Proceeding Series of the Brazilian Society of Applied and Computational Mathematics, Vol. 2, N. 1, 2014.
Trabalho apresentado no CMAC-Sul, Curitiba-PR, 2014.

Uma nova taxa de convergência para o Método do Gradiente aplicado à minimização de funções quadráticas

Tatiane Cazarin da Silva
Programa de Pós Graduação em Métodos Numéricos em Engenharia - PPGMNE/UFPR
81531-980, Curitiba, PR
UTFPR - Departamento de Matemática
Campus Campo Mourão
87301-899, Campo Mourão, PR
E-mail: tatianecazarin@utfpr.edu.br

Ademir Alves Ribeiro
UFPR - Departamento de Matemática
81531-980, Curitiba, PR
E-mail: ademir.ribeiro@ufpr.br

Gislaine Aparecida Periçaro
UNESPAR - Departamento de Matemática
Campus Campo Mourão
87303-100, Campo Mourão, PR
E-mail: gapericaro@fecilcam.br

Resumo: *Realizamos neste trabalho um estudo sobre a velocidade de convergência do Método do Gradiente aplicado à minimização de funções quadráticas convexas com busca exata. Destacamos a ocorrência de uma nova taxa de convergência para a sequência gerada pelo algoritmo, diferente das apresentadas na literatura, a qual demonstramos para o caso 2 × 2.*

Palavras-chave: *Gradiente, velocidade, convergência*

1. Introdução

Um problema de otimização irrestrita, matematicamente, é definido como

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

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

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.

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, que escolhe como direção de descida a direção oposta ao vetor gradiente da função objetivo avaliada no ponto corrente e definimos para o passo, a busca exata, que procura o minimizador da função ao longo desta direção. O algoritmo do Gradiente encontra-se estruturado a seguir.

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 $x^{k+1} = x^k + td$
 $k = k + 1$

A convergência global do Algoritmo 1 foi comprovado por Ribeiro e Karas (2013). Estamos particularmente interessados em estabelecer uma taxa de convergência linear para o algoritmo do Gradiente, quando aplicado a uma função quadrática dada por

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

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 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 realizadas algumas considerações finais.

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

Diversas são as taxas de convergência para o Método do Gradiente encontradas na literatura, como por exemplo nos trabalhos de Boyd e Vandenberghe (2004), Freund (2004) e Saad (2005). Destacamos duas abordagens em especial, 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:

$$\frac{\left\|x^{k+1} - x^*\right\|_2}{\left\|x^k - x^*\right\|_2} \leq \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}, \tag{3}$$

onde λ_1 e λ_n são respectivamente, o maior e menor autovalor de A, com tamanho de passo fixo, dado por $t = \dfrac{2}{\lambda_1 + \lambda_n}$.

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 $t = \dfrac{d^T d}{d^T Ad}$. Nesse contexto, estabeleceram então uma taxa para a velocidade de convergência do algoritmo do Gradiente com busca exata e norma euclidiana, dada por $\left\|x^{k+1} - x^*\right\|_2 \leq \gamma \left\|x^k - x^*\right\|_2$, onde $\gamma = \sqrt{1 - \dfrac{\lambda_1}{\lambda_n}}$. Esse 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

$$\sqrt{1 - \frac{\lambda_1}{\lambda_n}} \geq 1 - \frac{\lambda_1}{\lambda_n}, \tag{4}$$

pois $\left(1 - \dfrac{\lambda_1}{\lambda_n}\right) \in [0, 1)$. Nesse sentido, verificamos por meio de experimentos numéricos que a taxa de convergência do método proposto pode ser dada por um valor menor

$$\widetilde{\gamma} = 1 - \frac{\lambda_1}{\lambda_n}, \tag{5}$$

DOI: 10.5540/03.2014.002.01.0119

demonstrado algebricamente para o caso 2×2, conforme apresentado na seção a seguir.

3. A nova taxa de convergência

Considere o algoritmo do Gradiente aplicado à minimização da função quadrática dada em (2), para $n = 2$, 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

$$d = -Ax^k. \tag{6}$$

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 (6) e definindo $y^k = P^T x^k$, obtemos a seguinte expressão para o tamanho do passo determinado por busca exata

$$t = \frac{d^T d}{d^T A d} = \frac{(y^k)^T D^2 y^k}{(y^k)^T D^3 y^k} \tag{7}$$

e

$$x^{k+1} = x^k + td = P(y^k - tDy^k) = Py^{k+1}. \tag{8}$$

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

$$D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & \dfrac{\lambda_2}{\lambda_1} \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} = \lambda_1 B, \tag{9}$$

onde $\dfrac{\lambda_2}{\lambda_1} = \lambda$. Substituindo (9) em (7) obtemos $t = \dfrac{1}{\lambda_1} s$, 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

$$y^{k+1} = y^k - tDy^k = y^k - sBy^k. \tag{10}$$

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

Seja $y^k = \begin{pmatrix} p \\ q \end{pmatrix}$ o ponto corrente. Podemos supor, sem perda de generalidade, que $\left\|y^k\right\|_2 = 1$. Então, temos que $d = -\begin{pmatrix} p \\ \lambda q \end{pmatrix}$ e o tamanho do passo é dado por $t = \dfrac{d^T d}{d^T B d} = \dfrac{p^2 + \lambda^2 q^2}{p^2 + \lambda^3 q^2}$, 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, $\dfrac{\left\|y^{k+1}\right\|_2}{\|y^k\|_2} \leq 1 - \dfrac{1}{\lambda}$, 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 $\varphi : [0, 1] \to \mathbb{R}$ 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

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

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

$$\varphi(q) = 1 + aq^2 + bq^4 + cq^6. \tag{11}$$

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 1.

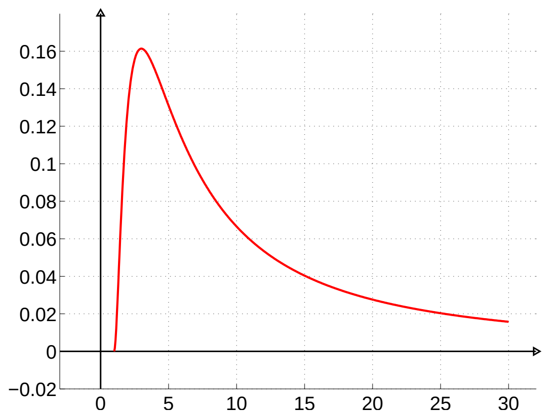

Figura 1: $\min \varphi(q)$

Dessa forma, vamos agora provar que o mínimo de φ é positivo partindo da análise de algumas propriedades das funções

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

Inicialmente, percebemos que cada uma das funções apresenta apenas uma raiz, denotadas respectivamente por λ_a e λ_b, com $0 < \lambda_a < \lambda_b$. 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* (11)*. 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

$$\Delta = 4(b^2 - 3ac). \tag{12}$$

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

Prova. Primeiramente, afirmamos que

$$9\lambda^6 - 20\lambda^4 + 3\lambda^2 + 12 > 0. \tag{13}$$

De fato, se considerarmos $\mu = \lambda^2$, a expressão (13) 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 (13), 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.

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

$$1 + a(q^*)^2 > 0. \tag{14}$$

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

$$3c + a^3 - 2ab > 0. \tag{15}$$

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

$$3c + a^3 - 2ab > 3c + 2a^2 - 1. \tag{16}$$

Note, na Figura 2, 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

$$3c + 2a^2 > 2a^2 > 1 \tag{17}$$

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

$$3c + 2a^2 > 3c > 1. \tag{18}$$

De (16)-(18), concluímos (15) e consequentente, (14). Além disso, usando o fato de que $b, c > 0$, obtemos

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

□

DOI: 10.5540/03.2014.002.01.0119

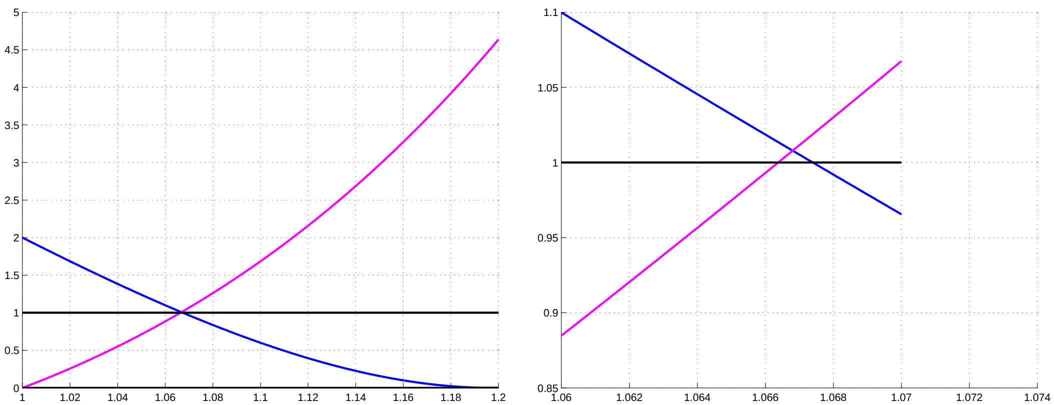

Figura 2: Gráficos de $\lambda \mapsto 2a^2(\lambda)$ e $\lambda \mapsto 3c(\lambda)$.

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

Teorema 1. *Considere a função quadrática dada em (2), com $n = 2$, e a sequência gerada pelo algoritmo do Gradiente com busca exata. Então, para todo $k \in \mathbb{N}$,*

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

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. A partir de inúmeros testes computacionais, conjecturamos que o Teorema 1 seja válido para todo $n \in \mathbb{N}$.

Referências

[1] A. A. Ribeiro e E. W. Karas, "Otimização Contínua: aspectos teóricos e computacionais", Cengage Learning Editora, 2013.

[2] A. Izmailov e M. Solodov, "Otimização", IMPA, 2007.

[3] D. P. Bertsekas, "Nonlinear Programing", Athena Scientific, 1995.

[4] E. W. Karas, A. M. Mota e A. A. Ribeiro, On the convergence rate of the Cauchy algorithm in the l2 norm, em "Technical Report", Dep. Mathematics, Federal University of Paraná, 2005.

[5] R. M. Freund, The Steepest Descent Algorithm for Unconstrained Optimization and a Bisection Line-search Method, *Journal of Massachusetts Institute of Technology*, 2004.

[6] S. Boyd e L. Vandenbergue, "Convex Optimization", Cambridge University, 2004.

[7] Y. Saad, "Iterative Methods for Sparse Linear Systems", PWS Publishing Company, 2000.

DOI: 10.5540/03.2014.002.01.0119