

PROPUESTO CLASE AUXILIAR 9: OPTIMIZACIÓN

PROFESOR : ALEJANDRO JOFRÉ

AUXILIARES : NICOLÁS HERNÁNDEZ & EMILIO VILCHES

22 de octubre de 2009

1. Preliminares

Para resolver el problema de minimización sin restricciones

$$\min\{f(x), x \in \mathbb{R}^n\} \quad (1)$$

consideramos el algoritmo

$$x^{k+1} = x^k + \alpha^k d^k, \quad k = 0, 1, \dots \quad (2)$$

donde α^k es llamado el paso y d^k es una dirección de descenso, es decir, satisface

$$f(x^k + \alpha^k d^k) < f(x^k), \quad k = 0, 1, \dots \quad (3)$$

Para escoger la dirección de descenso podemos considerar:

1. “Método del máximo descenso”

$$d^k = -\nabla f(x^k), \quad k = 0, 1, \dots$$

2. “Método de Newton”

$$d^k = -\left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k), \text{ siempre que } \nabla^2 f(x^k) \text{ sea definido positivo} \quad k = 0, 1, \dots$$

Para escoger el paso podemos considerar las siguientes reglas

1. **“Regla de minimización:”** En cada iteración α^k es tal que la función f es minimizada a lo largo de la dirección d^k , esto es, α^k satisface

$$f(x^k + \alpha^k d^k) = \min_{\alpha \geq 0} f(x^k + \alpha d^k). \quad (4)$$

2. **“Regla de minimización limitada:”** En cada iteración α^k es tal que

$$f(x^k + \alpha^k d^k) = \min_{\alpha \in [0, s]} f(x^k + \alpha d^k). \quad (5)$$

donde $s > 0$ es un escalar fijo.

3. **“Regla de Armijo o de reducciones sucesivas:”** Sean s, β y σ , con $s > 0$, $0 < \beta < 1$, y $0 < \sigma < 1$ escalares fijos. En cada iteración $\alpha^k = \beta^{m_k} s$ donde m_k es el primer entero no negativo m que satisface

$$f(x^k) - f(x^k + \beta^m s d^k) \geq -\sigma \beta^m s \nabla f(x^k)^t d^k. \quad (6)$$

4. **“Regla de Goldstein”** Sea $\sigma \in (0, 1/2)$ fijo. En cada iteración α^k se escoge para satisfacer

$$\sigma \leq \frac{f(x^k + \alpha^k d^k) - f(x^k)}{\alpha^k \nabla f(x^k)^t d^k} \leq 1 - \sigma.$$

2. Problema

Problema 1.

Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}$ una función dos veces continuamente diferenciable que satisface la condición

$$m\|y\|^2 \leq y^t \nabla^2 f(x) y \leq M\|y\|^2, \quad \forall x, y \in \mathbb{R}^n$$

donde m y M son constantes positivas. Sea x^* el mínimo de f .

Demuestre que

$$f(x) - f(x^*) \leq \frac{1}{2m} \|\nabla f(x)\|^2 \quad \forall x \in \mathbb{R}^n$$

$$\frac{m}{2} \|x - x^*\| \leq f(x) - f(x^*), \quad \forall x \in \mathbb{R}^n.$$

Problema 2.

Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}$ una función dos veces continuamente diferenciable que satisface la condición

$$m\|y\|^2 \leq y^t \nabla^2 f(x) y \leq M\|y\|^2, \quad \forall x, y \in \mathbb{R}^n$$

donde m y M son constantes positivas. Considere el método del máximo descenso $x^{k+1} = x^k - \alpha^k \nabla f(x^k)$ con α^k elegido con la regla de Armijo. Sea x^* el mínimo de f y definamos

$$r = 1 - \frac{4m\beta\sigma(1-\sigma)}{M}$$

Mostrar que para todo k , se tiene

$$f(x^{k+1}) - f(x^k) \leq r(f(x^k) - f(x^*))$$

y

$$\|x^k - x^*\| \leq qr^k,$$

donde q es alguna constante.

Solución:

Como α^k se determina por la regla de Armijo, sabemos que $\alpha^k = \beta^{m_k} s$, donde m_k es el primer índice m para el que

$$f(x^k) - f(x^k + \beta^m s d^k) \geq -\sigma \beta^m s \nabla f(x^k)^t d^k \quad (7)$$

y como estamos considerando el máximo descenso $d_k = -\nabla f(x^k)$, reemplazando esto en (7), obtenemos

$$f(x^k - \beta^m s \nabla f(x^k)) - f(x^k) \leq -\sigma \beta^m s \|\nabla f(x^k)\|^2. \quad (8)$$

La expansión de segundo orden de f entrega

$$f(x^k - \beta^i s \nabla f(x^k)) - f(x^k) = -\beta^i s \|\nabla f(x^k)\|^2 + \frac{(\beta^i s)^2}{2} \nabla f(x^k)^t \nabla^2 f(\bar{x}) \nabla f(x^k),$$

para algún $\bar{x}$ que está sobre el segmento uniendo los puntos $x^k - \beta^i s \nabla f(x^k)$ y x^k . Usando la propiedad de la función f , sigue que

$$f(x^k - \beta^i s \nabla f(x^k)) - f(x^k) \leq -\beta^i s \left(1 - \frac{\beta^i s M}{2}\right) \|\nabla f(x^k)\|^2. \quad (9)$$

Ahora, sea i_k el primer índice i para el cual $1 - \frac{M}{2}\beta^i s \geq \sigma$, es decir,

$$1 - \frac{M}{2}\beta^i s < \sigma \quad \forall i, 0 \leq i \leq i_k, \quad \text{y} \quad 1 - \frac{M}{2}\beta^{i_k} s \geq \sigma. \quad (10)$$

Usando (7)-(9), concluimos que $m_k \leq i_k$. Por lo tanto $\alpha^k \geq \hat{\alpha}^k$, donde $\hat{\alpha}^k = \beta^{i_k} s$. Así, tenemos

$$f(x^k - \alpha^k \nabla f(x^k)) - f(x^k) \leq -\sigma \hat{\alpha}^k \|\nabla f(x^k)\|^2 \quad (11)$$

Notemos que (10) implica

$$\sigma > 1 - \frac{M}{2}\beta^{i_k-1} s = 1 - \frac{M}{2\beta} \hat{\alpha}^k$$

Por lo tanto, $\hat{\alpha}^k \geq 2\beta(1 - \sigma)/M$. Reemplazando en (11), obtenemos

$$f(x^{k+1}) - f(x^*) \leq f(x^k) - f(x^*) - \frac{2\beta\sigma(1 - \sigma)}{M} \|\nabla f(x^k)\|^2. \quad (12)$$

La propiedad de f implica(ver problema anterior)

$$f(x) - f(x^*) \leq \frac{1}{2m} \|\nabla f(x)\|^2 \quad \forall x \in \mathbb{R}^n \quad (13)$$

$$\frac{m}{2} \|x - x^*\| \leq f(x) - f(x^*), \quad \forall x \in \mathbb{R}^n. \quad (14)$$

Combinando (12) y (13), obtenemos

$$f(x^{k+1}) - f(x^*) \leq r(f(x^k) - f(x^*)), \quad \forall k,$$

con $r = 1 - \frac{4m\beta\sigma(1-\sigma)}{M}$. Por lo tanto, tenemos

$$f(x^{k+1}) - f(x^*) \leq r^k(f(x^0) - f(x^*)), \quad \forall k,$$

que combinada con (14) entrega

$$\|x^k - x^*\|^2 \leq qr^k$$

con $q = \frac{2}{m}(f(x^0) - f(x^*))$

Problema 3.

Compare el resultado anterior con lo obtenido en la auxiliar 5.