José Luis Cordero Rodríguez 164860

Ⓘ Gradiente Conjugado
dem

Examen Final

① Supongamos que $p_0, \ldots, p_l$ son direcciones conjugadas. Expresemos una de esas direcciones, p_i, como una combinación lineal de las demás:

$$p_i = \sigma_0 p_0 + \cdots + \sigma_n p_m$$

para algunos coeficientes σ_k $(k = 0, \ldots, m)$. Notemos que la suma no incluye a p_i. Luego, por propiedades de direcciones conjugadas tenemos:

$$0 = p_0^T A p_i = \sigma_0 p_0^T A p_0 + \cdots + \sigma_n p_0^T A p_m$$
$$= \sigma_0 p_0^T A p_0$$

Lo anterior implica que $\sigma_0 = 0$ porque los vectores $p_0, \ldots, p_m$ son conjugados y la matriz A es positiva definida. Se puede usar el mismo argumento para demostrar que los escalares σ_k con $k = 0, 1, \ldots, m$ son cero.

Esto implica que $p_i = 0$ por la primera ecuación, lo cual contradice que p_i es un vector no nulo. Por lo que los vectores $p_0, \ldots, p_m$ son linealmente independientes

■

Ⓘⓘ Quasi-Newton
dem

① Recordemos que una de las condiciones fuertes de Wolfe es:

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

lo cual implica que:

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

porque p_k es una dirección de descenso. Por lo tanto:

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

Asumiendo que $c_2 < 1$. Finalmente obtenemos el resultado deseado multiplicando ambos lados por α_k y observando que $s_k = \alpha_k p_k$ y $y_k = \nabla f(x_k + \alpha_k p_k) - \nabla f(x_k)$ ■

dem Grediente Conjugado

② Tenemos que probar que B_{k+1} H_{k+1} son inversas la una de la otra. Basta demostrar que $B_{k+1}H_{k+1} = I$. Notemos que $B_{k+1}s_k = y_k$. Tenemos que multiplicando B_{k+1} por la solución H_{k+1} del BFGS es:

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

Para esta demostración se usó el hecho que

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

■

② Q uasi Newton

¿Por qué el gradiente conjugado converge en lo más n iteraciones?

Para checar esto hay que recordar que las direcciones conjugadas $\{p_i\}$ son linealmente independientes como ya habíamos demostrado. A partir de esto sabemos que generan a todo el espacio $\mathbb{R}^n$; es decir forman una base. Por esto, podemos escribir la la diferencia entre x_0 y la sol x^* como una combinación lineal de nuestras direcciones de la siguiente forma:

$$x^* - x_0 = \sigma_0 p_0 + \cdots + \sigma_{n-1} p_{n-1} \qquad (1)$$

con $\sigma_i \in \mathbb{R}$, $i \in \{0,\dots,n-1\}$. A partir de aquí es importante la propiedad de la conjugación pues multiplicando (1) por $p_k^T A$ obtenemos

$$\sigma_k = \frac{p_k^T A(x^* - x_0)}{p_k^T A p_k} \qquad \text{ya que} \quad p_i^T A p_j = 0 \;\; \forall i \neq j$$

¡Además σ_k coincide con los tamaños de paso α_k! Y habíamos visto que junto con las direcciones de descenso $\{p_0, \dots, p_{n-1}\}$ minimizaban $\phi(\cdot)$ en n pasos.

¿Cómo es que coinciden?

Pues tenemos que $x_k = x_0 + \alpha_0 p_0 + \alpha_1 p_1 + \dots + \alpha_{k-1} p_{k-1}$ y multiplicando por $p_k^T A$; tenemos que $p_k^T A(x_k - x_0) = 0$

$$\therefore \; p_k^T A(x^* - x_0) = p_k^T A(x^* - x_k) = p_k^T (b - A x_k) = -p_k^T r_k$$

Esto se puede ver gráficamente también porque las curvas de nivel de ϕ son elipses si A es diagonal. Podemos ir minimizando la función al momento a través de los ejes