Convergencia

José Luis Morales

Departamento de Matemáticas. ITAM. 2015.

Tasa de convergencia (local)

Se dice que $\{x_k\}$, una sucesión de puntos en $\mathbb{R}^n$, converge a x^*:

- Q-linealmente, si existe una constante r en $(0,1)$ tal que

$$\frac{||x_{k+1} - x^*||}{||x_k - x^*||} \leq r$$

  para toda k suficientemente grande.

- Q-superlinealmente, si

$$\lim_{k\to\infty} \frac{||x_{k+1} - x^*||}{||x_k - x^*||} = 0.$$

- Q-cuadráticamente, si existe una constante positiva M, tal que

$$\frac{||x_{k+1} - x^*||}{||x_k - x^*||^2} \leq M$$

  para toda k suficientemente grande.

Ejemplos

- El método de gradiente converge linealmente.

$$B_k = I.$$

$$\left\{1 + \frac{1}{2^k}\right\} \to 1.$$

- Los métodos cuasi-Newton convergen superlinealmente

$$\left\{1 + \frac{1}{k^k}\right\} \to 1.$$

- El método de Newton converge cuadráticamente.

$$B_k = \nabla^2 f(x_k)$$

$$\left\{1 + \frac{1}{2^{2^k}}\right\} \to 1.$$

Convergencia global

$$x_{k+1} = x_k + \alpha_k p_k \qquad (1)$$

- direcciones de descenso
  - $(-)$ gradiente
  - Newton (Hessiana spd)
  - gradiente conjugado
  - cuasi-Newton: DFP, BFGS, SR1
- pasos *apropiados*
  - condiciones de Wolfe
  - condiciones de Goldstein

Teorema de Zoutendijk

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

Consideremos una iteración de la forma (1), en donde p_k es una dirección de descenso y α_k satisface las condiciones de Wolfe. Supongamos que f es acotada por abajo en $\mathbb{R}^n$ y que f es continuamente diferenciable en un abierto $\mathcal{N}$ que contiene al conjunto

$$\mathcal{L} = \{x \,:\, f(x) \leq f(x_0)\},$$

en donde x_0 es el punto inicial de la iteración. Supongamos también que ∇f es una función de Lipschitz en $\mathcal{N}$. Entonces

$$\sum_{k\geq 0} \cos^2\theta_k\, ||\nabla f_k||^2 < \infty.$$

Continuidad de Lipschitz.

Existe una constante $L > 0$ tal que

$$||\nabla f(x) - \nabla f(\tilde{x})|| \leq L||x - \tilde{x}||, \quad \forall x, \tilde{x} \in \mathcal{N}.$$

<u>Prueba</u>: de la 2da condición de Wolfe

$$(\nabla f_{k+1} - \nabla f_k)^T p_k \geq (c_2 - 1)\nabla f_k^T p_k$$

utilizando la continuidad de Lipschitz, (1) + desigualdad de Schwartz

$$(\nabla f_{k+1} - \nabla f_k)^T p_k \leq \alpha_k L||p_k||^2$$

combinando los resultados

$$\alpha_k \geq \frac{c_2 - 1}{L} \frac{\nabla f_k^T p_k}{||p_k||^2}$$

sustituyendo en la 1ra condición de Wolfe

$$f_{k+1} \leq f_k - c_1 \frac{1 - c_2}{L} \frac{(\nabla f_k^T p_k)^2}{||p_k||^2}$$

continuación ...

$$f_{k+1} \leq f_k - c\cos^2\theta_k ||\nabla f_k||^2, \quad c = c_1(1-c_2)/L.$$

Finalmente, sumando sobre los índices $\leq k$ se tiene que

$$f_{k+1} \leq f_0 - c\sum_{j=0}^{k} \cos^2\theta_j ||\nabla f_j||^2$$

puesto que f es acotada por abajo, entonces la cantidad $f_0 - f_{k+1}$ está acotada por un número positivo para toda k. Por lo tanto tenemos que

$$\sum_{k\geq 0} \cos^2\theta_k \, ||\nabla f_k||^2 < \infty.$$

$\square$

Por lo tanto

$$\cos^2\theta_k \, ||\nabla f_k||^2 \to 0$$

método "ideal"

Ejemplo

$$p_k = -B_k^{-1} \nabla f_k,$$

en donde B_k es simétrica positiva definida (Newton, cuasi-Newton, ...)
Supongamos que la sucesión de matrices $\{B_k\}$ tiene número de condición uniformemente acotado. Es decir que existe una constante positiva M tal que :

$$\text{cond}\,(B_k) \equiv ||B_k||\,||B_k^{-1}|| \leq M, \quad \forall k.$$

Entonces, *se puede probar* que (Tarea)

$$\cos\theta_k \geq \frac{1}{M}.$$

Referencias

- Yu-Hong Dai, *Convergence properties of the BFGS algorithm.* SIAM Journal on Optimization, vol 13. No 3, pp. 693-701, 2002.
- Dong C. Liu and Jorge Nocedal, *On the limited memory BFGS method for large scale optimization.* Mathematical Programming 45, pp. 503-528, 1989.