Manuel García Garduño CLAVE: 162136

① Demuestre que si los vectores no nulos $p_1, p_2, \dots, p_\ell$ satisfacen que: $p_i^T A p_j = 0 \ \forall i \neq j$

y A es simétrica ~~[illegible]~~ y positiva definida entonces los vectores son linealmente independientes.

<u>Demo</u>: Para mostrar que el conjunto de vectores

$\{p_1, \dots, p_j, \dots, p_\ell\}$ es linealmente independiente

basta mostrar que:

$$\lambda_1 p_1 + \dots + \lambda_j p_j + \dots + \lambda_\ell p_\ell = 0 \iff \lambda_i = 0 \ \forall i \in \{1, \dots, \ell\}$$

Multiplicando por $p_j^T A$ por ambos lados tenemos que:

$$p_j^T A(\lambda_1 p_1 + \dots + \lambda_j p_j + \dots + \lambda_\ell p_\ell) = 0$$

$$\iff \lambda_1 \cancel{p_j^T A p_1}^{\to 0} + \dots + \lambda_j p_j^T A p_j + \dots + \lambda_\ell \cancel{p_j^T A p_\ell}^{\to 0} = 0$$

$$\iff \lambda_j p_j^T A p_j = 0$$

Como A es simétrica positiva definida entonces: $p_j^T A p_j > 0$

Con lo cual:

$$\lambda_j p_j^T A p_j = 0 \iff \lambda_j = 0$$

Y este razonamiento es válido para toda $j \in \{1, \dots, \ell\}$

$$\therefore \sum_{i=1}^{\ell} \lambda_i p_i = 0 \iff \lambda_i = 0 \ \forall i \in \{1, \dots, \ell\}$$

$\therefore \{p_1, \dots, p_\ell\}$ es un conjunto linealmente independiente ▩

Escaneado con CamScanner

② Dado este resultado, ¿Por qué el gradiente conjugado converge en a lo más n iteraciones?

<u>Demo</u>: Si el dominio de la función a optimizar es un subconjunto de $\mathbb{R}^n$ entonces sabemos que n vectores linealmente independientes son suficientes para generar a todo $\mathbb{R}^n$ y, en particular, a todo el dominio de la función.

Dado que el conjunto inicial de vectores $\{p_0, \dots, p_{n-1}\}$ es l.i entonces podemos afirmar que:

$$(x^* - x_0) = \lambda_1 p_1 + \dots + \lambda_n p_n$$ para una correcta elección de λ_i's.

¿Cuál es la elección de λ's correcta?

Podemos ver cuál es la elección de λ_j correcta si multiplicamos por $p_j^T A$ en ambos lados:

$$p_j^T A (x^* - x_0) = \lambda_j p_j^T A p_j$$

$$\iff \lambda_j = \frac{p_j^T A (x^* - x_0)}{p_j^T A p_j} = \frac{p_j^T A ((x^* - x_j) + \overset{0}{(x_j - x^*)})}{p_j^T A p_j}$$

$$= \frac{p_j^T A (x^* - x_j)}{p_j^T A p_j} = \frac{p_j^T (b - A x_j)}{p_j^T A p_j} = \frac{-r_j^T p_j}{p_j^T A p_j} \; !!$$

Esta elección de λ's coincide con la propuesta del método de Gradiente Conjugado.

<u>Conclusión</u>: Como las direcciones son l.i entonces podemos escribir al vector $(x^* - x_0)$ como una combinación lineal de esas n direcciones y eligiendo las λ's propuestas logramos escribir al vector $(x^* - x_0)$ en términos de <u>a lo más</u> n vectores, que requieren <u>a lo más</u> n pasos. ▨

Escaneado con CamScanner

1) Muestre que la condición fuerte de Wolfe implica la condición de curvatura.

Recordemos que $|a| \leq b \Rightarrow -b \leq a \leq b$

Entonces podemos afirmar lo siguiente:

$$|\nabla f(x_k + \alpha_k p_k)^T p_k| \leq c_2 |\nabla f(x_k)^T p_k| \rightarrow \text{Wolfe fuerte}$$

$$\Leftrightarrow \quad -c_2 |\nabla f(x_k)^T \cdot p_k| \leq \nabla f(x_k + \alpha_k p_k)^T p_k \leq c_2 |\nabla f(x_k)^T p_k|$$

En particular:

$$\nabla f^T(x_k + \alpha_k p_k) p_k \geq -c_2 |\nabla f(x_k)^T \cdot p_k|$$

Y como p_k es dirección de descenso:

$$\nabla f^T(x_k + \alpha_k p_k) p_k \geq c_2 \nabla f_k^T p_k$$

Restando $\nabla f_k^T p_k$ de ambos lados con $c_2 < 1$, tenemos que:

$$\nabla f(x_k + \alpha_k p_k)^T p_k - \nabla f_k^T p_k = (c_2 - 1) \nabla f_k^T p_k > 0$$

Multiplicando en ambos lados por α_k:

$$\nabla f^T(x_k + \alpha_k p_k) \overbrace{\alpha_k p_k}^{s_k} - \nabla f_k^T \alpha_k p_k = (c_2 - 1) \nabla f_k^T \alpha_k p_k > 0$$

$$\Leftrightarrow \quad s_k \underbrace{(\nabla f^T(x_k + \alpha p_k) - \nabla f_k^T)}_{y_k} = (c_2 - 1) \nabla f_k^T s_k > 0$$

$$\Leftrightarrow \quad s_k^T \cdot y_k > 0$$

■

② Verifique que B_{k+1} y H_{k+1} son inversa una de la otra.

$$B_{k+1} H_{k+1} = B_{k+1} \left((I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T \right)$$

$$= (B_{k+1} - \rho_k \underline{B_{k+1} s_k} y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k B_{k+1} s_k s_k^T$$

$$= (B_{k+1} - \rho_k y_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k y_k s_k^T$$

$$= \left(B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} - \rho_k y_k y_k^T \right) H_k (I - \rho_k y_k s_k^T) + \rho_k y_k s_k^T$$

$$= \left(B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \right) H_k (I - \rho_k y_k s_k^T) + \rho_k y_k s_k^T$$

$$= \left(I - \frac{B_k s_k s_k^T}{s_k^T B_k s_k} \right) (I - \rho_k y_k s_k^T) + \rho_k y_k s_k^T$$

$$= I - \rho_k y_k s_k^T - \frac{B_k s_k s_k^T}{s_k^T B_k s_k} + \frac{B_k s_k s_k^T}{s_k^T B_k s_k} (\rho_k y_k s_k^T) + \rho_k y_k s_k^T$$

$$= I - \rho_k y_k s_k^T - \frac{B_k s_k s_k^T}{s_k^T B_k s_k} + \frac{1}{s_k^T y_k} \frac{B_k s_k (s_k^T y_k) s_k^T}{s_k^T B_k s_k} + \rho_k y_k s_k^T$$

$$= I - \cancel{\frac{B_k s_k s_k^T}{s_k^T B_k s_k}} - \cancel{\rho_k y_k s_k^T} + \cancel{\frac{B_k s_k s_k^T}{s_k^T B_k s_k}} + \cancel{\rho_k y_k s_k^T}$$

$$= I \; \blacksquare$$