c, x \rangle + \gamma,$$

onde Q é simétrica não negativa e γ é um número real. Mostramos que a sequência $\{x^k\}$ tem uma expressão explícita, bem como os parâmetros β_k.

De fato, da definição do Algoritmo, temos que o iterado x^{k+1} é obtido como solução da equação

$$\nabla f(x^{k+1}) + \beta_k (X^k)^{-r} (x^{k+1} - x^k) = 0, \tag{4.1}$$

onde $\nabla f(x^{k+1}) = Qx^{k+1} + c$. Substituindo esta última igualdade em (4.1), obtemos

$$Qx^{k+1} + c + \beta_k (X^k)^{-r} (x^{k+1} - x^k) = 0. \tag{4.2}$$

Por outro lado, como $\nabla f(x^k) = Qx^k + c$, resulta que $c = \nabla f(x^k) - Qx^k$. Dessa forma a equação (4.2) torna-se

$$(Q + \beta_k (X^k)^{-r})(x^{k+1} - x^k) = -\nabla f(x^k). \tag{4.3}$$

Uma vez que Q é não negativa e $\beta_k(X^k)^{-r}$ é positiva, temos que $Q + \beta_k(X^k)^{-r}$ é positiva para todo $k \in \mathbb{N}$, consequentemente, invertível.

Portanto, a equação (4.3) torna-se

$$x^{k+1} = x^k - (Q + \beta_k(X^k)^{-r})^{-1}\nabla f(x^k).$$

Por outro lado, como

$$(Q + \beta_k(X^k)^{-r})^{-1} = \frac{1}{\beta_k}(1 + \frac{1}{\beta_k}(X^k)^r Q)^{-1}(X^k)^r,$$

segue-se que

$$x^{k+1} = x^k - \frac{1}{\beta_k}(I + \frac{1}{\beta_k}(X^k)^r Q)^{-1}(X^k)^r \nabla f(x^k). \tag{4.4}$$

Assim, na definição dos parâmetros β_k em (3.4)-(3.5), a constante $L = \|Q\|_2$.

A demonstração acima foi dada em [29].

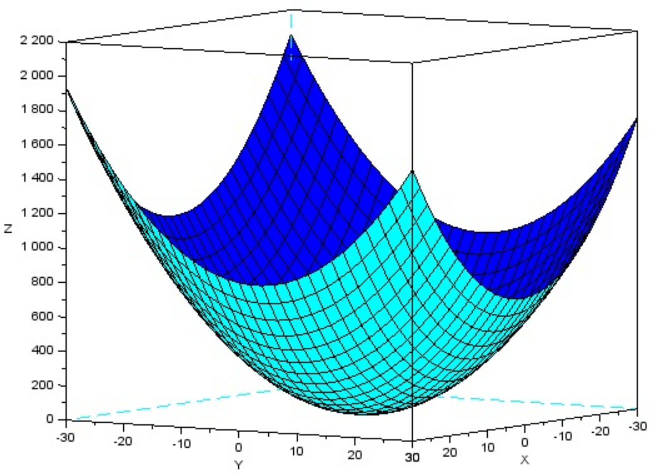
Para implementação, usamos computador com as seguintes configurações: Intel Core i5 2.30 GHz, 3GB de RAM, o sistema operacional Windows 7 SP1 32 bits e o software Scilab versão 5.4.1.

Exemplo 1:

$f : \mathbb{R}^2 \to \mathbb{R}$ definida por

$$f(x) = (x_1 - 1)^2 + (x_2 - 3)^2 + 10,$$

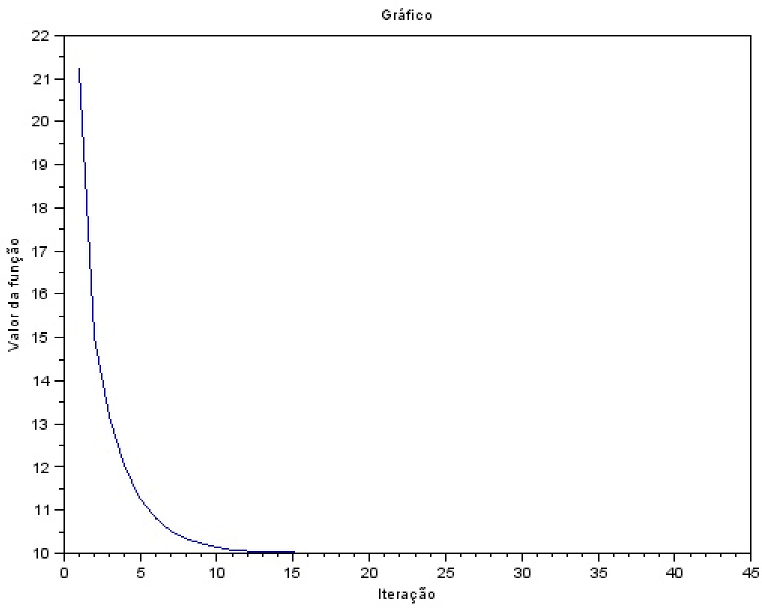
com $x = (x_1, x_2) \in \mathbb{R}^2$.



Fazendo $x^0 = (5, 5)$, $r = 1$ e $\epsilon = 10^{-3}$, temos a seguinte iteração:

k	x_1^k	x_2^k	$\|x^k - x^{k-1}\|$	$f(x^k)$
1.	4.	4.5	1.118034	21.25
2.	3.	4.	1.118034	15.
3.	2.6048159	3.7566032	0.4641255	13.147883
4.	2.3080572	3.5335249	0.3712542	11.995662
5.	2.0631274	3.3767465	0.2908093	11.272178
6.	1.8650176	3.2763147	0.2221127	10.824605
7.	1.6986286	3.1684869	0.1982729	10.51647
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
43.	1.0012305	2.9999965	0.0002051	10.000002
44.	1.0010545	2.9999974	0.0001760	10.000001
45.	1.0008987	2.9999981	0.0001558	10.000001

A tabela acima diz que o programa levou 45 iterações para se aproximar do ponto crítico para $r = 1$.

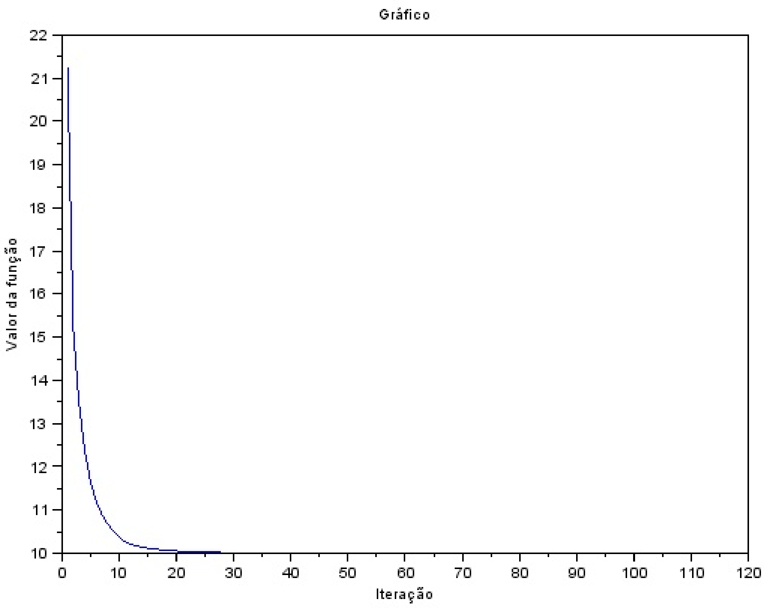


No gráfico podemos observar a relação entre o valor da função e a quantidade de iterações necessárias para se aproximar do mínimo da função.

Tomando $x^0 = (5,5)$, $r = 2$ e $\epsilon = 10^{-3}$, obtemos:

k	x_1^k	x_2^k	$\|x^k - x^{k-1}\|$	$f(x^k)$
1.	4.	4.5	1.118034	21.25
2.	3.0662888	3.9543116	1.0814769	15.18026
3.	2.7340502	3.681588	0.4298380	13.471492
4.	2.4718995	3.4639524	0.3407173	12.38174
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
36.	1.0691836	2.9999001	0.0042313	10.004786
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
100.	1.0015522	3.0000013	0.0001000	10.000002
101.	1.0014522	3.0000012	0.0001000	10.000002
102.	1.0013522	3.0000011	0.0001000	10.00000

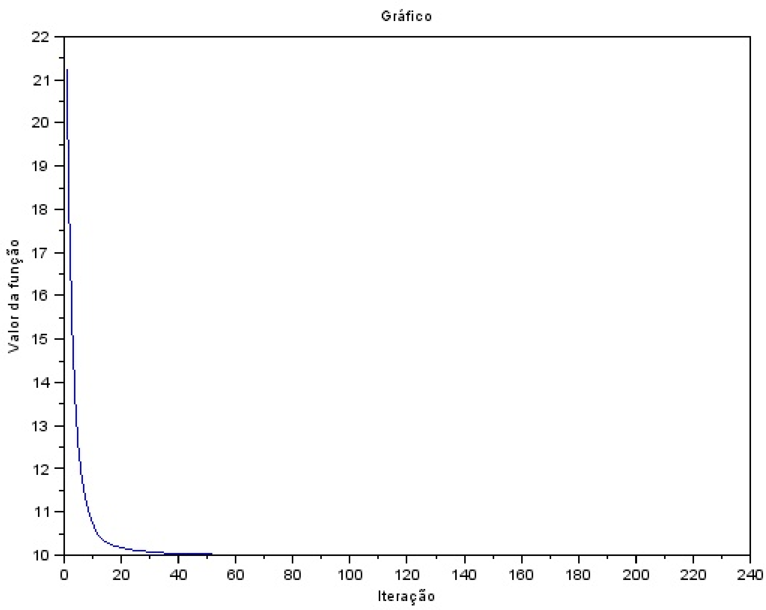
Fazendo $r = 2$ e mantendo os mesmos valor inicial e ϵ, temos que o programa leva 102 iterações para se aproximar da solução.



Agora fazendo $x^0 = (5,5)$, $r = 3$ e $\epsilon = 10^{-3}$, temos:

k	x_1^k	x_2^k	$\|x^k - x^{k-1}\|$	$f(x^k)$
1.	4.	4.5	1.118034	21.25
2.	3.5065646	4.125771	0.6192946	17.550226
3.	3.0995254	3.8415478	0.4964512	15.11621
4.	2.7916155	3.6043257	0.3886937	13.575096
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
80.	1.0680789	2.999602	0.0016916	10.004635
81.	1.0664905	2.9998714	0.0016111	10.004421
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
223.	1.0032286	2.9999982	0.0001000	10.00001
224.	1.0031286	2.9999982	0.0001000	10.00001

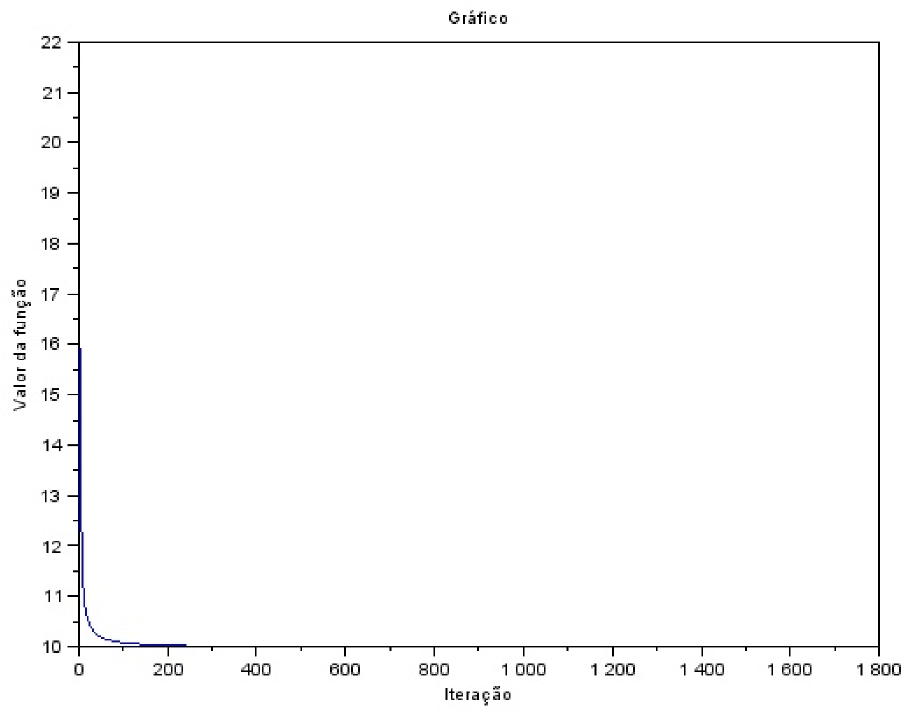
Desta vez o programa levou 224 iterações para se aproximar da solução para os mesmos ϵ e valor inicial dados.



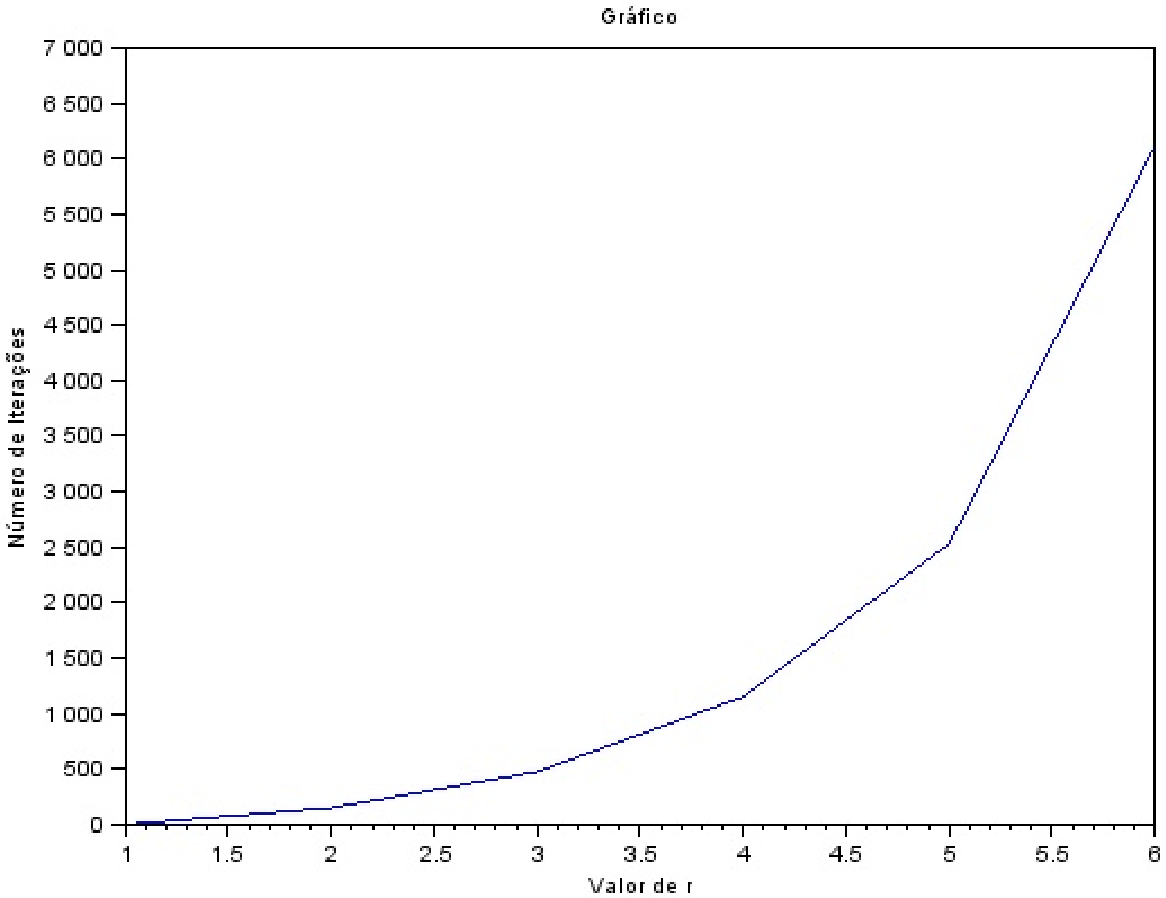
Fazendo $x^0 = (5,5)$, $r = 5$ e $\epsilon = 10^{-3}$, obtemos:

k	x_1^k	x_2^k	$\left\|x^k - x^{k-1}\right\|$	$f(x^k)$
1.	4.	4.5	1.118034	21.25
2.	3.5843507	4.1442809	0.5470835	17.988247
3.	3.2547946	3.8653175	0.4317729	15.832873
4.	2.9964192	3.6518514	0.3351501	14.4106
5.	2.788145	3.4778822	0.2713733	13.425834
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1727.	1.0029179	3.	0.00001	10.000009
1728.	1.0029079	3.	0.00001	10.000008
1729.	1.0028979	3.	0.00001	10.000008
1730.	1.0028879	3.	0.00001	10.000008

Veja que, aumentando o valor de r, o programa leva mais iterações para se aproximar da solução.



No gráfico abaixo vemos a relação entre **r** e o número de iterações necessárias para se aproximar da solução do problema, mantendo o valor inicial e o ϵ (erro).



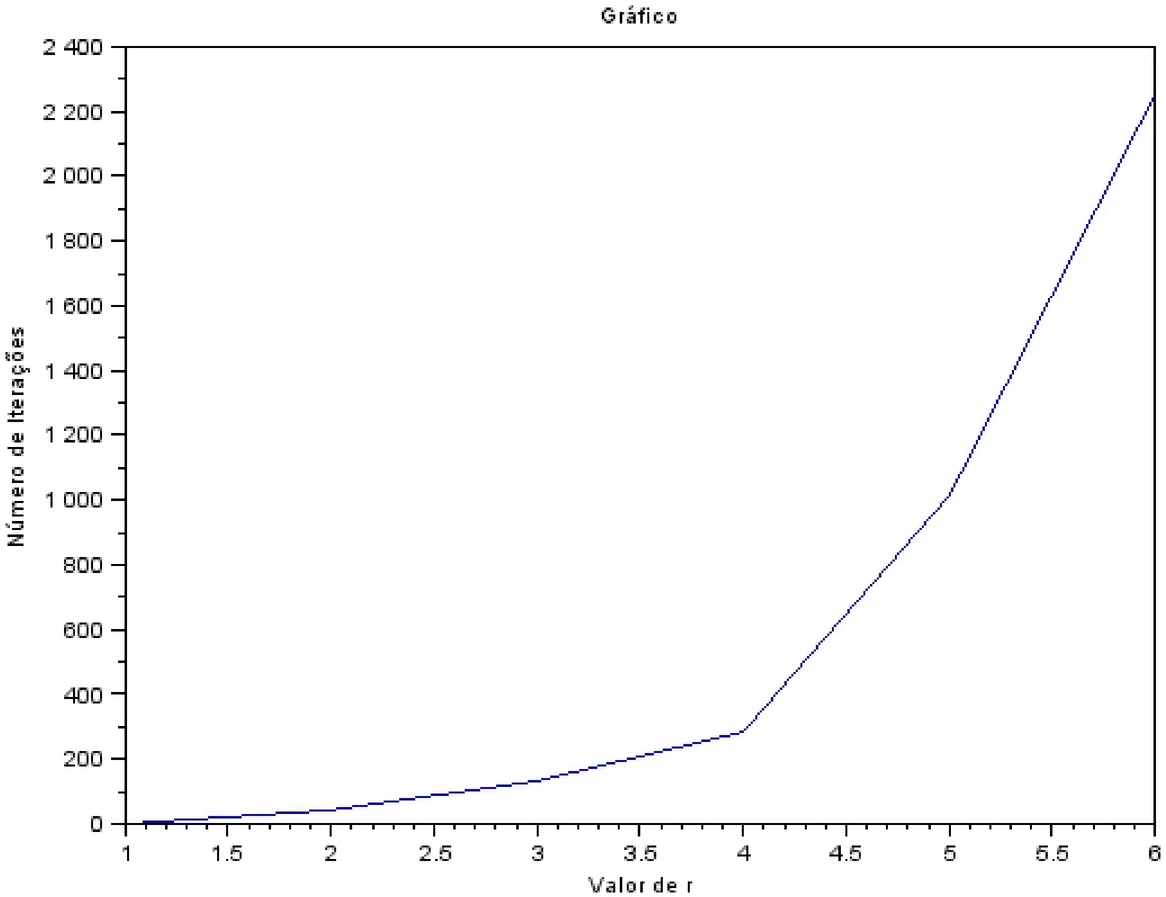
Agora analisamos o que acontece para a mesma função usando o Algoritmo II. Tomando $x^0 = (5, 5)$, $r = 2$ e $\epsilon = 10^{-3}$, temos:

k	x_1^k	x_2^k	$\|x^k - x^{k-1}\|$	$f(x^k)$
1.	3.9407385	4.4703692	1.1842904	20.809929
2.	2.9407385	3.9703692	1.118034	14.708082
3.	2.5485919	3.6314131	0.5183341	12.796819
4.	2.2218818	3.3733222	0.4163537	11.632365
5.	1.9811859	3.1903517	0.3023454	10.99896
6.	1.7932806	3.1162926	0.2019732	10.642818
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
40.	1.0077066	3.0000138	0.0010000	10.000059
41.	1.0067066	3.000012	0.0010000	10.000045
42.	1.0057066	3.0000102	0.0010000	10.000033
43.	1.0047066	3.0000084	0.0010000	10.000022

Mantendo o valor inicial e ϵ e, fazendo $r = 3$, obtemos:

k	x_1^k	x_2^k	$\|x^k - x^{k-1}\|$	$f(x^k)$
1.	3.9407385	4.4703692	1.1842904	20.809929
2.	2.9578971	3.9086626	1.1320298	14.659029
3.	2.6241721	3.6126177	0.4461109	13.013236
4.	2.3578285	3.3922936	0.3456612	11.997593
5.	2.149826	3.2205323	0.2697535	11.370734
6.	1.9761764	3.1391021	0.1917944	10.97227
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
130.	1.0021193	3.0000008	0.0001000	10.000004
131.	1.0020193	3.0000007	0.0001000	10.000004
132.	1.0019193	3.0000007	0.0001000	10.000004
133.	1.0018193	3.0000007	0.0001000	10.000003

No gráfico abaixo vemos a relação entre **r** e o número de iterações necessárias para se aproximar da solução do problema de minimização, para a mesma função, para o mesmo valor inicial e o mesmo ϵ, usando o Algoritmo II.

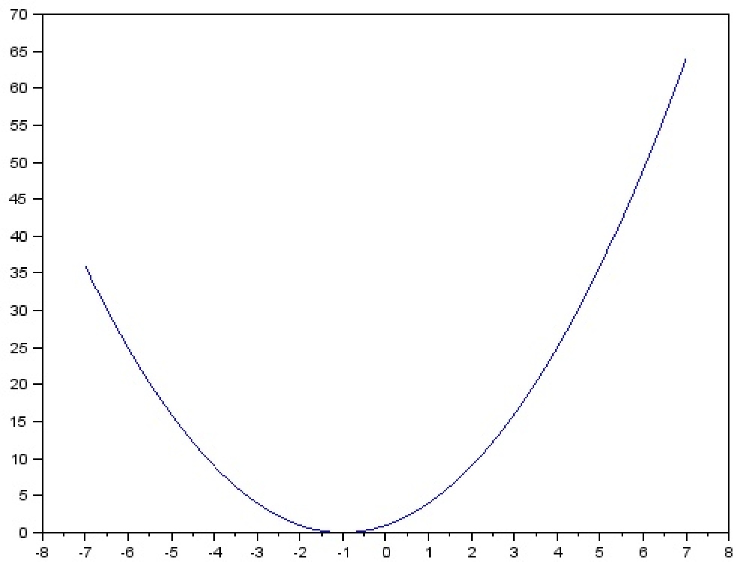


Considere agora uma função em que o mínimo global não se encontra no ortante não negativo.

Exemplo 2:

$$f : \mathbb{R} \to \mathbb{R} \text{ definida por}$$

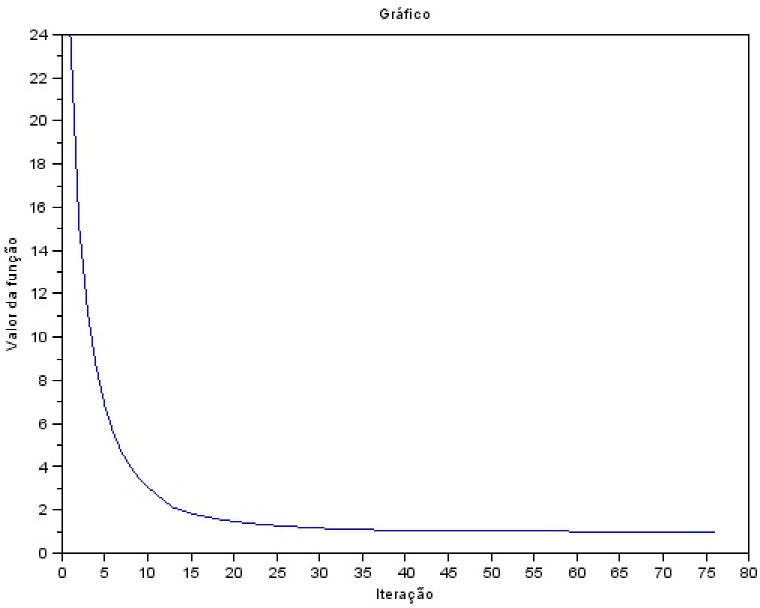
$$f(x) = (x + 1)^2.$$



Veja que o mínimo global dessa função é o ponto $(-1, 0)$, que não se encontra no ortante não negativo, mas o menor valor da função no ortante não negativo é o ponto $(0, 1)$, que será encontrado por qualquer um dos algoritmos.

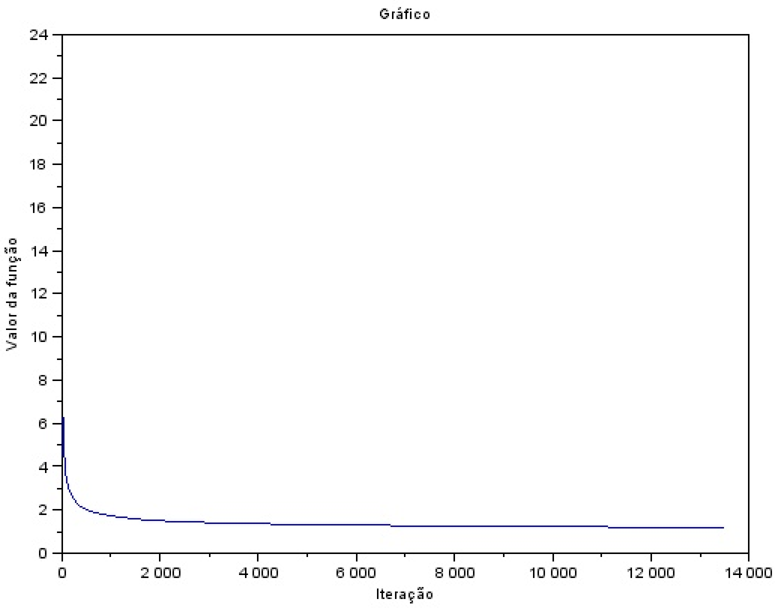
Fazendo $x^0 = 5$, $r = 1$, $\epsilon = 10^{-3}$ e usando o Algoritmo I, obtemos a seguinte iteração:

k	x^k	$\|\|x^k - x^{k-1}\|\|$	$f(x^k)$
1.	3.8948192	1.1051808	23.959255
2.	2.8948192	1.	15.169616
3.	2.3489589	0.5458602	11.215526
4.	1.9363295	0.4126294	8.622031
5.	1.6165816	0.3197480	6.8464991
$\vdots$	$\vdots$	$\vdots$	$\vdots$
73.	0.0012622	0.0001171	1.002526
74.	0.0011569	0.0001053	1.0023152
75.	0.0010569	0.0001	1.002115
76.	0.0009569	0.0001	1.0019148

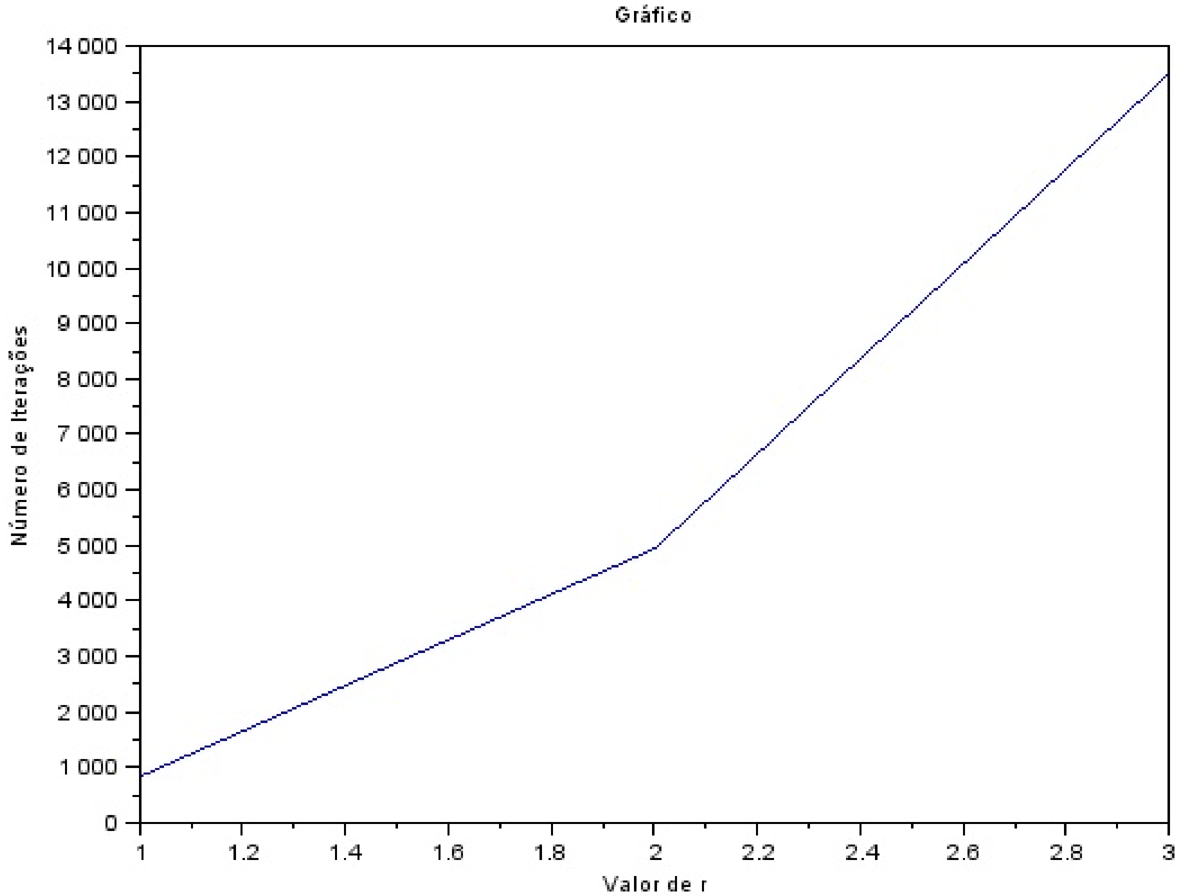


Para $x^0 = 5$, $r = 3$ e $\epsilon = 10^{-3}$, temos:

k	x^k	$\left\|x^k - x^{k-1}\right\|$	$f(x^k)$
1.	3.8900026	1.1099974	23.912125
2.	2.8900026	1.	15.13212
3.	2.5557798	0.3342228	12.64357
4.	2.2969077	0.2588722	10.8696
5.	2.0997007	0.1972069	9.6081446
⋮	⋮	⋮	⋮
13496.	0.0946175	0.0000030	1.1981874
13497.	0.0946144	0.0000030	1.1981807
13498.	0.0946114	0.0000030	1.1981741
13499.	0.0946083	0.0000030	1.1981674
13500.	0.0946053	0.0000030	1.1981608



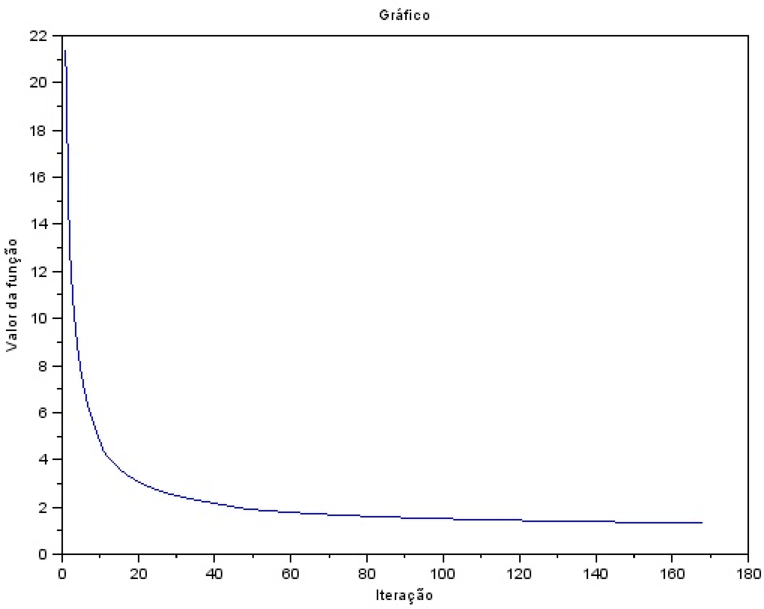
Relação entre **r** e o número de iterações.



Agora usamos o Algoritmo II para a mesma função.

Para $x^0 = 5$, $r = 2$ e $\epsilon = 10^{-1}$:

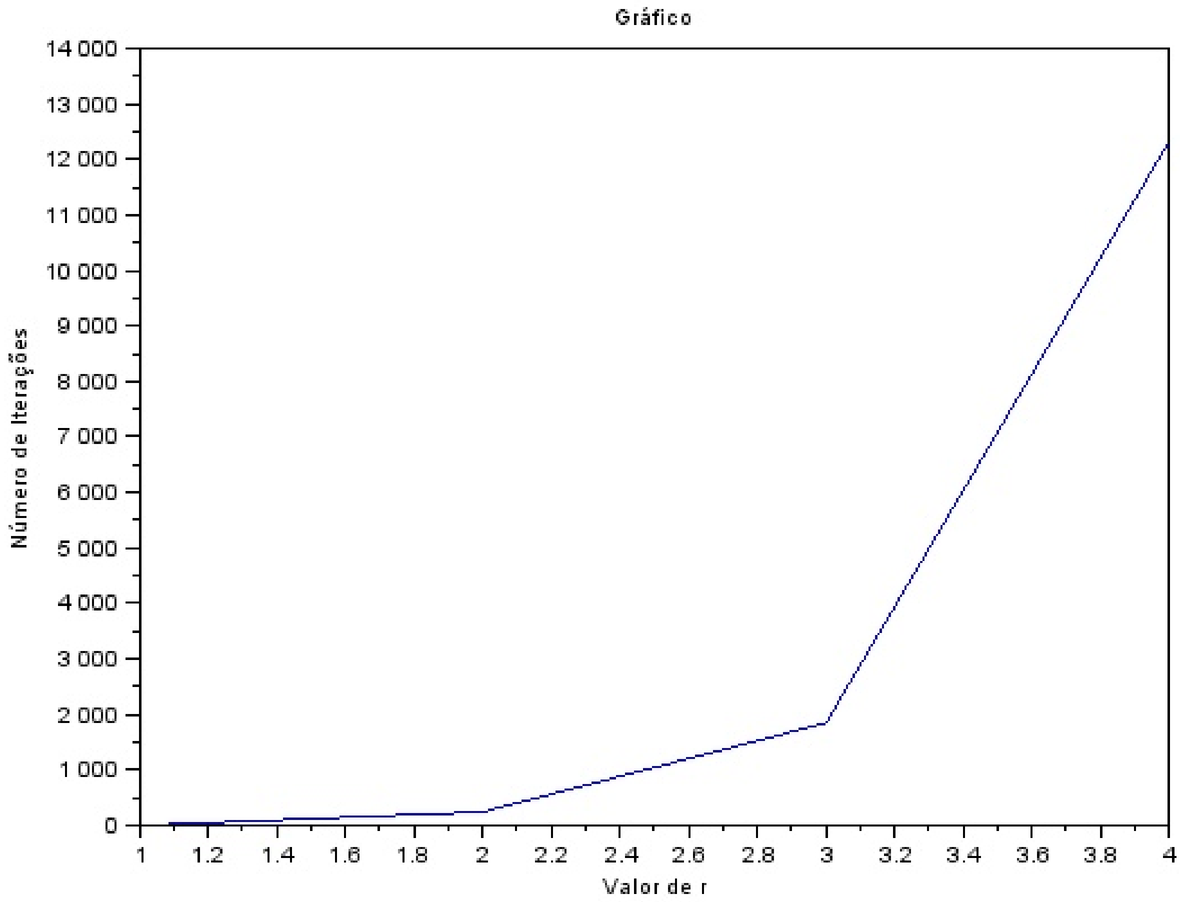
k	x^k	$\|x^k - x^{k-1}\|$	$f(x^k)$
1.	3.6230669	1.3769331	21.372748
2.	2.6230669	1.	13.126614
3.	2.256848	0.3662189	10.607059
4.	1.986861	0.2699871	8.9213385
5.	1.7864691	0.2003919	7.7644098
$\vdots$	$\vdots$	$\vdots$	$\vdots$
165.	0.1510057	0.001	1.3248141
166.	0.1500057	0.001	1.322513
167.	0.1490057	0.001	1.320214
168.	0.1480057	0.001	1.317917



Para $x^0 = 5$, $r = 4$ e $\epsilon = 10^{-1}$, obtemos, usando o Algoritmo II:

k	x^k	$\left\|x^k - x^{k-1}\right\|$	$f(x^k)$
1.	3.6230669	1.3769331	21.372748
2.	3.0773643	0.5457027	16.624899
3.	2.7450681	0.3322962	14.025535
4.	2.5129109	0.2321572	12.340543
5.	2.3229109	0.19	11.041737
⋮	⋮	⋮	⋮
12326.	0.2578028	0.00001	1.5820678
12327.	0.2577928	0.00001	1.5820427
12328.	0.2577828	0.00001	1.5820175
12329.	0.2577728	0.00001	1.5819924
12330.	0.2577628	0.00001	1.5819672

Relação entre **r** e o número de iterações.



Em todos os exemplos acima vimos que, à medida que aumentamos o valor de r, a sequência gerada pelo algoritmo demora mais a convergir para o mínimo da função no ortante não negativo. Além disso no Algoritmo I a sequência gerada converge mais rápido para o mínimo da função no ortante não negativo do que a sequência gerada pelo Algoritmo II.

Referências Bibliográficas

[1] Auslender, A., Teboulle M. e Ben-Tiba, S.: A logarithmic-quadratic proximal method for variational inequalities. Computational Optimization and Aplications 12, 31-40 (1999)

[2] Auslender, A., Teboulle M. e Ben-Tiba, S.: Interior proximal and multiplier methods based on second order homogeneous kernels. Math. Oper. Res. 24, 645-668 (1999)

[3] Bonnans J. F., Gilbert J. C., Lemarechal C., Sagastizabal C. A. A family of variable metric proximal methods. Mathematical Programming 68 (1995) 15-47.

[4] Burachik, R. S. and Iusem, A. N.: A generalized proximal point algorithm for the variational inequality problem in a Hilbert space. SIAM Journal on Optimization 8, 197-216 (1998)

[5] Censor, Y.; Zenios, S. A. The proximal minimization algorithm with D-functions. J. Optim. Theory Appl. 73, 451-464 (1992)

[6] Chen, G.; Teboulle, M. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. SIAM J. Optim. 3, 538-543 (1993)

[7] Cunha, F.G.M., Cruz Neto, J. X., Oliveira, P. R.: A proximal point algorithm with ϕ-divergence to quasiconvex programming, to appear in Optimization.

[8] Filho, D. C. de M. Um Convite à Matemática. 1ª ed. Rio de Janeiro. SBM, 2012.

[9] Filho, D. C. de M. Manual de Redação Matemática: com um dicionário etimológico-explicativo de palavras usadas na matemática e um capítulo especial sobre como se escreve uma dissertação. 1ª ed. Edição do Autor. Campina Grande, PB, 2010.

[10] Eckstein, J.: Approximate iterations in Bregman-function-based proximal algorithms. Math. Program. 83, 113-123 (1998)

[11] Eggermont, P. P. B. Multiplicative Iterative Algorithms for Convex-Programming. Linear Algebra Appl. 130 (1990), 25-42.

[12] Ferreira, O. P. e Oliveira, P. R.: Proximal point algorithm on Riemannian manifolds. Optimization 51, 257-270 (2002)

[13] Horn, R. A. and Johnson, C. R. Matrix Analysis. Cambridge Univ. Press, Cambridge (1985)

[14] Ismailov, A.; Solodov, M. Otimização - volume 1. Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade. 2ªed. Rio de Janeiro. IMPA, 2009.

[15] Iusem, A. N.: An interior multiplicative method for optimization under positivity constraints. Acta Appl. Math. 38, 163-184 (1995)

[16] Iusem, A. N. and Teboulle, M.: On the Convergence Rate of Entropic Proximal Optimization Methods. Comp. Appl. Math., 12(2), 153-168 (1993)

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

[18] Iusem, A. N.; Svaiter, B.; Teboulle, M. Entropy-like proximal methods in convex programming. Math. Oper. Res. 19, 790-814 (1994)

[19] Kiwiell, K. C.: Proximal minimization methods with generalized Bregman functions. SIAM J. Control Optim. 35, 1142-1168 (1997)

[20] Lima, E. L. Álgebra Linear. Coleção Matemática Universitária. 7ª ed. Rio de Janeiro. IMPA, 2008.

[21] Lima, E. L. Análise Real volume 2. Coleção Matemática Universitária. 4ª ed. IMPA, 2009.

[22] Lima, E. L. Curso de Análise vol.2. 10ªed. Coleção Projeto Euclides. IMPA, 2010.

[23] Louredo, A. T.; Oliveira, A. M.; Lima, O. A. Cálculo Avançado. Campina Grande: Eduepb, 2010.

[24] Luenberger, D. G. Introduction to Linear and Nonlinear Programming. Springer. 2nd edition. Stanford University, 2003.

Opér. 4, 154-158 (1970)

[25] Manzano, J. A. N. G.; Oliveira, J. F. de. Algoritmos: Lógica para Desenvolvimento de Programação de Computadores. 21ª ed. São Paulo. Érica, 2008.

[26] Martinet, B.: Regularization d'inequations variationelles par approximations sucessives. Rev. Franc. Inform. Rech. Oper. 4, 154-159 (1970)

[27] Moreau, J. J. Fonctions convexes duales et points proximaux dans un espace hilbertien. Comptes Rendus de l'Aacademie des Sciences de Paris 225 (1962) 2897-2899

[28] Moreau, J. J. Proximité et dualité dans un espace hilbertien. Bulletin de la Societé Mathématique de France, 93, (1965) 273-299.

[29] Oliveira, G.L. e Oliveira, P.R., A new class of proximal point methods with variable metric for optimization under positivity constraints, ES-570/02, 2002, PESC/COPPE, Federal University of Rio de Janeiro.

[30] Oliveira, A. B. de; Boratti, J. C. Introdução a Programação - Algoritmos. Visual Books. Florianópolis, 1999.

[31] Papa Quiroz, E. e Oliveira, P. R.: Proximal Point Methods for Quasiconvex and Convex Functions with Bregman Distances on Hadamard Manifolds. Journal of Convex Analysis 16, (2008)

[32] Rockafellar, R. T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. 14, 877-898 (1976)

[33] Silva, G. J. P. e Oliveira, P. R.: A new class of proximal algorithms for the nonlinear complementarity problema. Optimization and Control With Applications, L. Qi, K. Teo and X. Yang (Eds.), Applied Optimization Series 96, 549-561, Springer (2005)

[34] Souza, S. da S. Uma Classe de Métodos Interior-Proximais com Métrica Variável Para Problemas Convexos e um Algoritmo Interior-Proximal com Distância de Bregman para Otimização Quase-Convexa. Tese - PESC/COPPE, Universidade Federal do Rio de Janeiro.

[35] Souza, S. Oliveira, P. R., Cruz Neto, J. X. e Soubeyran, A.: A Proximal Point Algorithm with Bregman Distances for Quasiconvex Optimiation over the Positive Orthant. European Journal of Operational Research (2009).

[36] Teboulle, M. Entropic proximal mapping with applications to nonlinear programming. Math. Oper. Res. 17, 670-690 (1992)

[37] Teboulle, M. Convergence of proximal-like algorithms. SIAM J. Optim. 7, 1069-1083 (1997)