Maya Denti 181158

1.1 Gradiente Conjugado

- Demuestre que si los vectores no nulos $p_1, p_2, ..., p_l$ satisfacen que :

$$p_i^T A p_j = 0, \forall i \neq j,$$

y A es simétrica y positiva definida, entonces los vectores son linealmente independientes.

Supongamos que son linealmente independientes entonces existen escalares $\alpha_k \neq 0$ para $k \in \{1, 2, \dots i, j\}$ tales que $\alpha_1 p_1 + \alpha_2 p_2 + \dots + \alpha_i p_i + \dots + \alpha_j p_j = 0$.

Multiplico la expresión anterior

$\alpha_1 A p_1 + \alpha_2 A p_2 + \dots + \alpha_i A p_i + \dots + \alpha_j A p_j + \dots = 0$

Premultiplico por p_i^T

$\alpha_1 p_i^T A p_1 + \alpha_2 p_i^T A p_2 + \alpha_1 p_i^T A p_i + \dots + \alpha_j p_j^T A p_j = 0$

Vemos que $p_i^T A p_j = 0 \quad \forall i \neq j$ entonces

$\alpha_i p_i^T A p_i = 0$

Como A es positiva definida $p_i^T A p_i > 0$ entonces podemos dividir y nos queda.

$\alpha_i = 0$!

Así que el conjunto de vectores es linealmente independiente.

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

Tenemos un teorema visto en clase que dice que para cualquier $x_0 \in \mathbb{R}$ la sucesion $\{x_k\}$ que se genera mediante el algoritmo de gradiente conjugado converge a x^* en a lo más n pasos.

Una parte de la justificación de este teorema es que como el conjunto de direcciones $\{p_i\}$ es linealmente independiente (lo que probamos arriba) entonces estas direcciones generan a todo $\mathbb{R}^n$ y en particular $x^* - x_k \in \mathbb{R}^n$ o sea $x^* - x_k \in \text{span}\{p_1, \dots p_n\}$ con $\alpha_k = \dfrac{p_k^T A(x^* - x_0)}{p_k^T A p_k}$.

Así que por propiedad de conjugación podemos minimizar $\phi(\cdot)$ en a lo más n pasos a lo largo de las direcciones conjugadas y entonces G.C. converge en a lo más n pasos.

1.2 BFGS

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

Hace la demostración por inducción. Sea $H_0 B_0 = I$ Supongamos que se cumple que $B_k H_k = I$ (HI)

p.d. $B_{k+1} H_{k+1} = I$

Recordemos que
$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}$$
$$H_{k+1} = (I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T \quad \text{con} \quad \rho_k = \frac{1}{y_k^T s_k}$$

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

ρ_k es escalar
$$= \left[(B_{k+1} - \rho_k B_{k+1} s_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k B_{k+1} s_k s_k^T\right] =$$

Ahora por la ecuación secante $B_{k+1} s_k = y_k$

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

Sustituyo B_{k+1}
$$= \left[\left(B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} - \rho_k y_k y_k^T\right) H_k (I - \rho_k y_k s_k^T) + \rho_k y_k s_k^T\right] =$$

sustituyo ρ_k
$$= \left[\left(B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \cancel{\frac{y_k y_k^T}{y_k^T s_k}} - \cancel{\frac{y_k y_k^T}{y_k^T s_k}}\right) H_k \left(I - \frac{y_k s_k^T}{y_k^T s_k}\right) + \frac{y_k s_k^T}{y_k^T s_k}\right] =$$

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

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

HI

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

multiplico

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

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

$$= I - \frac{B_k s_k s_k^T}{s_k^T B_k s_k} + \frac{B_k s_k s_k^T y_k s_k^T}{s_k^T B_k s_k y_k^T s_k} \underset{☆}{=} I$$

☆ Esto se cumple pues $\frac{B_k s_k s_k^T y_k s_k^T}{s_k^T B_k s_k y_k^T s_k} = \frac{B_k s_k s_k^T}{s_k^T B_k s_k}$

Partimos del lado derecho

$\frac{B_k s_k s_k^T \cancel{y_k s_k^T}}{s_k^T B_k s_k \cancel{y_k^T s_k}} = \frac{B_k s_k s_k^T}{s_k^T B_k s_k}$ pues $y_k s_k^T = y_k^T s_k$ pues y_k, s_k son vectores tales que es un escalar

$\therefore$ $B_{k+1} H_{k+1} = I$ y entonces $B_{k+1}^{-1} = H_{k+1}$ son inversas.