

## **CONVERGENCIA DE BÚSQUEDA EN LÍNEA**

ALAN REYES-FIGUEROA  
MÉTODOS NUMÉRICOS II

(AULA 24) 05.OCTUBRE.2023

# Búsqueda en Línea

Estudiamos ahora la convergencia global para el caso del algoritmo de búsqueda en línea, usando las condiciones de Wolfe o de Goldstein. La propiedad clave es estudiar el ángulo entre  $\mathbf{d}_k$  y  $-\nabla f(\mathbf{x}_k)$ :

$$\cos \varphi_k = -\frac{\nabla f(\mathbf{x}_k)^T \mathbf{d}_k}{\|\nabla f(\mathbf{x}_k)\| \|\mathbf{d}_k\|}.$$

El siguiente resultado cuantifica el efecto de elegir apropiadamente el tamaño de paso  $\alpha_k$ . También describe qué tan lejos  $\mathbf{d}_k$  puede desviarse de  $-\nabla f(\mathbf{x}_k)$ , y aún producir convergencia global.

## Teorema (Zoutendijk)

Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  diferenciable, con  $\nabla f$  Lipschitz sobre un abierto  $U$  que contiene al conjunto de subnivel  $S_{f_0} = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$ , con constante de Lipschitz  $\gamma$ .

Entonces

$$\sum_{k \geq 0} \cos^2 \varphi_k \|\nabla f(\mathbf{x}_k)\|^2 < \infty,$$

# Búsqueda en Línea

si  $\mathbf{x}_k$  se construye con descenso y búsqueda en línea usando Wolfe-Backtracking.

Prueba: De la segunda condición de Wolfe (6),

$$\nabla f(\mathbf{x}_k + \alpha_k \mathbf{d}_k)^T \mathbf{d}_k \geq c_2 \nabla f(\mathbf{x}_k)^T \mathbf{d}_k,$$

tenemos que

$$(\nabla f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) - \nabla f(\mathbf{x}_k))^T \mathbf{d}_k \geq c_2 \nabla f(\mathbf{x}_k)^T \mathbf{d}_k - \nabla f(\mathbf{x}_k)^T \mathbf{d}_k = (c_2 - 1) \nabla f(\mathbf{x}_k)^T \mathbf{d}_k.$$

Por otro lado, la condición de Lipschitz (+ Cauchy-Schwarz) implican

$$\begin{aligned} (\nabla f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) - \nabla f(\mathbf{x}_k))^T \mathbf{d}_k &\leq \|\nabla f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) - \nabla f(\mathbf{x}_k)\| \cdot \|\mathbf{d}_k\| \\ &\leq \gamma \|\alpha_k \mathbf{d}_k\| \cdot \|\mathbf{d}_k\| = \gamma \alpha_k \|\mathbf{d}_k\|^2. \end{aligned}$$

Combinando ambos resultados,

$$\alpha_k \geq \frac{(\nabla f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) - \nabla f(\mathbf{x}_k))^T \mathbf{d}_k}{\gamma \|\mathbf{d}_k\|^2} \geq \frac{(c_2 - 1) \nabla f(\mathbf{x}_k)^T \mathbf{d}_k}{\gamma \|\mathbf{d}_k\|^2}.$$

Sustituyendo esta última desigualdad en la primer condición de Wolfe, obtenemos

$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) \leq f(\mathbf{x}_k) + c_1 \alpha_k \nabla f(\mathbf{x}_k)^T \mathbf{d}_k \leq f(\mathbf{x}_k) - \frac{c_1(1-c_2)}{\gamma \|\mathbf{d}_k\|^2} (\nabla f(\mathbf{x}_k)^T \mathbf{d}_k)^2.$$

# Búsqueda en Línea

Haciendo  $A = \frac{c_1(1-c_2)}{\gamma}$ , podemos escribir lo anterior como

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - A \cos^2 \varphi_k \|\nabla f(\mathbf{x}_k)\|^2, \quad k = 0, 1, 2, \dots$$

Sumando sobre todos los índices  $\leq k$ , resulta

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_0) - A \sum_{j=0}^k \cos^2 \varphi_j \|\nabla f(\mathbf{x}_j)\|^2, \quad k = 0, 1, 2, \dots$$

Como  $f$  es limitada inferiormente, si  $b$  es una cota inferior para  $f$ , entonces  $f(\mathbf{x}_0) - f(\mathbf{x}_{k+1}) < f(\mathbf{x}_0) - b < \infty$ . Luego,

$$A \sum_{j=0}^k \cos^2 \varphi_j \|\nabla f(\mathbf{x}_j)\|^2 \leq f(\mathbf{x}_0) - f(\mathbf{x}_{k+1}) < \infty, \quad \forall k \in \mathbb{N}.$$

Tomando  $k \rightarrow \infty$ , la serie  $\sum_{j=0}^{\infty} \cos^2 \varphi_j \|\nabla f(\mathbf{x}_j)\|^2$  converge.  $\square$

**Obs!** Un resultado similar se obtiene para las condiciones fuertes de Wolfe, y para las condiciones de Goldstein. En todos los casos, la selección del tamaño de paso implica la **condición de Zoutendijk**:

$$\sum_{k \geq 0} \cos^2 \varphi_k \|\nabla f(\mathbf{x}_k)\|^2 < \infty.$$

La condición de Zoutendijk implica que  $\lim_{k \rightarrow \infty} \cos^2 \varphi_k \|\nabla f(\mathbf{x}_k)\|^2 = 0$ , y portanto

$$\lim_{k \rightarrow \infty} \cos \varphi_k \|\nabla f(\mathbf{x}_k)\| = 0.$$

Si la elección de  $\mathbf{d}_k$  asegura en cada paso que el ángulo  $\varphi_k$  está lejos de  $90^\circ$ , entonces existe  $\delta > 0$  tal que

$$\cos \varphi_k \geq \delta > 0, \quad \forall k \geq 0.$$

De ahí,  $\delta \lim_{k \rightarrow \infty} \|\nabla f(\mathbf{x}_k)\| \leq \lim_{k \rightarrow \infty} \cos \varphi_k \|\nabla f(\mathbf{x}_k)\| = 0 \implies \lim_{k \rightarrow \infty} \|\nabla f(\mathbf{x}_k)\| = 0.$

## Corolario (Convergencia Global Wolfe/Goldstein-Backtracking)

Si  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es diferenciable, con  $\nabla f$  Lipschitz en el conjunto de subnivel  $S_{f_0} = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$ , con  $f$  limitada inferiormente; y la dirección de descenso  $\mathbf{d}_k$  se elige de modo que  $\cos \varphi_k \geq \delta > 0$ , para todo  $k \geq 0$ , entonces el algoritmo de Backtracking con condiciones de Wolfe (o de Goldstein) converge a un punto estacionario  $\mathbf{x}^*$ , con  $\nabla f(\mathbf{x}^*) = \mathbf{0}$ .  $\square$

### Observaciones:

- Para métodos de búsqueda en línea, la convergencia global (i.e.  $\nabla f(\mathbf{x}_k) \rightarrow \mathbf{0}$ ) es el mejor resultado de convergencia que puede obtenerse.
- No se puede garantizar convergencia a un mínimo de  $f$ , sólo a un punto estacionario. Salvo que se introduzca alguna condición sobre la curvatura sobre  $\mathbf{d}_k$  (e.g. curvatura positiva o convexidad en la dirección de  $\mathbf{d}_k$ , a partir del Hessiano  $D^2f(\mathbf{x}_k)$ ), puede fortalecerse el resultado anterior para asegurar convergencia a un mínimo.

# Búsqueda en Línea

Para los métodos de tipo Newton o quasi-Newton,

$$\mathbf{x}_{k+1} = \mathbf{x}_k - B_k^{-1} \nabla f(\mathbf{x}_k),$$

suponga que  $B_k$  es positiva definida y con número de condición limitado

$$\kappa = \kappa(B_k) = \|B_k\| \cdot \|B_k^{-1}\| \leq M, \quad \forall k \in \mathbb{N}.$$

Entonces de las propiedades de norma matricial,  $\|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\|$ , y de Cauchy-Schwarz  $|\mathbf{y}^T \mathbf{x}| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$ , tenemos

$$\begin{aligned} \cos \varphi_k &= -\frac{\nabla f(\mathbf{x}_k)^T \mathbf{d}_k}{\|\nabla f(\mathbf{x}_k)\| \cdot \|\mathbf{d}_k\|} = \frac{|\nabla f(\mathbf{x}_k)^T \mathbf{d}_k|}{\|B_k \mathbf{d}_k\| \cdot \|B_k^{-1} \nabla f(\mathbf{x}_k)\|} \geq \frac{\|\nabla f(\mathbf{x}_k)\| \cdot \|\mathbf{d}_k\|}{\|B_k\| \cdot \|\nabla f(\mathbf{x}_k)\| \cdot \|B_k^{-1}\| \cdot \|\mathbf{d}_k\|} \\ &\geq \frac{1}{\|B_k\| \cdot \|B_k^{-1}\|} = \frac{1}{M}. \end{aligned}$$

Luego, de la condición de Zoutendijk, tenemos

## Corolario (Convergencia Global de los métodos quasi-Newton)

Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  diferenciable, con  $\nabla f$  Lipschitz en el conjunto de subnivel  $S_{f_0} = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$ , y con  $f$  limitada inferiormente. Si para todo  $k \in \mathbb{N}$ ,  $B_k$  es positiva definida y tiene número de condición limitado  $\kappa(B_k) \leq M$ , y  $\alpha_k$  satisface las condiciones de Wolfe o de Goldstein, entonces la iteración

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k B_k^{-1} \nabla f(\mathbf{x}_k),$$

satisface  $\lim_{k \rightarrow \infty} \|\nabla f(\mathbf{x}_k)\| = 0$ , y el método converge a un punto estacionario.  $\square$

Para otros métodos como el gradiente conjugado, veremos que es posible probar la condición débil  $\liminf_{k \rightarrow \infty} \|\nabla f(\mathbf{x}_k)\| = 0$ , y que sólo una subsecuencia de gradientes converge a 0.

Usando la condición de Zoutendijk, aún es posible probar algo útil:

Prueba: Suponga que  $\liminf_{k \rightarrow \infty} \|\nabla f(\mathbf{x}_k)\| = \gamma > 0$ . Entonces existe  $k_0 \in \mathbb{N}$  tal que



$$\nabla f(\mathbf{x}_k) \geq \gamma, \quad \text{para } k \geq 0.$$

Como

$$\sum_{k=0}^{\infty} \cos^2 \varphi_k \|\nabla f(\mathbf{x}_k)\|^2 < \infty \implies \lim_{k \rightarrow \infty} \cos^2 \varphi_k \|\nabla f(\mathbf{x}_k)\|^2 = 0,$$

luego se tiene que  $\cos^2 \varphi_k \rightarrow 0$ , y portanto  $\cos \varphi_k \rightarrow 0$ .

Para mostrar tal afirmación, basta que una subsecuencia  $\{\cos \varphi_{k_j}\}$  esté limitada lejos de 0.  $\square$

## Definición

Sea  $\{\mathbf{x}_k\} \subset \mathbb{R}^n$  una secuencia de puntos tales que  $\mathbf{x}_k \rightarrow \mathbf{x}^* \in \mathbb{R}^n$ , para la cual existen constantes  $M > 0$ ,  $\alpha > 0$ , tales que

$$\frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|^\alpha} \leq M, \quad \text{para todo } k \in \mathbb{N}, \quad (1)$$

o equivalentemente

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\| \leq M \|\mathbf{x}_k - \mathbf{x}^*\|^\alpha, \quad \text{para todo } k \in \mathbb{N}. \quad (2)$$

Decimos que  $\{\mathbf{x}_k\}$  converge a  $\mathbf{x}^*$  **con orden**  $\alpha$ , o que  $\{\mathbf{x}_k\}$  tiene **orden de convergencia**  $\alpha$ .

- Si  $\alpha = 1$ , decimos que  $\{\mathbf{x}_k\}$  tiene convergencia lineal.
- Si  $\alpha = 2$ , decimos que  $\{\mathbf{x}_k\}$  tiene convergencia cuadrática.
- Si  $0 < \alpha < 1$ , decimos que  $\{\mathbf{x}_k\}$  tiene convergencia sub-lineal.
- Si  $1 < \alpha < 2$ , decimos que  $\{\mathbf{x}_k\}$  tiene convergencia super-lineal.

# Búsqueda en Línea

Hemos visto que tomar  $\mathbf{d}_k$  de modo que  $\cos \varphi_k$  sea mayor que cierta constante  $\delta > 0$ . Sin embargo estos criterios no son deseables por varias razones:

- impiden una tasa de convergencia rápida,
- mala elección de  $\delta$  en caso de Hessianos mal condicionados,
- estos test angulares destruyen las propiedades de invarianza en los métodos quasi-Newton.

## Tasa de Convergencia para Descenso Gradiente:

Consideremos el caso ideal donde la función objetivo es cuadrática

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} - \mathbf{b}^T \mathbf{x},$$

con  $\mathbf{Q} \in \mathbb{R}^{n \times n}$  simétrica y positiva definida. El gradiente es  $\nabla f(\mathbf{x}) = \mathbf{Q} \mathbf{x} - \mathbf{b}$ , y en el mínimo global  $\mathbf{x}^*$ , se satisface  $\mathbf{Q} \mathbf{x}^* = \mathbf{b}$ .

Calculamos el tamaño de paso óptimo  $\alpha_k$  que minimiza la función

$$\varphi(\alpha) = f(\mathbf{x}_k - \alpha \nabla f(\mathbf{x}_k)).$$

# Búsqueda en Línea

$$\begin{aligned}\varphi'(\alpha) &= Df(\mathbf{x}_k - \alpha \nabla f(\mathbf{x}_k)) (-\nabla f(\mathbf{x}_k)) = [Q(\mathbf{x}_k - \alpha \nabla f(\mathbf{x}_k)) - \mathbf{b}](-\nabla f(\mathbf{x}_k)) \\ &= -\nabla f(\mathbf{x}_k)^T (Q\mathbf{x}_k - \alpha Q \nabla f(\mathbf{x}_k) - \mathbf{b}) = -\nabla f(\mathbf{x}_k)^T (\nabla f(\mathbf{x}_k) - \alpha Q \nabla f(\mathbf{x}_k)) = 0.\end{aligned}$$

Luego

$$\alpha_k = \frac{\nabla f(\mathbf{x}_k)^T \nabla f(\mathbf{x}_k)}{\nabla f(\mathbf{x}_k)^T Q \nabla f(\mathbf{x}_k)}.$$

Usando este tamaño de paso óptimo, la iteración de descenso máximo resulta

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left( \frac{\nabla f(\mathbf{x}_k)^T \nabla f(\mathbf{x}_k)}{\nabla f(\mathbf{x}_k)^T Q \nabla f(\mathbf{x}_k)} \right) \nabla f(\mathbf{x}_k). \quad (3)$$

Como  $\nabla f(\mathbf{x}_k) = Q\mathbf{x}_k - \mathbf{b}$ , la ecuación (3) produce una fórmula cerrada para  $\mathbf{x}_{k+1}$ .

Para calcular la tasa de convergencia, usamos la norma  $\|\mathbf{x}\|_Q^2 = \mathbf{x}^T Q \mathbf{x}$ . De  $Q\mathbf{x}^* = \mathbf{b}$ ,

$$\begin{aligned}\frac{1}{2} \|\mathbf{x} - \mathbf{x}^*\|_Q^2 &= \frac{1}{2} (\mathbf{x} - \mathbf{x}^*)^T Q (\mathbf{x} - \mathbf{x}^*) = \frac{1}{2} \mathbf{x}^T Q \mathbf{x} - \mathbf{x}^T Q \mathbf{x}^* + \frac{1}{2} (\mathbf{x}^*)^T Q \mathbf{x}^* \\ &= \frac{1}{2} \mathbf{x}^T Q \mathbf{x} - \mathbf{x}^T Q \mathbf{x}^* - \frac{1}{2} (\mathbf{x}^*)^T Q \mathbf{x}^* + (\mathbf{x}^*)^T Q \mathbf{x}^* = \frac{1}{2} \mathbf{x}^T Q \mathbf{x} - \mathbf{b}^T \mathbf{x} - \frac{1}{2} (\mathbf{x}^*)^T Q \mathbf{x}^* + \mathbf{b}^T \mathbf{x}^* \\ &= f(\mathbf{x}) - f(\mathbf{x}^*).\end{aligned}$$

# Búsqueda en Línea

Esta norma mide la diferencia entre el valor actual de  $f$  y el valor óptimo. Ahora, como  $\nabla f(\mathbf{x}_k) = Q\mathbf{x}_k - \mathbf{b} = Q(\mathbf{x}_k - \mathbf{x}^*)$ , entonces

$$\begin{aligned} \|\mathbf{x}_{k+1} - \mathbf{x}^*\|_Q^2 - \|\mathbf{x}_k - \mathbf{x}^*\|_Q^2 &= 2(f(\mathbf{x}_{k+1}) - f(\mathbf{x}^*)) - 2(f(\mathbf{x}_k) - f(\mathbf{x}^*)) = 2(f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k)) \\ &= (\mathbf{x}_{k+1}^T Q \mathbf{x}_{k+1} - 2\mathbf{b}^T \mathbf{x}_{k+1}) - (\mathbf{x}_k^T Q \mathbf{x}_k - 2\mathbf{b}^T \mathbf{x}_k) \\ &= (\mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k))^T Q (\mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k)) - 2\mathbf{b}^T (\mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k)) \\ &\quad - \mathbf{x}_k^T Q \mathbf{x}_k + 2\mathbf{b}^T \mathbf{x}_k \\ &= \mathbf{x}_k^T Q \mathbf{x}_k - 2\alpha_k \nabla f(\mathbf{x}_k)^T Q \mathbf{x}_k + \alpha_k^2 \nabla f(\mathbf{x}_k)^T Q \nabla f(\mathbf{x}_k) \\ &\quad + 2\alpha_k \mathbf{b}^T \nabla f(\mathbf{x}_k) - \mathbf{x}_k^T Q \mathbf{x}_k \\ &= \alpha_k^2 \nabla f(\mathbf{x}_k)^T Q \nabla f(\mathbf{x}_k) - 2\alpha_k \nabla f(\mathbf{x}_k)^T \underbrace{(Q\mathbf{x}_k - \mathbf{b})}_{=\nabla f(\mathbf{x}_k)} \\ &= -\frac{(\nabla f(\mathbf{x}_k)^T \nabla f(\mathbf{x}_k))^2}{\nabla f(\mathbf{x}_k)^T Q \nabla f(\mathbf{x}_k)}. \end{aligned}$$

# Búsqueda en Línea

y como

$$\|\mathbf{x}_k - \mathbf{x}^*\|_Q^2 = (\mathbf{x}_k - \mathbf{x}^*)^T Q (\mathbf{x}_k - \mathbf{x}^*) = \nabla f(\mathbf{x}_k)^T Q^{-1} Q Q^{-1} \nabla f(\mathbf{x}_k) = \nabla f(\mathbf{x}_k)^T Q^{-1} \nabla f(\mathbf{x}_k),$$

entonces

$$\begin{aligned} \|\mathbf{x}_{k+1} - \mathbf{x}^*\|_Q^2 &= \|\mathbf{x}_k - \mathbf{x}^*\|_Q^2 + (\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_Q^2 - \|\mathbf{x}_k - \mathbf{x}^*\|_Q^2) \\ &= \nabla f(\mathbf{x}_k)^T Q^{-1} \nabla f(\mathbf{x}_k) - \frac{(\nabla f(\mathbf{x}_k)^T \nabla f(\mathbf{x}_k))^2}{\nabla f(\mathbf{x}_k)^T Q \nabla f(\mathbf{x}_k)} \\ &= \left(1 - \frac{(\nabla f(\mathbf{x}_k)^T \nabla f(\mathbf{x}_k))^2}{(\nabla f(\mathbf{x}_k)^T Q \nabla f(\mathbf{x}_k))(\nabla f(\mathbf{x}_k)^T Q^{-1} \nabla f(\mathbf{x}_k))}\right) (\nabla f(\mathbf{x}_k)^T Q^{-1} \nabla f(\mathbf{x}_k)) \\ &= \left(1 - \frac{(\nabla f(\mathbf{x}_k)^T \nabla f(\mathbf{x}_k))^2}{(\nabla f(\mathbf{x}_k)^T Q \nabla f(\mathbf{x}_k))(\nabla f(\mathbf{x}_k)^T Q^{-1} \nabla f(\mathbf{x}_k))}\right) \|\mathbf{x}_k - \mathbf{x}^*\|_Q^2. \end{aligned} \quad (4)$$

Esta ecuación describe la tasa exacta de descenso de  $f$  en cada iteración. El término entre paréntesis es difícil de interpretar, y es más útil acotar éste por una condición más simple.

## Teorema (Desigualdad de Kantorovich)

Sea  $Q \in \mathbb{R}^{n \times n}$  simétrica y positiva definida, con autovalores  $\lambda_1 \geq \dots \geq \lambda_n > 0$ . Para todo  $\mathbf{x} \in \mathbb{R}^n$ ,  $\mathbf{x} \neq \mathbf{0}$ , vale

$$\frac{(\mathbf{x}^T \mathbf{x})^2}{(\mathbf{x}^T Q \mathbf{x})(\mathbf{x}^T Q^{-1} \mathbf{x})} \geq \frac{4\lambda_1 \lambda_n}{(\lambda_1 + \lambda_n)^2}.$$

Prueba: Por el Teorema Espectral,  $Q$  admite una descomposición  $Q = U \Lambda U^T$ , con  $\Lambda = (\lambda_1, \dots, \lambda_n)$  y  $U \in O(n)$  ortogonal.

Haciendo el cambio de coordenadas  $\mathbf{y} = U^T \mathbf{x}$ , tenemos

$$\mathbf{y}^T \mathbf{y} = \mathbf{x}^T U U^T \mathbf{x} = \mathbf{x}^T \mathbf{x}, \quad \mathbf{x}^T Q \mathbf{x} = \mathbf{x}^T U \Lambda U^T \mathbf{x} = \mathbf{y}^T \Lambda \mathbf{y}, \quad \mathbf{x}^T Q^{-1} \mathbf{x} = \mathbf{x}^T U \Lambda^{-1} U^T \mathbf{x} = \mathbf{y}^T \Lambda^{-1} \mathbf{y}. \quad (5)$$

En este nuevo sistema coordenado, la expresión en el lado izquierdo de (5) es

$$\frac{(\mathbf{x}^T \mathbf{x})^2}{(\mathbf{x}^T Q \mathbf{x})(\mathbf{x}^T Q^{-1} \mathbf{x})} = \frac{(\mathbf{y}^T \mathbf{y})^2}{(\mathbf{y}^T \Lambda \mathbf{y})(\mathbf{y}^T \Lambda^{-1} \mathbf{y})} = \frac{(\sum_i y_i^2)^2}{(\sum_i \lambda_i y_i^2)(\sum_i y_i^2 / \lambda_i)}.$$

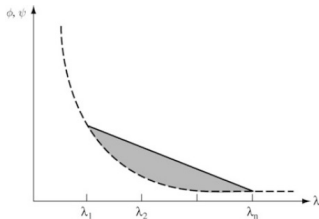
# Búsqueda en Línea

Haciendo  $z_i = y_i^2 / \sum_j y_j^2$  (normalización en los  $y_j^2$ ), tenemos

$$\frac{(\mathbf{x}^T \mathbf{x})^2}{(\mathbf{x}^T \mathbf{Q} \mathbf{x})(\mathbf{x}^T \mathbf{Q}^{-1} \mathbf{x})} = \frac{(\sum_i y_i^2)^2}{(\sum_i \lambda_i y_i^2)(\sum_i y_i^2 / \lambda_i)} = \frac{1 / \sum_i \lambda_i z_i}{\sum_i z_i / \lambda_i} = \frac{u(\mathbf{z})}{v(\mathbf{z})}.$$

Así, la expresión de interés es cociente de dos funciones convexas: una combinación de los  $\lambda_i$ ; la otra, combinación de los  $1/\lambda_i$ . Consideramos la función  $\frac{1}{\lambda}$ . Como  $\lambda_n \leq \sum_i z_i \lambda_i \leq \lambda_1$ , entonces  $u(\mathbf{z}) = 1 / \sum_i z_i \lambda_i$  es un punto arriba de esa curva.

Por otro lado,  $v(\mathbf{z}) = \sum_i z_i / \lambda_i$  es una combinación conveza de puntos sobre esa curva  $\Rightarrow$  el valor está restringido al área sombreada.





# Búsqueda en Línea

Para  $\mathbf{x} \in \mathbb{R}^n$  fijo, el mínimo de  $\frac{u(\mathbf{z})}{v(\mathbf{z})}$  corresponde a un valor  $\lambda = t\lambda_1 + (1-t)\lambda_n$ , con  $0 \leq t \leq 1$ . Usando la relación  $\frac{t}{\lambda_1} + \frac{1-t}{\lambda_n} = \frac{t\lambda_n + (1-t)\lambda_1}{\lambda_1\lambda_n}$ , podemos escribir

$$\frac{u(\mathbf{z})}{v(\mathbf{z})} \geq \liminf_{\lambda_1 \leq \lambda \leq \lambda_n} \frac{1/(t\lambda_1 + (1-t)\lambda_n)}{(t\lambda_n + (1-t)\lambda_1)/\lambda_1\lambda_n}.$$

El mínimo de esta relación se alcanza cuando  $t = \frac{1}{2}$  (basta derivar en  $t$ ). Así,

$$\lambda = \frac{\lambda_1 + \lambda_n}{2} \quad \Rightarrow \quad \frac{u(\mathbf{z})}{v(\mathbf{z})} \geq \frac{2/(\lambda_1 + \lambda_n)}{(\lambda_1 + \lambda_n)/2\lambda_1\lambda_n} = \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2}. \quad \square$$

## Teorema (Error en Descenso Gradiente, caso cuadrático)

Sea  $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q \mathbf{x} - \mathbf{b}^T \mathbf{x}$ , con  $Q \succ 0$  y autovalores  $\lambda_1 \geq \dots \geq \lambda_n > 0$ . Para todo  $\mathbf{x}_0 \in \mathbb{R}^n$ , el método de descenso gradiente converge a un mínimo  $\mathbf{x}^*$  de  $f$ . Más aún, si se toma  $\alpha_k$  el tamaño de paso óptimo, entonces

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_Q^2 \leq \left( \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} \right)^2 \|\mathbf{x}_k - \mathbf{x}^*\|_Q^2. \quad (6)$$

# Búsqueda en Línea

Prueba: De la observación anterior (4), y la desigualdad de Kantorovich (5),

$$\begin{aligned}\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_Q^2 &\leq \left(1 - \frac{(\nabla f(\mathbf{x}_k)^T \nabla f(\mathbf{x}_k))^2}{(\nabla f(\mathbf{x}_k)^T Q \nabla f(\mathbf{x}_k))(\nabla f(\mathbf{x}_k)^T Q^{-q} \nabla f(\mathbf{x}_k))}\right) \|\mathbf{x}_k - \mathbf{x}^*\|_Q^2 \\ &\leq \left(1 - \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2}\right) \|\mathbf{x}_k - \mathbf{x}^*\|_Q^2 = \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2 \|\mathbf{x}_k - \mathbf{x}^*\|_Q^2. \quad \square\end{aligned}$$

## Observaciones:

- La convergencia se sigue del hecho que  $\left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2 < 1$ . (Recordar el Teorema de Punto Fijo de Banach, para contracciones). En consecuencia  $\|\mathbf{x}_k - \mathbf{x}^*\|_Q \rightarrow 0$ .
- La tasa de convergencia depende sólo de la cantidad  $r = \frac{\lambda_n}{\lambda_1}$ , pues

$$\left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2 = \left(\frac{1-r}{1+r}\right)^2.$$

- La ecuación (6) indica que el método converge linealmente a convergencia a una tasa menor que  $\left(\frac{1-r}{1+r}\right)^2$ . Cuando  $\lambda_1 = \lambda_n$ , el método converge en 1 paso.

# Búsqueda en Línea

Para funciones no cuadráticas, establecemos estimativas para cuando  $D^2f$  es limitada inferiormente y superiormente, es positiva definida y  $aI \leq D^2f(\mathbf{x}) \leq AI$ .

## Búsqueda en línea exacta:

Para cualquier  $\mathbf{x}_k \in \mathbb{R}^n$ ,  $\alpha > 0$ , tenemos

$$f(\mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k)) \leq f(\mathbf{x}_k) - \alpha_k \nabla f(\mathbf{x}_k)^T \nabla f(\mathbf{x}_k) + \frac{A\alpha^2}{2} \nabla f(\mathbf{x}_k)^T \nabla f(\mathbf{x}_k). \quad (7)$$

Minimizando ambos lados respecto de  $\alpha$ , la desigualdad se mantiene. Del lado derecho

$$\nabla_{\alpha} = (A\alpha - 1) \nabla f(\mathbf{x}_k)^T \nabla f(\mathbf{x}_k) = \mathbf{0} \Rightarrow \alpha = \frac{1}{A}. \text{ Así,}$$
$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \frac{1}{2A} \|\nabla f(\mathbf{x}_k)\|^2.$$

Restando el valor óptimo  $f^* = f(\mathbf{x}^*)$  en ambos lados,

$$f(\mathbf{x}_{k+1}) - f^* \leq f(\mathbf{x}_k) - f^* - \frac{1}{2A} \|\nabla f(\mathbf{x}_k)\|^2. \quad (8)$$

Haciendo lo mismo para  $f(\mathbf{x}) \geq f(\mathbf{x}_k) - \nabla f(\mathbf{x}_k)^T (\mathbf{x} - \mathbf{x}_k) + \frac{a}{2} (\mathbf{x} - \mathbf{x}_k)^T (\mathbf{x} - \mathbf{x}_k)$ , y minimizando ambos lados, resulta  $\nabla_{\alpha} = \nabla f(\mathbf{x}_k) - a(\mathbf{x} - \mathbf{x}_k) = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{x}_k - \frac{1}{a} \nabla f(\mathbf{x}_k) =$ .

De ahí que

$$f^* \geq f(\mathbf{x}_k) - \frac{1}{2a} \|\nabla f(\mathbf{x}_k)\|^2. \quad (9)$$

# Búsqueda en Línea

Despejando  $-\|\nabla f(\mathbf{x}_k)\|^2$  en (9) y sustituyendo en (8), resulta

$$f(\mathbf{x}_{k+1}) - f^* \leq f(\mathbf{x}_k) - f^* - \frac{1}{2A} (2a) (f^* - f(\mathbf{x}_k)) \leq \left(1 - \frac{a}{A}\right) (f(\mathbf{x}_k) - f^*).$$

Portanto, obtenemos

$$f(\mathbf{x}_{k+1}) - f^* \leq \left(1 - \frac{a}{A}\right) (f(\mathbf{x}_k) - f^*). \quad (10)$$

## Otros casos:

En el caso de descenso gradiente con la condición de Armijo:

$$f(\mathbf{x}_{k+1}) - f^* \leq \left(1 - \frac{2c_1 a \rho}{A}\right) (f(\mathbf{x}_k) - f^*). \quad (11)$$

En el caso de descenso gradiente con la dirección de Newton:

## Teorema (Convergencia del método de Newton)

Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  clase  $C^2$ , con  $D^2f$  Lipschitz de constante  $L$  en una vecindad  $U$  del mínimo  $\mathbf{x}^*$ , con  $\nabla f(\mathbf{x}^*) = \mathbf{0}$  y  $D^2f(\mathbf{x}^*) \succ \mathbf{0}$ . En la iteración de descenso gradiente con la dirección de Newton, tenemos:

# Búsqueda en Línea

1. si  $\mathbf{x}_0$  está suficientemente cerca de  $\mathbf{x}^*$  ( $\mathbf{x}_0 \in U$ ), entonces  $\mathbf{x}_k \rightarrow \mathbf{x}^*$ ,
2. la tasa de convergencia de  $\{\mathbf{x}_k\}$  es cuadrática, con

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\| \leq \tilde{L} \|\mathbf{x}_k - \mathbf{x}^*\|^2, \quad \text{con } \tilde{L} = L \|D^2f(\mathbf{x}^*)^{-1}\|. \quad (12)$$

3. la secuencia  $\{\|\nabla f(\mathbf{x}_k)\|\}$  converge cuadráticamente a 0.

En el caso de descenso los métodos quasi-Newton:

## Teorema (Convergencia de los métodos quasi-Newton)

Sea  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  clase  $C^2$ , y considere la iteración  $\mathbf{x}_{k+1} = \mathbf{x}_k - B_k^{-1} \nabla f(\mathbf{x}_k)$ , con tamaño de paso  $\alpha_k = 1$ . Suponga que  $\mathbf{x}_k \rightarrow \mathbf{x}^*$ , con  $\nabla f(\mathbf{x}^*) = \mathbf{0}$  y  $D^2f(\mathbf{x}^*) \succ \mathbf{0}$ . Entonces el método converge super-linealmente si, y sólo si,

$$\lim_{k \rightarrow \infty} \frac{\|B_k - D^2f(\mathbf{x}_k)B_k^{-1}\nabla f(\mathbf{x}_k)\|}{\|B_k^{-1}\nabla f(\mathbf{x}_k)\|} = 0. \quad (13)$$