Gradiente Conjugado

Optimización no restringida

Gamaliel Moreno

Maestría en Ciencias del Procesamiento de la Información

Ene-Julio, 2020

El método de gradiente conjugado es una técnica matemática que puede ser útil para optimizar problemas lineales y no lineales. Este método fue desarrollado por Magnus Hestenes y Eduard Stiefel.

Hay un conjunto de ecuaciones lineales que queremos resolver representadas en notación vectorial como:

$$\mathbf{Ax} = \mathbf{b}$$

A es una matriz definida simétrica, real y positiva conocida nxn, b es un vector conocido y queremos resolver el vector x. Una solución única a este problema está representada por el vector x^*.
Recuerda

$$f(x) \approx f(x^k) + (x - x^k)^T \nabla f(x^k) + \frac{1}{2}(x - x^k)^T H(x^k)(x - x^k)$$

Derivando

$$\nabla f(x) \approx \nabla f(x^k) + H(x^k)(x - x^k)$$

Igualando a cero y siendo H sdp ($x^T Hx > 0$)

$$H(x^k)(x - x^k) = -\nabla f(x^k)$$

Hay un conjunto de vectores conjugados P donde $P = \{p_k : \forall i \neq k, i, k \in [1, n], \langle p_i, p_k \rangle\}$. Usando este conjunto P l solución x^* puede ser expresado como

$$x^* = \sum_{i=1}^{n} \alpha_i p_i$$

y de esto se obtiene

$$b = Ax^* = \sum_{i=1}^{n} \alpha_i A p_i$$

sustituyendo

$$p_k^T b = \alpha_k p_k^T A p_k$$

entonces el coeficiente α puede calcularse con

$$\alpha_k = \frac{\langle p_k, b \rangle}{\|p_k\|^2}$$

Para el caso especifico de minimización

$$H(x^k)(x - x^k) = -\nabla f(x^k)$$

Esto sugiere tomar el primer vector base $p0$ como negativo del gradiente de f en $x = x_0$. Comenzando con una suposición inicial x_0, esto significa que tomamos $p_0 = b - Ax_0$. Los otros vectores en la base se conjugarán con el gradiente, de ahí el nombre de método de *gradiente conjugado*.

El método de gradiente conjugado es un método de primer orden, pero tiene la propiedad de la convergencia cuadrática y, por lo tanto, tiene una ventaja significativa sobre los métodos de segundo orden. Se dice que dos direcciones, S_1 y S_2, se conjugan si

$$S_1^T H S_2 = 0$$

donde H es una matriz simétrica.

Método gradiente conjunado Fletcher-Reeves

Las direcciones ortogonales son direcciones conjugadas. Iniciando en el punto x_{1a}, La dirección de búsqueda S_1 resulta en el punto mínimo x_a^*. Similarmente, iniciando en el punto x_{1b}. El resultado de la búsqueda en dirección S_1 en el punto mínimo x_b^*. La linea que une x_a^* y x_b^* es la búsqueda de dirección S_2, entonces S_1 y S_2 son direcciones conjugadas.

El método del paso más descendente fue modificado por Fletcher y Reeves en el método del gradiente conjugado. Iniciando con la dirección de búsqueda

$$S_1 = -\nabla f(x_1)$$

y subsecuentemente la dirección de búsqueda es tomado como combinanción lineal de S_1 y $-\nabla f(x_2)$. Eso es

$$S_2 = -\nabla f(x_2) + \alpha S_1$$

Usando la propiedad de $S_1^T H S_2 = 0$ de las direcciones conjugadas, α puede ser evaluada con

$$\alpha = \frac{||\nabla f(x_{i+1})||^2}{||\nabla f(x_i)||^2}$$

Iniciando con $S_1 = -\nabla f(x_1)$, la dirección de búsqueda en cada iteración es calculada usando la ecuación

$$S_{i+1} = -\nabla f(x_i) + \frac{||\nabla f(x_{i+1})||^2}{||\nabla f(x_i)||^2} S_i$$

Método gradiente conjunado Fletcher-Reeves