

VELOCIDADE DE CONVERGÊNCIA DO MÉTODO DO GRADIENTE APLICADO À MINIMIZAÇÃO DE FUNÇÕES QUADRÁTICAS

Tatiane Cazarin da Silva

Programa de Pós Graduação em Métodos Numéricos em Engenharia - PPGMNE
Universidade Federal do Paraná - UFPR
Av. Cel. Francisco H. Dos Santos, Jardim Das Américas, Curitiba - PR.
Universidade Tecnológica Federal do Paraná - UTFPR
BR 369 - Km 0,5, Campo Mourão - PR
tatianecazarin@utfpr.edu.br

Ademir Alves Ribeiro

Universidade Federal do Paraná - UFPR
Av. Cel. Francisco H. Dos Santos, Jardim Das Américas, Curitiba - PR.
ademir.ribeiro@ufpr.br

Gislaine Aparecida Pericaro

Universidade Estadual do Paraná - UNESPAR
Av. Comendador Noberto Marcondes, Campo Mourão - PR
gapericaro@fecilcam.br

RESUMO

Este trabalho propõe uma discussão sobre a velocidade de convergência do método do Gradiente aplicado à minimização de funções quadráticas convexas com busca exata. Apresentamos uma nova taxa de convergência para a sequência gerada pelo algoritmo considerando a norma euclidiana, o que a diferencia das demais taxas discutidas na literatura.

PALAVRAS-CHAVE. Otimização, Gradiente, Convergência.

PM Programação Matemática

ABSTRACT

This paper proposes a discussion on the convergence rate of the Gradient method applied to the minimization of convex quadratic functions with exact line search. We present a new rate of the convergence generated by the algorithm that considers the Euclidean norm, which differentiates it from other rates discussed in the literature.

KEYWORDS. Optimization. Gradient. Convergence.

MP Mathematical Programming

1. Introdução

Neste trabalho estudamos um método clássico empregado na resolução de problemas de otimização irrestrita, cujo objetivo é minimizar uma função sem restrições sobre as variáveis. Matematicamente, um problema de otimização irrestrita é definido como

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && x \in \mathbb{R}^n, \end{aligned} \tag{1}$$

em que assumimos que a função objetivo $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é continuamente diferenciável.

Surge então uma questão inicial e de grande importância para o problema: o que garante a existência de uma solução ótima? A função objetivo f pode ser ilimitada inferiormente, o que faz com que o problema não admita solução ou, permita encontrar apenas minimizadores locais, ou seja, pontos que possuem menor valor funcional apenas quando comparados com pontos situados em uma vizinhança destes.

Os pontos x^* candidatos a minimizadores locais da função objetivo são chamados de pontos estacionários, e satisfazem a condição $\nabla f(x^*) = 0$, denominada condição necessária de otimalidade. Muitos métodos de otimização se contentam em encontrar pontos estacionários para o problema, empregando um processo iterativo que consiste em, a partir de um ponto inicial, determinar uma direção na qual a função decresça para dar o próximo passo. O processo continua até que um ponto estacionário seja obtido.

Dentre os métodos clássicos empregados na resolução do problema (1) temos o Método do Gradiente, o Método de Newton, o Método das Direções Conjugadas e os Métodos de Quase-Newton (IZMAILOV e SOLODOV, 2007; RIBEIRO e KARAS, 2013). Neste trabalho estudamos o Método do Gradiente, também conhecido como Método de Cauchy, que escolhe como direção de descida a direção oposta ao vetor gradiente da função objetivo avaliada no ponto corrente. Após determinar essa direção, uma outra questão que surge é o quanto caminhar nela. Para isso existem métodos denominados métodos de busca, que determinam o tamanho do passo a ser dado na direção de descida. A busca considerada nesse trabalho é a que procura o minimizador da função ao longo desta direção, denominada busca exata. O procedimento empregado pelo método do Gradiente é descrito no seguinte algoritmo.

Algoritmo 1: Algoritmo do Gradiente

Dado: $x^0 \in \mathbb{R}^n$
 $k = 0$
 REPITA enquanto $\nabla f(x^k) \neq 0$
 Defina $d = -\nabla f(x^k)$
 Obtenha $t > 0$ tal que $f(x^k + td) < f(x^k)$
 Faça $x^{k+1} = x^k + td$
 $k = k + 1$

Na Figura 1, apresentamos a trajetória da sequência gerada pelo algoritmo quando aplicado à uma função quadrática convexa. Vale ressaltar que duas direções consecutivas geradas pelo algoritmo são ortogonais.

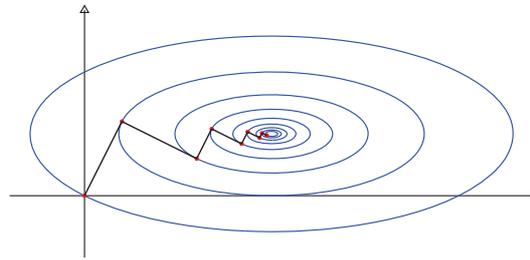


Figura 1: Representação gráfica do Algoritmo do Gradiente.

Ribeiro e Karas (2013) provam que o Algoritmo 1 é globalmente convergente, ou seja, o algoritmo gera uma sequência $(x^k) \subset \mathbb{R}^n$ que converge para um ponto estacionário $x^* \in \mathbb{R}^n$, o que equivale a dizer que $\|x^k - x^*\| \rightarrow 0$. Uma vez estabelecida a convergência global, podemos pensar na convergência local do Algoritmo 1, que diz respeito à velocidade de convergência deste para um ponto estacionário.

Dizemos que uma sequência $(x^k) \subset \mathbb{R}^n$ converge linearmente para $x^* \in \mathbb{R}^n$ quando existe $r \in [0, 1)$, tal que

$$\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \leq r \quad (2)$$

para todo k suficientemente grande, em que r é denominado taxa de convergência. Quanto mais próximo r estiver de zero, mais rápido se dará a convergência para x^* .

Neste trabalho estamos particularmente interessados em estabelecer uma taxa de convergência linear para o Algoritmo 1 quando aplicado a uma função quadrática dada por

$$f(x) = \frac{1}{2}x^T Ax + b^T x + c, \quad (3)$$

com $A \in \mathbb{R}^{n \times n}$, $A \succ 0$, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$, com o tamanho do passo t determinado por busca exata. A taxa de convergência discutida aqui não está presente na literatura e se diferencia das demais, apresentadas adiante, por considerar a norma euclidiana em (2) e a busca exata.

Este trabalho está estruturado da seguinte forma: na Seção 2 são apresentadas diferentes abordagens para a taxa de convergência para o Método do Gradiente, na Seção 3 é discutida a conjectura proposta para a taxa de convergência e na Seção 4 algumas considerações finais.

2. Abordagens para a taxa de convergência do Método do Gradiente

Diversas são as caracterizações da taxa de convergência do método do Gradiente. Alguns trabalhos destacam a forma dita clássica e outros propõem novas taxas, que se menores, garantem uma maior rapidez na convergência e eficiência do método, como os trabalhos de Boyd e Vandenberghe (2004), Freund (2004) e Saad (2005). A seguir são apresentadas duas abordagens sobre a taxa de convergência propostas por Bertsekas (1995) e Karas, Mota e Ribeiro (2005), respectivamente.

Bertsekas (1995) apresenta a taxa de convergência clássica para o método do Gradiente, tomando a função quadrática e a norma euclidiana, dada por:

$$\frac{\|x^{k+1} - x^*\|_2}{\|x^k - x^*\|_2} \leq \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}, \quad (4)$$

onde λ_1 e λ_n são respectivamente, o maior e menor autovalor de A , com tamanho de passo fixo, dado por

$$t = \frac{2}{\lambda_1 + \lambda_n}. \quad (5)$$

Karas, Mota e Ribeiro (2005) verificaram que essa taxa clássica não é satisfeita quando a busca exata é aplicada, ou seja, quando t é calculado por

$$t = \frac{d^T d}{d^T A d}. \quad (6)$$

Considerando $A = \begin{pmatrix} 1 & 0 \\ 0 & 50 \end{pmatrix}$ e $x^0 = \begin{pmatrix} 20 \\ 0 \end{pmatrix}$, mostraram que

$$\frac{\|x^{k+1}\|_2}{\|x^k\|_2} = \frac{\|x^1\|_2}{\|x^0\|_2} > 0,9753 \quad (7)$$

e a taxa da velocidade de convergência é dada por $\frac{\lambda - 1}{\lambda + 1} < 0,9608$. Ou seja, podemos verificar que a taxa de convergência clássica não é satisfeita quando consideramos a norma euclidiana e busca exata, já que pelo exemplo apresentado, a taxa superou a apresentada em (4).

Nesse contexto, Karas, Mota e Ribeiro (2005) estabeleceram uma taxa para a velocidade de convergência do algoritmo do Gradiente com busca exata e norma euclidiana, a qual apresentamos no teorema a seguir.

Teorema 1. *Considere a função quadrática dada em (3) e a sequência gerada pelo algoritmo do Gradiente com busca exata. Então, para todo $k \in \mathbb{N}$, $\|x^{k+1} - x^*\|_2 \leq \gamma \|x^k - x^*\|_2$, onde*

$$\gamma = \sqrt{1 - \frac{\lambda_1}{\lambda_n}}.$$

Tal resultado estabelece uma relação entre a velocidade de convergência associada à norma euclidiana dos termos da sequência gerada pelo Algoritmo 1. Analisando a taxa γ , temos que

$$\sqrt{1 - \frac{\lambda_1}{\lambda_n}} \geq 1 - \frac{\lambda_1}{\lambda_n}, \quad (8)$$

pois $\left(1 - \frac{\lambda_1}{\lambda_n}\right) \in [0, 1)$. Nesse sentido, verificamos a possibilidade de que a taxa de convergência do método proposto seja dada por um valor menor

$$\tilde{\gamma} = 1 - \frac{\lambda_1}{\lambda_n}, \quad (9)$$

resultado esse comprovado por experimentos numéricos e verificado algebricamente para o caso 2×2 , como apresentado na próxima seção.

3. Conjectura para a nova taxa de convergência

Considere o algoritmo do Gradiente aplicado à minimização da função quadrática dada em (3), onde assumimos, sem perda de generalidade, que $b = 0$ e $c = 0$. Assim, dado um

ponto corrente x^k , temos $x^{k+1} = x^k + td$ em que

$$d = -Ax^k. \quad (10)$$

A decomposição espectral da matriz A nos permite escrever $A = PDP^T$ em que D é uma matriz diagonal formada pelos autovalores de A e P é a matriz cujas colunas são os respectivos autovetores. Assim, substituindo $A = PDP^T$ em (10) e definindo $y^k = P^T x^k$, obtemos a seguinte expressão para o tamanho do passo determinado por busca exata

$$t = \frac{d^T d}{d^T A d} = \frac{(y^k)^T D^2 y^k}{(y^k)^T D^3 y^k} \quad (11)$$

e

$$x^{k+1} = x^k + td = P(y^k - tDy^k) = Py^{k+1}. \quad (12)$$

Sendo assim, tomando a matriz diagonal formada pelos autovalores de A , temos:

$$D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & \frac{\lambda_2}{\lambda_1} \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} = \lambda_1 B, \quad (13)$$

onde $\frac{\lambda_2}{\lambda_1} = \lambda$. Substituindo (13) em (11) obtemos $t = \frac{1}{\lambda_1} s$, onde s é o tamanho do passo pela busca exata aplicado na quadrática definida pela matriz B a partir do ponto y e o novo iterando passa a ser definido como

$$y^{k+1} = y^k - tDy^k = y^k - sBy^k. \quad (14)$$

Assim, para o caso 2×2 o nosso objetivo se torna verificar que $\frac{\|y^{k+1}\|_2}{\|y^k\|_2} \leq 1 - \frac{1}{\lambda}$.

Seja $y^k = \begin{pmatrix} p \\ q \end{pmatrix}$ o ponto corrente. Podemos supor, sem perda de generalidade, que $\|y^k\|_2 = 1$. Então, temos que $d = -\begin{pmatrix} p \\ \lambda q \end{pmatrix}$ e o tamanho do passo é dado por $t = \frac{d^T d}{d^T B d} = \frac{p^2 + \lambda^2 q^2}{p^2 + \lambda^3 q^2}$, permitindo escrever:

$$y^{k+1} = y^k + td = \frac{(\lambda - 1)pq}{p^2 + \lambda^3 q^2} \begin{pmatrix} \lambda^2 q \\ -p \end{pmatrix}.$$

Portanto, a nossa conjectura, $\frac{\|y^{k+1}\|_2}{\|y^k\|_2} \leq 1 - \frac{1}{\lambda}$, pode ser reescrita como

$$\frac{p^2 q^2}{(p^2 + \lambda^3 q^2)^2} (p^2 + \lambda^4 q^2) \leq \frac{1}{\lambda^2}.$$

Desde que $p^2 + q^2 = 1$, definimos a função $\varphi : [0, 1] \rightarrow \mathbb{R}$ por

$$\varphi(q) = (1 - q^2 + \lambda^3 q^2)^2 - \lambda^2 q^2 (1 - q^2) (1 - q^2 + \lambda^4 q^2),$$

e o objetivo torna-se provar que $\varphi(q) \geq 0$ para todo $q \in [0, 1]$. Reescrevendo $\varphi(q)$, temos que

$$\varphi(q) = 1 + (2\lambda^3 - \lambda^2 - 2)q^2 + (-2\lambda^3 + 2\lambda^2 + 1)q^4 + (\lambda^2(\lambda^4 - 1))q^6,$$

e, tomando $a = 2\lambda^3 - \lambda^2 - 2$, $b = -2\lambda^3 + 2\lambda^2 + 1$ e $c = \lambda^2(\lambda^4 - 1)$, podemos escrever

$$\varphi(q) = 1 + aq^2 + bq^4 + cq^6. \tag{15}$$

A função φ é apresentada na Figura 2 por meio de dois casos: à esquerda temos a representação para $\lambda = 1.5$ e à direita para $\lambda = 1.05$.

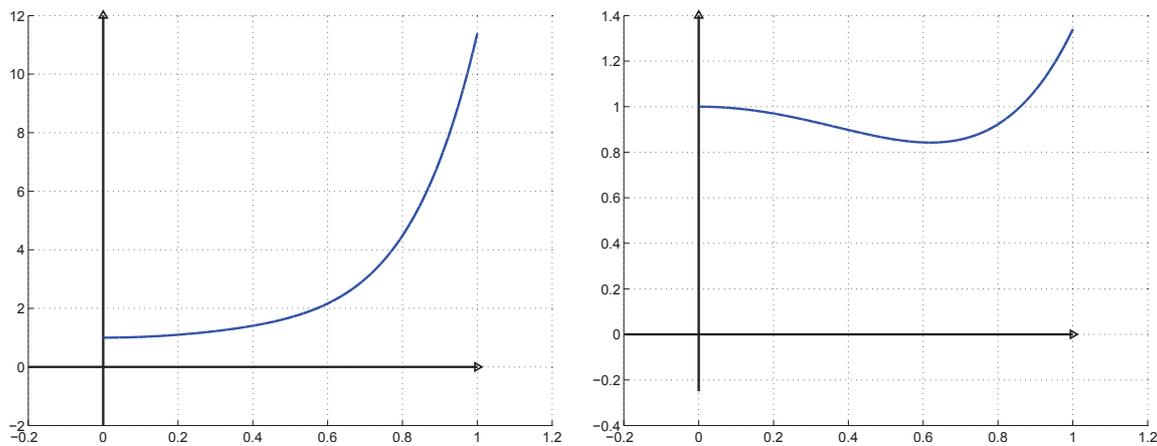


Figura 2: Gráficos de φ para $\lambda = 1.5$ e $\lambda = 1.05$, respectivamente.

Analisando os valores mínimos que a função assume para diferentes valores de λ , podemos verificar que $\min \varphi(q) \geq 0$, para todo λ , conforme pode ser observado na Figura 3.

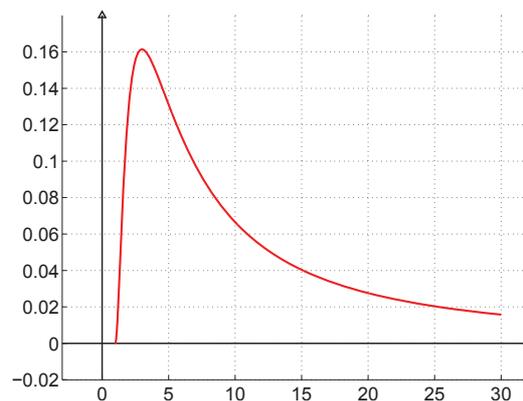


Figura 3: $\min \varphi(q)$

Os testes numéricos realizados indicaram que a função é sempre positiva, e que o mínimo se aproxima de zero para valores de λ suficientemente grandes, reforçando a idéia de que existe uma melhor taxa de convergência para o Método do Gradiente, dada em (9). Dessa forma, vamos agora provar que o mínimo de φ é positivo, sendo discutido com maiores detalhes a seguir.

O estudo será iniciado a partir da análise de algumas propriedades das funções

$$a(\lambda) = 2\lambda^3 - \lambda^2 - 2 \quad \text{e} \quad b(\lambda) = -2\lambda^3 + 2\lambda^2 + 1.$$

Percebemos, por meio da Figura 4, que cada uma das funções apresenta apenas uma raiz, denotadas respectivamente por λ_a e λ_b , com $0 < \lambda_a < \lambda_b$.

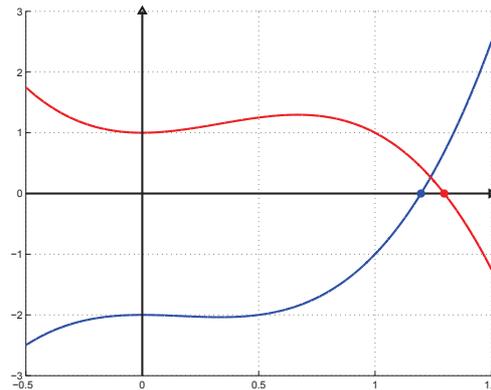


Figura 4: Gráficos de $\lambda \mapsto a(\lambda)$ e $\lambda \mapsto b(\lambda)$.

A fim de verificar que $\varphi(q) \geq 0$ para todo $q \in [0, 1]$, analisaremos a função considerando 3 subintervalos, sendo estes $\lambda \in [\lambda_a, \lambda_b]$, $\lambda < \lambda_a$ e $\lambda > \lambda_b$, apresentados nos Lemas a seguir.

Lema 1. Considere $\lambda \in [\lambda_a, \lambda_b]$ e a corresponde função φ dada por (15). Então $\varphi(q) \geq 0$ para todo $q \in [0, 1]$.

Prova. Primeiro, note que $a, b \geq 0$ para $\lambda_a \leq \lambda \leq \lambda_b$. Além disso, $c = \lambda^2(\lambda^4 - 1) > 0$ para todo $\lambda > 1$. Então,

$$\varphi(q) = 1 + aq^2 + bq^4 + cq^6 > 0$$

para todo $q \in [0, 1]$. □

Agora, vamos ver o que acontece quando $\lambda \geq \lambda_b$. Note que $\varphi'(q) = 2q(a + 2bq^2 + 3cq^4)$ e considere o discriminante

$$\Delta = 4(b^2 - 3ac). \tag{16}$$

Lema 2. Se $\lambda > \lambda_b$, então $\Delta < 0$.

Prova. Primeiramente, afirmamos que

$$9\lambda^6 - 20\lambda^4 + 3\lambda^2 + 12 > 0. \tag{17}$$

De fato, se considerarmos $\mu = \lambda^2$, a expressão (17) pode ser reescrita por $9\mu^3 - 20\mu^2 + 3\mu + 12$ o que é fácil verificar ser positivo. Agora, já que $\lambda_b > 54$, nós temos $\lambda > 54$, $6\lambda^7 > 152\lambda^6$ e $2\lambda^3 > 52\lambda^2$. Isto por sua vez, implica que

$$6\lambda^7 + 2\lambda^3 + 4\lambda + 1 > 152\lambda^6 + 52\lambda^2 + 6.$$

Por (17), concluímos que

$$152\lambda^6 + 52\lambda^2 + 6 > 3\lambda^6 + 10\lambda^4 + \lambda^2.$$

Assim,

$$6\lambda^7 + 2\lambda^3 + 4\lambda + 2 > 3\lambda^6 + 10\lambda^4 + \lambda^2 + 1.$$

Multiplicando por λ^2 , obtemos

$$6\lambda^9 + 2\lambda^5 + 4\lambda^3 + 2\lambda^2 > 3\lambda^8 + 10\lambda^6 + \lambda^4 + \lambda^2 > 3\lambda^8 + 10\lambda^6 + \lambda^4 + 1,$$

o que significa que

$$\Delta = 4(-6\lambda^9 + 3\lambda^8 + 10\lambda^6 - 2\lambda^5 + \lambda^4 - 4\lambda^3 - 2\lambda^2 + 1) < 0,$$

completando a prova. □

Lema 3. Se $\lambda > \lambda_b$, então $\varphi(q) > 0$ para todo $q \in [0, 1]$.

Prova. Pelo Lema 2, concluímos que $a + 2bq^2 + 3cq^4 > 0$ para todo $q \in \mathbb{R}$, uma vez que o discriminante é negativo. Assim,

$$\varphi'(q) = 2q(a + 2bq^2 + 3cq^4) > 0$$

para todo $q > 0$, o que nos permite concluir que a função φ é crescente nesse intervalo. Desde que $\varphi(0) = 1$, concluímos que $\varphi(q) \geq 0$ para todo $q \in [0, 1]$. □

Para finalizar a análise, devemos verificar o caso em que $\lambda \in (1, \lambda_a)$. Isso será feito para provar que o mínimo global de φ é positivo. Veja o gráfico de $\varphi(q)$ para $\lambda = 1.05$, na Figura 2.

Lema 4. Suponha que $\lambda \in (1, \lambda_a)$ e considere o mínimo global positivo, q^* , de φ . Então $\varphi(q^*) \geq 0$.

Prova. Note primeiro que neste caso temos $a < 0$ e $b > 0$. Além disso, $c = \lambda^2(\lambda^4 - 1) > 0$. Então, $\Delta' = b^2 - 3ac > b^2 > 0$. Além do mais, devido ao fato de que q^* é uma raiz de $a + 2bq^2 + 3cq^4 = 0$, temos uma expressão explícita, ou seja,

$$q^* = \sqrt{\frac{-b + \sqrt{\Delta'}}{3c}}.$$

Afirmamos então que

$$1 + a(q^*)^2 > 0. \tag{18}$$

O que é equivalente a provar que $3c - ab > -a\sqrt{\Delta'}$, que é,

$$3c + a^3 - 2ab > 0. \tag{19}$$

Uma vez que $a + b > 0$, temos $-ab > a^2$. Então, usando o fato de que $a > -1$, obtemos

$$3c + a^3 - 2ab > 3c + 2a^2 - 1. \tag{20}$$

Note, na Figura 5, que a função $\lambda \mapsto a^2(\lambda)$ é decrescente e $\lambda \mapsto c(\lambda)$ é crescente em $(1, \lambda_a)$.

Então, para $1 < \lambda \leq 1.067$, temos

$$3c + 2a^2 > 2a^2 > 1 \tag{21}$$

e, para $1.067 \leq \lambda < \lambda_a$

$$3c + 2a^2 > 3c > 1. \tag{22}$$

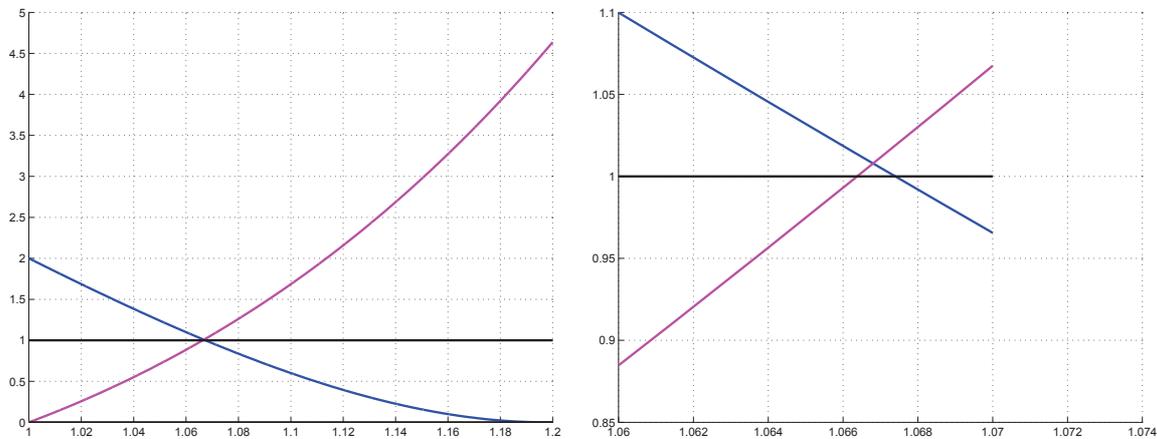


Figura 5: Gráficos de $\lambda \mapsto 2a^2(\lambda)$ e $\lambda \mapsto 3c(\lambda)$.

De (20)-(22), concluímos (19) e conseqüente, (18). Além disso, usando o fato de que $b, c > 0$, obtemos

$$\varphi(q^*) = 1 + a(q^*)^2 + b(q^*)^4 + c(q^*)^6 > 0$$

□

Podemos resumir o que foi provado nos lemas anteriores no seguinte teorema:

Teorema 2. *Considere a função quadrática dada em (3) e a sequência gerada pelo algoritmo do Gradiente com busca exata. Então, para todo $k \in \mathbb{N}$,*

$$\|x^{k+1} - x^*\|_2 \leq \left(1 - \frac{\lambda_1}{\lambda_n}\right) \|x^k - x^*\|_2.$$

Para evidenciar a diferença entre a taxa de convergência proposta pelo Teorema 1 e a apresentada no Teorema 2, em um problema de otimização irrestrito, apresentamos um exemplo que estabelece uma diferença significativa quanto ao número de iterações estimado para a convergência do Método do Gradiente.

Seja a função $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ dada por $f(x) = \frac{1}{2}x^T Ax + b^T x$ onde $A = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$ e $b = -\begin{pmatrix} 4 \\ 15 \end{pmatrix}$. Temos que o minimizador de f satisfaz $\nabla f(x) = 0$, ou seja, $Ax + b = 0$.

Obtemos então que o minimizador é dado por $x^* = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$.

Iniciando o Método do Gradiente em $x^0 = 0$, podemos obter uma estimativa para o número de iterações necessárias para garantir que a distância do ponto corrente ao minimizador seja inferior, por exemplo, a 10^{-6} . Para tanto, observe que

$$\|x^{k+1} - x^*\|_2 \leq \tilde{\gamma}^k \|x^k - x^*\|_2, \quad (23)$$

e portanto

$$\|x^{k+1} - x^*\|_2 \leq \tilde{\gamma}^k \|x^0 - x^*\|_2 = 5\tilde{\gamma}^k. \quad (24)$$

Tomando $\tilde{\gamma} = 1 - \frac{\lambda_1}{\lambda_n}$ e impondo $5\tilde{\gamma}^k < 10^{-6}$ obtemos

$$\tilde{k} > -\frac{\ln(5 \times 10^6)}{\ln(\tilde{\gamma})} \approx \frac{15,424948}{0,223143} > 69. \quad (25)$$

Por outro lado, se considerarmos $\gamma = \sqrt{1 - \frac{\lambda_1}{\lambda_n}}$ obteremos

$$k > -\frac{\ln(5 \times 10^6)}{\ln(\gamma)} \approx \frac{15,424948}{0,111571} > 138. \quad (26)$$

Podemos então verificar que a nova taxa é alcançada com aproximadamente metade do número de iterações necessárias quando consideramos a taxa dada no Teorema 1.

4. Considerações finais

O Método do Gradiente aplicado à minimização de funções quadráticas é linearmente convergente. A taxa que verifica a convergência local do algoritmo é apresentada e discutida na literatura fazendo associações aos autovalores da matriz A . Neste trabalho apresentamos uma nova taxa de convergência para o algoritmo do Gradiente, com busca exata, aplicado à minimização de funções quadráticas convexas, na qual identificamos ser satisfeita em um menor número de iterações quando comparado às taxas apresentadas na literatura associadas à velocidade de convergência do Método do Gradiente.

Referências

- Bertsekas, D. P.**, *Nonlinear Programming*, Athena Scientific, 1995.
- Freund, R. M.**, *The Steepest Descent Algorithm for Unconstrained Optimization and a Bisection Line-search Method*, Massachusetts Institute of Technology, 2004.
- Boyd, S. e Vandenberghe, L.**, *Convex Optimization*, Cambridge University, 2004.
- Izmailov, A. e Solodov, M.**, *Otimização*, vol. 2, IMPA, 2007.
- Karas E. W., Mota A. M. e Ribeiro A. A.** (2005), On the convergence rate of the Cauchy algorithm in the l2 norm, *Technical Report*, Dep. Mathematics, Federal University of Paraná.
- Ribeiro, A. A. e Karas E. W.**, *Um Curso de Otimização*, Cengage Learning Editora, a ser lançado, 2013.
- Saad, Y.**, *Iterative Methods for Sparse Linear Systems*, PWS Publishing Company, 2000.