Final análisis aplicado 173199

① Gradiente conjugado:

a) P.D: si $p_1, \dots, p_t$ (vectores no nulos) satisfacen que $p_i^T A p_j = 0 \ \forall i \neq j$, A simétrica y definida positiva $\Rightarrow \{p_1, \dots, p_t\}$ es LI

Para que sean LI querríamos que si $\alpha_1 p_1 + \dots + \alpha_t p_t = 0 \Rightarrow \alpha_1 = \alpha_2 = \dots = \alpha_t = 0$

Tomemos algún p_i y expresémoslo como conjugación lineal de los otros $t-1$ vectores. Esto, partiendo de la idea de que ↓ y que $p_i^T A p_j = 0$ si $i \neq j$

$\Rightarrow p_j = \beta_1 p_1 + \beta_2 p_2 + \dots + \beta_{j-1} p_{j-1} + \beta_{j+1} p_{j+1} + \dots + \beta_t p_t$ (con $j = 1, \dots, t$) (si $i+1 > t$, pues $i = t$ entonces no hay problema)

$\Rightarrow A p_j = A(\beta_1 p_1 + \dots + \beta_t p_t) = \beta_1 A p_1 + \dots + \beta_t A p_t$

$\Rightarrow p_i^T A p_j = p_i^T(\beta_1 A p_1 + \dots + \beta_t A p_t) = \underbrace{\beta_1 p_i^T A p_1 + \dots + \beta_t p_i^T A p_t}$

Sabemos que $p_i^T A p_j = 0 \ \forall i \neq j$, entonces todo número en ↓ es cero, a excepción de $\beta_i p_i^T A p_i$ (que existe en la sucesión, pues $i \neq j$ y tomamos todo $i \in \{1, \dots, t\}$ excepto j)

$\therefore p_i^T A p_j = \beta_i p_i^T A p_i$, pero $p_i^T A p_j = 0$ (por ser conjugados)

$\Rightarrow \beta_i p_i^T A p_i = 0$. Ahora bien, $\forall z \in \mathbb{R}^n \setminus \{0\}$, $z^T A z > 0$ por ser definida positiva

$\therefore p_i^T A p_i > 0 \Rightarrow \beta_i = 0$

Pero esto pasa $\forall i \Rightarrow p_j = 0 p_1 + \dots + 0 p_t = 0$ ↯ (son no nulos)

$\therefore$ Son LI (no podemos expresar uno como conjugación lineal de los otros) ■

c) De clase, sabemos que $\forall x_0 \in \mathbb{R}^n$, $\{x_k\}_{k \geq 0}$ converge en a lo más n pasos a x^*, si usamos el método del gradiente conjugado.

¿Cómo, y por qué, se relaciona esto con que los p's sean LI? Pues, como son LI, son una base para $\mathbb{R}^n$, entonces podemos expresar todo vector como conjugación lineal de ellos, y por ende proceder con la demostración hecha en clase (en ese entonces creo que asumimos que eran LI, sin demostrarlo).

② Quasi-Newton

a) P.D: Wolfe fuerte 2 $\Rightarrow s_k^T y_k > 0$

Recordando el 1er parcial, la 2a condición fuerte de Wolfe es:

$|\nabla f(x_k + \alpha_k p_k)^T p_k| \leq c_2 |\nabla f(x_k)^T p_k|$

sabemos que Wolfe fuerte 2 implica Wolfe débil 2 $(\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f(x_k)^T p_k)$,

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

$= c_2 \nabla f(x_k)^T p_k$

Como α_k es cte, $\alpha_k \nabla f(x_k + \alpha_k p_k)^T p_k \geq \alpha_k c_2 \nabla f(x_k)^T p_k$

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

$\Rightarrow \nabla f(x_k + \alpha_k p_k)^T s_k \geq c_2 \nabla f(x_k)^T s_k \quad (*)$

Sabemos que $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$, pero $\nabla f(x_{k+1}) = \nabla f(x_k + \alpha_k p_k)$

$\Rightarrow$ restemos $\nabla f(x_k)^T s_k$ de ambos lados de $(*)$: $\nabla f(x_{k+1})^T s_k - \nabla f(x_k)^T s_k \geq c_2 \nabla f(x_k)^T s_k - \nabla f(x_k)^T s_k$

$\Rightarrow (\nabla f(x_{k+1}) - \nabla f(x_k))^T s_k = y_k^T s_k \geq (c_2 - 1) \nabla f(x_k)^T s_k$

Regresando s_k a $\alpha_k p_k$, $y_k^T s_k \geq (c_2 - 1) \nabla f(x_k)^T (\alpha_k p_k)$. Como $c_2 < 1$, $c_2 - 1 > 0$; como

p_k es la dirección de descenso, $\nabla f(x_k)^T (\alpha_k p_k) > 0 \Rightarrow y_k^T s_k > 0 \Rightarrow s_k^T y_k > 0$ ∎

b) P.D: B_{k+1} y H_{k+1} son inversas, sabiendo que $H_k = B_k^{-1}$

Sabemos que $H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k^T s_k}{y_k s_k}$

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

Vamos a multiplicar H_{k+1} y B_{k+1} y llegar a la identidad, entonces, por simplicidad, olvidaremos de los subíndices k

$$H_{k+1} B_{k+1} = \left(H - \frac{H y y^T H}{y^T H y} + \frac{s^T s}{y s}\right)\left(B - \frac{B s s^T B}{s^T B s} + \frac{y y^T}{y^T s}\right)$$

$$= HB - \frac{HBss^TB}{s^TBs} + \frac{Hyy^T}{y^Ts} - \frac{Hyy^THB}{y^THy} + \frac{Hyy^THBss^TB}{y^THys^TBs} - \frac{Hyy^THyy^T}{y^THyy^Ts} + \frac{s^TsB}{ys} - \frac{s^TsBss^TB}{yss^TBs} + \frac{s^Tsyy^T}{ysy^Ts}$$

Sabemos que $HB = I$

$$\Rightarrow I - \frac{ss^TB}{s^TBs} + \frac{Hyy^T}{y^Ts} - \frac{Hyy^T}{y^THy} + \frac{Hyy^Tss^TB}{y^THys^TBs} - \frac{Hyy^THyy^T}{y^THyy^Ts} + \frac{s^TsB}{ys} - \frac{s^TsBss^TB}{yss^TBs} + \frac{s^Tsyy^T}{ysy^Ts}$$

⇒ queda muy feo ⇒ probemos otra manera :'v

También sabemos que $H_{k+1} = (I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T$, donde

$s_k = x_{k+1} - x_k = \alpha_k p_k$, $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$, $B_{k+1} s_k = y_k$, $H_k B_k = I$, $\rho_k = \frac{1}{y_k^T s_k}$. Además,

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

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

$$= (I - \rho y s^T) B (I - \rho s y^T)(I - \rho s y^T) H (I - \rho y s^T) + (I - \rho y s^T) B (I - \rho s y^T) \rho s s^T$$
$$+ \rho y y^T (I - \rho s y^T) H (I - \rho y s^T) + \rho y y^T \rho s s^T$$

que también está feo :c

¿Quizá solo sustituyendo uno de los dos?

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

No nos gusta que sigue B_{k+1} presente, pues si hay B_k y se multiplica por H_k da I.

Maneras de quitar B_{k+1} : $B_{k+1} s_k = y_k$, $B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \rho_k y_k y_k^T$ (*),

En $B_{k+1} (I - \rho_k s_k y_k^T)$ hay un ↗, y también hay en $B_{k+1} \rho_k s_k s_k^T$ → ¿servirá?

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

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

De hecho, (se parece a (*), pero pasando el $\rho_k y_k y_k^T$ del otro lado

→

(9)

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

Por simplicidad, ignoramos el K

$$\Rightarrow B_{k+1} H_{k+1} = \left(BH - \frac{Bss^TBH}{s^TBs}\right)\left(I - \rho y s^T\right) + \rho y s^T$$

$$= \left(I - \frac{Bss^T}{s^TBs}\right)\left(I - \rho y s^T\right) + \rho y s^T$$

$$= I - \rho y s^T - \frac{Bss^T}{s^TBs} + \frac{Bss^T \rho y s^T}{s^TBs} + \rho y s^T$$

$\Rightarrow$ P.D: $-\cancel{\rho y s^T} - \frac{Bss^T}{s^TBs} + \frac{Bss^T \rho y s^T}{s^TBs} + \cancel{\rho y s^T} = 0$

$\Rightarrow$ P.D: $\frac{Bss^T \rho y s^T}{s^TBs} = \frac{Bss^T}{s^TBs}$

$\Rightarrow$ P.D: $Bss^T \rho y s^T = Bss^T \iff ss^T \rho y s^T = ss^T$ ... Hm ?

¡Ah! ρ es constante e igual a $\frac{1}{y^Ts}$ $\Rightarrow ss^T \rho y s^T = \rho s s^T y s^T = \frac{1}{y^Ts} ss^Tys^T = \frac{1}{y^Ts} s(s^Ty)s^T$ (*)

Es lo mismo $\frac{1}{y^Ts}$ que $\frac{1}{s^Ty}$ $\Rightarrow$ (*) $= \frac{1}{\cancel{s^Ty}} s(\cancel{s^Ty})s^T = ss^T$, que es lo que queríamos demostrar.

Hice un despapaye, entonces va resumen:

Como $\frac{Bss^T}{s^TBs} = \frac{Bss^T \rho y s^T}{s^TBs}$, $I - \rho y s^T - \frac{Bss^T}{s^TBs} + \frac{Bss^T \rho y s^T}{s^TBs} + \rho y s^T = I \Rightarrow B_{k+1} H_{k+1} = I \Rightarrow$ Inversas