

Convergencia del método BFGS

El objetivo de esta nota es presentar los antecedentes teóricos y prácticos de las pruebas de convergencia del método que utiliza la fórmula de actualización BFGS. Ambas pruebas hacen uso de la función

$$\Psi(B) = \text{traza}(B) + \ln(\det(B)),$$

en donde B es una matriz obtenida mediante la fórmula de actualización BFGS. El análisis de convergencia local hace uso del teorema de caracterización de Dennis y Moré.

Teorema. Supongamos que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es tres veces continuamente diferenciable, ahora consideremos una iteración general de la forma

$$x_{k+1} = x_k + \alpha_k p_k,$$

en donde p_k es una dirección de descenso que satisface las condiciones fuertes de Wolfe con $c_1 \leq \frac{1}{2}$; Si la sucesión $\{x_k\}$ converge a x^* , * en donde $\nabla f(x^*) = 0$, $\nabla^2 f(x^*)$ es positiva definida, y

$$\lim_{k \rightarrow \infty} \frac{\|\nabla f_k + \nabla^2 f_k p_k\|}{\|p_k\|} = 0 \quad (1)$$

entonces

1. existe un índice $k_0 > 0$ para el cual se acepta la longitud de paso $\alpha_k = 1$ con $k > k_0$
2. si $\alpha_k = 1 \quad \forall k > k_0$, $\{x_k\}$ converge a x^* superlinealmente.

Actividades

1. Es importante notar que el numerador en (1) implica que la dirección generada por el método cuasi Newton se aproxima a la dirección de Newton. Esta observación se puede hacer en el método de Broyden, el caso más elemental de un método cuasi-Newton. Verificar la observación anterior para los siguientes ejemplos:

$$\begin{aligned} F(1) &= x(1)^2 + x(2)^2 - 2; & \% \text{ x0} &= [1.5; 2]; \text{ B0} = J(\text{x0}) \\ F(2) &= \exp(x(1)-1) + x(2)^3 - 2; \end{aligned}$$

$$\begin{aligned} F(1) &= x(1) + x(2) - 3; & \% \text{ x0} &= [1; 5]; \text{ B0} = J(\text{x0}) \\ F(2) &= x(1)^2 + x(2)^2 - 9; \end{aligned}$$

2. Una de las hipótesis para la prueba de convergencia global es la siguiente

El conjunto $\mathcal{L} = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$ es convexo y existen constantes positivas m y M tales que

$$m\|z\|^2 \leq z^T G(x)z \leq M\|z\|^2 \quad \forall z \in \mathbb{R}^n, x \in \mathcal{L},$$

en donde $G(x)$ representa a la Hessiana de f evaluada en el punto x .

Interpretar la doble desigualdad $m\|z\|^2 \leq z^T G(x)z \leq M\|z\|^2$ en términos del cociente de Rayleigh.

3. Una referencia excelente sobre métodos cuasi-Newton es

JE Dennis, Jr, JJ Moré, *Quasi-Newton methods, motivation and theory*. SIAM Review, 19 (1), 46-89, 1977.

http://www.jstor.org/stable/2029325?seq=1#page_scan_tab_contents

4. Resolver los ejercicios 6.8-6.11 de J Nocedal and SJ Wright, *Numerical Optimization*, Second Edition. Springer.
5. Probar el

Teorema. Supongamos que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es dos veces continuamente diferenciable y una iteración de la forma $x_{k+1} = x_k + p_k$, en donde

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

con B_k simétrica positiva definida. Supongamos también que la sucesión $\{x_k\}$ converge a x^* en el cual $\nabla f(x^*) = 0$ y $\nabla^2 f(x^*)$ es spd. Entonces $\{x_k\}$ converge superlinealmente si y sólo si se cumple (1).