José Ángel Rodríguez

Examen Final

① Gradiente Conjugado

1.- $p_1, p_2, \dots, p_l \neq 0$ satisfacen: $p_i^T A p_j = 0 \quad \forall i \neq j$ (conjugados)

y A simétrica y positiva definida entonces los vectores son l.i.

Tomemos una combinación lineal de estos vectores:

$\sum_{i=1}^{l} a_i p_i$ esto es igual a 0 si y sólo si: $p_j^T A \sum_{i=1}^{l} a_i p_i = 0 \quad \forall j \in \{1, \dots, l\}$ $(*)$

Por la propiedad de conjugado:

$(*) \Rightarrow \sum_{i=1}^{l} a_i p_j^T A p_i = 0 \Rightarrow a_j p_j^T A p_j = 0 \quad \forall j \in \{1, \dots, l\}$ como

$p_j \neq 0$ y A es positiva definida $a_j p_j^T A p_j = 0 \Rightarrow a_j = 0$

Entonces sólo nos quedamos con $\sum_{i \neq j}^{l} a_i p_i$ aplicando el mismo método

para todas las "j"'s restantes obtenemos que $a_j = 0 \quad \forall j \in \{1, \dots, l\}$

$\therefore$ el conjunto $\{p_1, \dots, p_l\}$ es l.i.

2° ¿Por qué tenemos convergencia en a lo más l iteraciones?

Como $\{p_1, \dots, p_l\}$ es un conjunto l.i. (1.1) $\operatorname{span}\{p_1, \dots, p_l\} = \mathbb{R}^l$

Entonces cualquier punto en $\mathbb{R}^l$ puede ser escrito como combinación lineal de los p_i's. En particular para x_0 - el punto inicial y x^* el óptimo:

$$x^* - x_0 = \sum_{i=1}^{l} a_i p_i \Rightarrow p_j^T A(x^* - x_0) = p_j^T A \sum_{i=1}^{l} a_i p_i$$

$$\Rightarrow p_j^T A(x^* - x_0) = a_j p_j^T A p_j \quad \text{(conjugados)}$$

$$\Rightarrow a_j = \frac{p_j^T A(x^* - x_0)}{p_j^T A p_j},$$ que está bien definido ya que $p_j^T A p_j \neq 0$ por hipótesis.

Ahora si vamos dando pasos de la forma $x_{k+1} = x_k + \alpha_k p_k$

con $\alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k}$ con $r_k = Ax_k - b$ (del problema de minimización original)

Tenemos que $x_2 = x_1 + \alpha_1 p_1$, $x_3 = x_2 + \alpha_2 p_2 = x_1 + \alpha_1 p_1 + \alpha_2 p_2$

en general $x_j = x_1 + \alpha_1 p_1 + \alpha_2 p_2 + \dots + \alpha_{j-1} p_{j-1}$ (*)

$$\Rightarrow p_j^T A x_j = p_j^T A x_1$$ (*) ya que los otros términos se anulan por ser conjugados

$$\Rightarrow p_j^T A(x_j - x_1) = 0$$

Por (*) $p_j^T A(x^* - x_1) = p_j^T A(x^* - x_j) = p_j^T(Ax^* - Ax_j) = p_j^T(b - Ax_j) = p_j^T(-r_j)$

Por la forma de r_k tenemos que $\frac{p_j^T A(x^* - x_1)}{p_j^T A p_j} = \alpha_j$, entonces: continúa en la próxima

$a_j = \alpha_j$ esto nos dice que $x^* - x_1 = \alpha_1 p_1 + \dots + \alpha_l p_l$ (8.8)

entonces si empezamos en x_1 después de aplicar l pasos

(es posible que terminemos antes en cuyo caso podemos considerar a las α sucesivas como 0)

tenemos $x_l = x_1 + \alpha_1 p_1 + \dots + \alpha_l p_l \overset{(8.9)}{=} x^*$ ↯

2 Quasi Newton

1. La segunda condición fuerte de Wolfe implica la condición de curvatura

$$s_k^T y_k > 0$$

Recordamos las definiciones de s_k, y_k

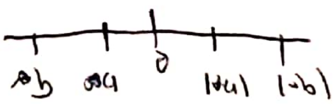
$s_k = x_{k+1} - x_k$ pero $x_{k+1} = x_k + \alpha_k p_k$ entonces $s_k = \alpha_k p_k$

$y_k = \nabla f_{k+1} - \nabla f_k$

Segunda condición fuerte de Wolfe (*)

$$|\nabla f_{k+1}^T p_k| \le c_2 |\nabla f_k^T p_k| \qquad (0 < c_2 < 1)$$

Como p_k es dirección de descenso $\nabla f_{k+1}^T p_k, \nabla f_k^T p_k < 0$



para dos números negativos si $|a| < |b|$ entonces $a > b$

entonces $\nabla f_{k+1}^T p_k \ge c_2 \nabla f_k^T p_k \Rightarrow \nabla f_{k+1}^T p_k - \nabla f_k^T p_k \ge c_2 \nabla f_k^T p_k - \nabla f_k^T p_k$

$\Rightarrow y_k^T p_k \ge (c_2 - 1) \nabla f_k^T p_k$

$\Rightarrow y_k^T s_k \ge \underbrace{(c_2 - 1)}_{<0 \text{ por } (*)} \underbrace{\nabla f_k^T s_k}_{<0}$ (multiplicando por $\alpha_k > 0$)

por p_k ser dirección de descenso y $s_k = \alpha_k p_k$

$\Rightarrow y_k^T s_k \ge (c_2 - 1) \nabla f_k^T s_k > 0$ $\quad \therefore \; s_k^T y_k = (y_k^T s_k)^T > 0$

2.2 B_{n+1} y H_{n+1} son inversos una de la otra.

Recordamos las definiciones de B_{n+1} y H_{n+1} de acuerdo a BFGS con $B_n = H_n^{-1}$

$$B_{n+1} = B_n - \frac{B_n s_n s_n^T B_n}{s_n^T B_n s_n} + \frac{y_n y_n^T}{y_n^T s_n}$$

$$H_{n+1} = (I - \rho_n s_n y_n^T) H_n (I - \rho_n y_n s_n^T) + \rho_n s_n s_n^T$$

con $\rho_n = \frac{1}{y_n^T s_n}$ y H_{n+1} satisface que: $H_{n+1} y_n = s_n$

También vemos que $y_n^T s_n$ y $s_n^T y_n$ son escalares y $\therefore (y_n^T s_n) = (y_n^T s_n)^T = s_n^T y_n$

Buscamos probar que $B_{n+1} H_{n+1} = I$ o equivalentemente que $B_{n+1} H_{n+1} y_n = y_n$

es decir $B_{n+1} s_n = y_n$

$$B_{n+1} s_n = B_n s_n - \frac{B_n s_n s_n^T B_n s_n}{s_n^T B_n s_n} + \frac{y_n y_n^T s_n}{y_n^T s_n}$$

$$= B_n s_n - \frac{\overset{\text{escalar}}{(s_n^T B_n s_n)} B_n s_n}{s_n^T B_n s_n} + \frac{(y_n^T s_n)}{(y_n^T s_n)} y_n$$

$$= B_n s_n - B_n s_n + y_n = y_n \quad \therefore B_{n+1} s_n = y_n \Rightarrow B_{n+1} H_{n+1} y_n = y_n \Rightarrow B_{n+1} H_{n+1} = I$$