Karla Susana Alva Ortiz
174144

Examen Final

Gradiente Conjugado.

1.- Debido a que se cumple que $p_i^T A p_j = 0 \ \forall i \neq j$, entonces las direcciones $p_1, p_2, \dots, p_\ell$ son conjugadas con respecto a A.

Supongamos que p_k se puede escribir como combinación lineal de $p_1, p_2, \dots, p_\ell$, es decir:

$p_k = a_1 p_1 + a_2 p_2 + \dots + a_\ell p_\ell$, donde $a_i \in \mathbb{R}$, es decir es un coeficiente, con $i = 1, 2, \dots, \ell$.

En particular, como estamos escribiendo p_k como una combinación lineal, entonces $a_1 p_1 + a_2 p_2 + \dots + a_\ell p_\ell$ no tiene al elemento p_k. Ahora determinaremos los valores de a_i

Como tenemos que las direcciones son conjugadas, y $k \neq 1$, entonces se cumple que: $p_1^T A p_k = 0$

$$\Rightarrow 0 = p_1^T A p_k = p_1^T A [a_1 p_1 + a_2 p_2 + \dots + a_\ell p_\ell] = a_1 p_1^T A p_1 + a_2 \underbrace{p_1^T A p_2}_{0} + \dots + a_\ell \underbrace{p_1^T A p_\ell}_{0} \underset{\uparrow}{=} a_1 p_1^T A p_1$$

porque son conjugadas

Entonces tenemos que: $a_1 p_1^T A p_1 = 0 \Rightarrow a_1 = 0$ y $p_1^T A p_1 \neq 0$, ya que A es positiva definida, y p_1 es conjugado respecto a A

El proceso mediante el cual se determinó a_1, se sigue para determinar el resto de los a_i, en particular, tenemos que $k \neq i$, y como las direcciones son conjugadas se cumple que $p_i^T A p_k = 0$, lo que implica que: $a_i p_i^T A p_i = 0 \Rightarrow a_i = 0$, ya que $p_i^T A p_i \neq 0$ debido a que A es positiva definida y p_i es conjugado respecto de A.

Como $a_i = 0 \ \forall i$, entonces tenemos que $p_k = 0 p_1 + 0 p_2 + \dots + 0 p_\ell = 0$. Sin embargo, por hipotesis, tenemos que p_k es un vector no nulo, por lo tanto, p_k no puede ser escrito como una combinación lineal

Debido a que p_k no se puede escribir como una combinación lineal, entonces $p_1, p_2, \dots, p_\ell$ son linealmente independientes.

2.- Como $p_1, p_2, \ldots, p_k$ son linealmente independientes, entonces estos vectores pueden generar a todo $\mathbb{R}^n$.

Sabemos que el tamaño del paso en el gradiente conjugado es $\alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k}$, además, de que $x_{k+1} = x_k + \alpha_k p_k$, de este modo, podemos encontrar x_{k+1} iterativamente, encontrado que:

$x_{k+1} = x_1 + \alpha_1 p_1 + \alpha_2 p_2 + \ldots + \alpha_k p_k$, donde x_1 es el punto inicial.

En particular, si suponemos que el optimo x^* se encuentra en la iteración n, entonces:

$x^* = x_1 + \alpha_1 p_1 + \alpha_2 p_2 + \ldots + \alpha_{n-1} p_{n-1}$

Es decir, dado a que $p_1, p_2, \ldots, p_k$ son linealmente independientes, entonces, el optimo x^* se encontrará gracias a una combinación lineal de los mismos. En este caso, en una combinación lineal de las n-1 direcciones de descenso, por lo cual, se hará en la iteración n.

Por lo tanto, el gradiente conjugado converge en a lo más n iteraciones, ya que se va llegando al optimo minimizando cada una de las direcciones conjugadas p_i.

Quasi - Newton

1.- La segunda condición fuerte de Wolfe, nos dice que $|\nabla f^T(x_k+\alpha_k p_k) p_k| \leq c_2 |\nabla f^T(x_k) p_k|$, donde $0<c_2<1$

De aquí que; quitando el valor absoluto, tenemos que: $\nabla f(x_k+\alpha_k p_k)^T p_k \geq -c_2 |\nabla f^T(x_k) p_k| = c_2 \nabla f^T(x_k) p_k$, ya que como p_k es una dirección de descenso, entonces es negativa.

Si restamos ambos lados por $\nabla f^T(x_k) p_k$, tenemos que:

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

$$\parallel$$

$$[\nabla f(x_k+\alpha_k p_k) - \nabla f^T(x_k)] p_k$$

Como $c_2<1$ y p_k es negativo, entonces tenemos que $^{(1)}[\nabla f^T(x_k+\alpha_k p_k) - \nabla f^T(x_k)] p_k = (c_2-1)\nabla f^T(x_k) p_k > 0$, ya que $c_2-1<0$.

En particular, tenemos que $x_{k+1} = x_k + \alpha_k p_k$. Y por definición del algoritmo BFGS, tenemos que: $y_k = \nabla f_{k+1} - \nabla f_k$ y $s_k = x_{k+1} - x_k = x_k + \alpha_k p_k - x_k = \alpha_k p_k$

Regresando a (1), tenemos que: $[\nabla f^T(x_{k+1}) - \nabla f^T(x_k)] p_k > 0$, multiplicando por α_k, tenemos que: $[\nabla f^T(x_{k+1}) - \nabla f^T(x_k)] \alpha_k p_k > \alpha_k(0) = 0$.

Sustituyendo los valores de y_k, s_k, encontramos que: $y_k^T s_k > 0$, lo cual es equivalente a que $s_k^T y_k > 0$.

$\therefore$ La condición de curvatura cumple que $s_k^T y_k > 0$

2.- Supongamos que B_k y H_k son inversas es decir, se cumple que $B_k H_k = H_k B_k = I$.

Además, tenemos que: $\rho_k = \frac{1}{s_k^T y_k}$ y que $B_{k+1} s_k = y_k$, la cual es la ecuación secante.

Notemos que: (*) $\rho_k \frac{B_k s_k s_k^T}{s_k^T B_k s_k} y_k s_k^T = \frac{1}{s_k^T y_k} \frac{B_k s_k s_k^T}{s_k^T B_k s_k} y_k s_k^T = \frac{B_k s_k s_k^T y_k s_k^T}{(s_k^T y_k) s_k^T B_k s_k} =$

$= \frac{(s_k^T y_k) B_k s_k s_k^T}{(s_k^T y_k) s_k^T B_k s_k} = \frac{B_k s_k s_k^T}{s_k^T B_k s_k}$, ya que $s_k^T y_k$ es un escalar.

Para verificar que B_{k+1} y H_{k+1} son inversas, basta probar que $B_{k+1} H_{k+1} = I$.

Por definición, tenemos que: $H_{k+1} = (I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T$

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \rho_k y_k y_k^T$$

De aquí que:

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

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

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

$\to$ Tenemos que $B_{k+1} s_k = y_k$

$\Rightarrow s_k = \frac{y_k}{B_{k+1}}$ y $B_{k+1} = \frac{y_k}{s_k}$

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

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

$\to$ Usando la definición de B_{k+1}

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

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

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

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

$\to$ Utilizando (*)

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

$$= I$$

Como $B_{k+1} H_{k+1} = I$, entonces B_{k+1} y H_{k+1} son inversas.