Tonantzin Real Rojas
163491

Examen Final Análisis Aplicado

December 12, 2020

1 Gradiente Conjugado

1. 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.

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

2 Quasi-Newton

1. Muestre que la segunda condición fuerte de Wolfe implica la condición de curvatura.

$$s_k^T y_k > 0.$$

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

1 Gradiente Conjugado

1. 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.

sean $p_0, p_1, \ldots, p_l$ vectores conjugados

Expresemos a uno de ellos, p_i, como una combinación lineal de los demás

$$\Rightarrow \quad p_i = \sigma_0 p_0 + \sigma_1 p_1 + \cdots + \sigma_l p_l \qquad (1)$$

para σ_k con $k = 0, 1, \ldots, l$

Notemos que la suma de p_i no incluye a p_i como sumando

ie $p_i = \sum_{\bar{k}=1}^{l} \sigma_{\bar{k}} p_{\bar{k}}$ donde $\bar{k} = \{0, \ldots, l\} \setminus \{i\}$

Entonces como las p_s son conjugadas

$$\Rightarrow \quad 0 = p_0^T A p_i = \sigma_0 p_0^T A p_0 + \cdots + \sigma_l p_0^T A p_l$$

$$= \sigma_0 p_0^T A p_0$$

$\Rightarrow \sigma_0 = 0$ pues $p_0, \ldots, p_l$ son conjugados y A es positiva definida

La demostración es analoga para el resto de los coeficientes σ_k $(k = 0, 1, \ldots, l)$ que son cero

$$\therefore \sigma_k = 0 \quad \forall k = 0, \ldots, l$$

Pero por (1) tenemos que $p_i = 0$! pues p_i no es el vector cero

$\therefore p_0, p_1, \ldots, p_l$ son linealmente independientes $\blacksquare$

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

Recordemos el teorema visto en clase que dice que:

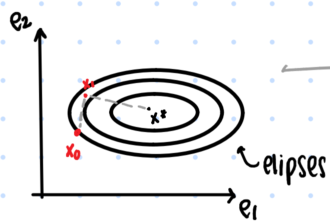
Para cualquier $x_0 \in \mathbb{R}$ en la sucesión $\{x_k\}$ que se genera mediante el algoritmo de dirección conjugada, el α_k converge a la sln. x^* en a lo más n pasos

La idea detrás de esto es que como las direcciones $\{p_i\}$ son linealmente independientes (aquí es donde usamos el resultado anterior) entonces lo que pasa es que estas direcciones generan a todo $\mathbb{R}^n$ y como sabemos que la sucesión de x_k converge a x^* entonces sabemos que $x^* - x_0$ está expresada como una combinación lineal de $\sigma_i p_i$ y las σ_i las expresamos como

$$\sigma_k = \frac{p_k^T A(x^* - x_0)}{p_k^T A p_k}$$

que de hecho coinciden con la longitud del paso mencionada en el teorema

La explicación de por qué σ_k y α_k coinciden es parte de la demostración del teorema entonces para que quede más claro haré una gráfica que ayude a entender mejor cómo se llega a x^* (converge) en n pasos.



y así es como se realizan minimizaciones sucesivas a lo largo de las direcciones coordenadas y converge en n pasos

En otras palabras, la razón por la que, a partir del resultado anterior, el gradiente conjugado converge en a lo más n iteraciones es que por la propiedad de conjugación, podemos minimizar $\phi(\cdot)$ en $\underline{n}$ pasos al minimizarlo sucesivamente a lo largo de cada una de las direcciones conjugadas del conjunto conjugado. Pero esto sucede, como explicamos antes, por la propiedad de conjugación y recordemos que esta está definida en el ejercicio anterior, de ahí que a partir del resultado anterior podamos decir que GC converge en a lo más n pasos pues las p_i son li

2 Quasi-Newton

1. Muestre que la segunda condición fuerte de Wolfe implica la condición de curvatura.

$$s_k^T y_k > 0. \qquad (*)$$

Recordemos que la segunda condición de Wolfe está dada por

$$|\nabla f(x_k + \alpha_k p_k)^T p_k| \leq c_2 |\nabla f(x_k)^T p_k| \qquad \text{con} \quad c_2 < 1$$

lo cual implica que

$$\begin{aligned} \nabla f(x_k + \alpha_k p_k)^T p_k &\geq -c_2 |\nabla f(x_k)^T p_k| \\ &= c_2 \nabla f(x_k)^T p_k \end{aligned}$$

puesto que p_k es una dirección de descenso

$$\begin{aligned} \Longrightarrow \quad \nabla f(x_k + \alpha_k)^T p_k - \nabla f(x_k)^T p_k &= (c_2 - 1) \nabla f(x_k)^T p_k \\ &> 0 \end{aligned}$$

pues habíamos asumido que $c_2 < 1$

por último, obtenemos $(*)$ al multiplicar ambos lados por α_k notando que

$$s_k = \alpha_k p_k$$
$$y_k = \nabla f(x_k + \alpha_k p_k) - \nabla f(x_k)$$

$$\therefore \quad s_k^T y_k > 0$$

∎

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

P.D. $B_{k+1}H_{k+1} = I$

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

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

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

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

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

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

$$= I$$

y recurrimos al hecho de que

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

$$= \frac{B_k s_k s_k^T}{s_k^T B_k s_k}$$

$\therefore B_{k+1}H_{k+1} = I$

$\therefore B_{k+1}$ y H_{k+1} son inversas la una de la otra

$\blacksquare$