

# Método dos gradientes (ou método de máxima descida)

Marina Andretta

ICMC-USP

13 de setembro de 2010

O **método dos gradientes** (ou **método de máxima descida**) para resolução de problemas de minimização irrestrita é um método de busca linear que usa a direção

$$p_k = -\nabla f_k.$$

(1)

Note que, neste caso,

$$p_k^T \nabla f_k = -\|\nabla f_k\|_2^2 < 0.$$

Ou seja,  $p_k$  é uma direção de descida.

Além disso,

$$\cos \theta_k = \frac{-\nabla f_k^T p_k}{\|\nabla f_k\| \|p_k\|} = 1.$$

Pela [condição de Zoutendijk](#), temos que

$$\lim_{k \rightarrow \infty} \|\nabla f_k\| = 0.$$

Portanto, o método dos gradientes produz uma sequência de gradientes que converge para zero, dado que ele use uma busca linear que satisfaça as condições de Wolfe.

Pode-se mostrar também que a sequência gerada pelo método dos gradientes converge quando o tamanho de passo é calculado usando *backtracking* e satisfaz a condição de Armijo.

# Ordem de convergência

Para estudar a ordem de convergência do **método dos gradientes**, vamos nos concentrar primeiro no caso ideal, no qual a função objetivo é uma quadrática e a busca linear é exata.

Considere, então, o problema

$$\text{Minimizar } f(x) = \frac{1}{2}x^T Qx - b^T x, \quad (2)$$

com  $Q$  simétrica e definida positiva (veja figura 3.7 de Numerical Optimization, de J. Nocedal e S. J. Wright).

# Ordem de convergência

Neste caso, o gradiente de  $f$  é dado por

$$\nabla f(x_k) = Qx - b.$$

Assim, o **minimizador**  $x^*$  é a solução única do sistema linear  $Qx = b$ .

# Ordem de convergência

Calculamos o tamanho de passo  $\alpha_k$  que minimiza a função

$$\phi(\alpha) = f(x_k + \alpha p_k) = f(x_k - \alpha \nabla f_k).$$

Derivando  $\phi$  em relação a  $\alpha$ , temos que

$$\alpha_k = \frac{\nabla f_k^T \nabla f_k}{\nabla f_k^T Q \nabla f_k}.$$



# Ordem de convergência

Se usamos o minimizador exato  $\alpha_k$  como tamanho de passo no método dos gradientes para a função  $f$ , temos que

$$x_{k+1} = x_k - \left( \frac{\nabla f_k^T \nabla f_k}{\nabla f_k^T Q \nabla f_k} \right) \nabla f_k. \quad (3)$$

Como  $\nabla f_k = Qx_k - b$ , esta equação fornece uma fórmula fechada para  $x_{k+1}$  em relação a  $x_k$ .

# Ordem de convergência

Para quantificar a ordem de convergência, introduzimos a norma

$$\|x\|_Q^2 = x^T Q x.$$

Usando a relação  $Qx^* = b$ , temos que

$$\frac{1}{2} \|x - x^*\|_Q^2 = f(x) - f(x^*). \quad (4)$$

# Ordem de convergência

Usando a igualdade (3) e o fato de que  $\nabla f_k = Q(x_k - x^*)$ , temos que

$$\|x_{k+1} - x^*\|_Q^2 = \left\{ 1 - \frac{(\nabla f_k^T \nabla f_k)^2}{(\nabla f_k^T Q \nabla f_k)(\nabla f_k^T Q^{-1} \nabla f_k)} \right\} \|x_k - x^*\|_Q^2.$$

Esta equação descreve o decréscimo exato de  $f$  a cada iteração.

Como é difícil interpretar a expressão dentro das chaves, é mais conveniente limitá-la em termos do número de condição do problema.

**Teorema 1:** Quando o método dos gradientes com busca linear exata é aplicado na resolução do problema de minimizar a função fortemente convexa (2), a norma do erro (4) satisfaz

$$\|x_{k+1} - x^*\|_Q^2 \leq \left( \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)^2 \|x_k - x^*\|_Q^2, \quad (5)$$

onde  $0 < \lambda_1 \leq \dots \leq \lambda_n$  são os autovalores de  $Q$ .

# Ordem de convergência

As inequações (5) e (4) mostram que os valores  $f_k$  da função convergem ao mínimo com uma **razão linear**.

Um caso especial deste resultado é aquele em que **todos os autovalores da matriz  $Q$  são iguais** ( $Q$  é um múltiplo da matriz identidade). Neste caso, a convergência do método é obtida em **apenas uma iteração**.

Quando  $Q$  é um múltiplo da identidade, as curvas de nível de  $f$  são círculos e a direção de máxima descida sempre aponta para o minimizador.

# Ordem de convergência

Em geral, conforme o número de condição  $\kappa(Q) = \lambda_n/\lambda_1$  aumenta, as curvas de nível da quadrática se tornam mais alongadas e o zigue-zague obtido pela sequência de pontos calculada pelo método dos gradientes se pronuncia. Neste caso, a inequação (5) implica que a convergência se degrada.

Apesar de a inequação (5) ser um limitante para o pior caso do algoritmo, ela fornece uma indicação precisa do comportamento do algoritmo quando  $n > 2$ .

**Teorema 2:** *Suponha que  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  tenha segunda derivada contínua e que os iterandos gerados pelo método dos gradientes com busca linear exata converjam a um ponto  $x^*$  no qual a Hessiana  $\nabla^2 f(x^*)$  seja definida positiva.*

*Então,*

$$f(x_{k+1}) - f(x^*) \leq \left( \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)^2 [f(x_k) - f(x^*)],$$

*onde  $0 < \lambda_1 \leq \dots \leq \lambda_n$  são os autovalores de  $\nabla^2 f(x^*)$ .*

# Ordem de convergência

Em geral, não podemos esperar que a ordem de convergência melhore se uma busca linear inexata for usada. Portanto, o Teorema 2 mostra que o método dos gradientes pode ter a convergência inaceitavelmente lenta mesmo quando a Hessiana da função objetivo é bem-condicionada.

Por exemplo, se  $\kappa(Q) = 800$ ,  $f(x_1) = 1$  e  $f(x^*) = 0$ , o Teorema 2 sugere que o valor da função ainda será aproximadamente 0.08 depois de mil iterações do método dos gradientes.