Tarea 7 - Optimización 1

1st Daniel Vallejo Aldana
Departamento de Matemáticas
Universidad de Guanajuato
daniel.vallejo@cimat.mx

***Resumen*—En el presente trabajo se presentará el algoritmo de gradiente conjugado no lineal así como sus modificaciones como lo es el caso del algoritmo de Polak-Ribiere y el algoritmo de Hestenes-Stiefel.**

***Index Terms*—Gradiente conjugado no lineal, Fletcher-Rieves, Polak-Ribiere,Hestenes-Stiefel.**

I. Introducción

El método de gradiente conjugado no lineal busca adaptar el método de gradiente conjugado lineal utilizado para minimizar la función cuadrática convexa $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^t Q\mathbf{x} - b^t\mathbf{x}$ para cualquier función convexa o incluso funciones no lineales f.

Existen diversos métodos que buscan resolver lo anterior, el primero de ellos es el método de Fletcher-Reeves, la idea de este algoritmo es modificar el algoritmo de gradiente conjugado lineal y, para una longitud de paso α_k se necesita calcular una búsqueda en linea que encuentre un mínimo aproximado de la función no lineal f a lo largo de la dirección p_k, luego reemplazamos el residuo debe de ser reemplazado por el gradiente de la función no lineal objetivo f.

Para que la dirección p_k sea una dirección de descenso se necesita que el tamaño de paso α_k sea exacto o que satisfaga las condiciones fuertes de Wolfe, las cuales son las siguientes.

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f(x_k)^t p_k$$
$$|\nabla f(x_k + \alpha_k p_k)^t p_k| \leq -c_2 \nabla f(x_k)^t p_k$$

Para constantes $0 < c_1 < c_2 < \frac{1}{2}$ de esta forma se garantiza que la dirección p_k es realmente de descenso.

Existen variaciones del método de Fletcher-Reeves, las cuales son el método de Polak-Ribier y el método de Hestenes-Stiefel donde el cambio se encuentra en el cálculo de β_{k+1}

II. Método/Algoritmo

Se dará el esquema general del algoritmo de gradiente conjugado no lineal, y posteriormente se describirán cada una de sus variantes como lo son Fletcher-Reeves, Polak-Ribiere y Hestenes-Stiefel. El esquema general de los algoritmos de gradiente conjugado no lineal es el siguiente

Input: x_0
Output: x^*
$d_0 = -g_0$, $k = 0$
while $\|g_k\| \neq 0$ **do**
    Calcular α_k con un método de búsqueda en linea
    $x_{k+1} = x_k + \alpha x_k$
    Calcular $\nabla f(x_{k+1})$
    Calcular β_{k+1}
    $d_{k+1} = -\nabla f(x_{k+1}) + \beta_{k+1} d_k$
    $k = k + 1$
end while

Cada algoritmo de los probados en esta tarea, varía de los demás por la consideración de β en el caso del algoritmo de Fletcher-Reeves se tiene el siguiente cálculo para β. Definamos $g_k = \nabla f(x_k)$

$$\beta_{k+1}^{FR} = \frac{g_{k+1}^t g_{k+1}}{g_k^t g_k}$$

Para el caso de la modificación al algoritmo de gradiente conjugado de Polak-Ribiere se tiene la siguiente consideración para β.

$$\beta_{k+1}^{PR} = \frac{g_{k+1}^t (g_{k+1} - g_k)}{g_k^t g_k}$$

Para asegurar que la β anterior garantice descenso se toma la consideración $\beta_{k+1}^{+PR} = max(0, \beta_{k+1}^{PR})$

Finalmente para la implementación del algoritmo de Hestenes-Stiefel se tiene la siguiente consideración para β.

$$\beta_{k+1}^{HS} = \frac{g_{k+1}^t (g_{k+1} - g_k)}{(g_{k+1} - g_k)^t d_k}$$

Para probar los algoritmos anteriores se consideraron dos funciones, la función Rosenbrock con n=100 y la función Wood, la expresión de la función Rosenbrock es la siguiente.

$$f(\mathbf{x}) = \sum_{i=1}^{n-1} \left(100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2\right)$$

Con punto inicial $x_0 = [-1{,}2, 1, 1, ..., 1, -1{,}2, 1]$. Se propuso un punto inicial dado por lo siguiente $x_0 = [1{,}2, 1, 1, ..., 1, -1{,}2, 1]$.
Y la función Wood se define de la siguiente forma

$$\begin{aligned} f(\mathbf{x}) = {} & 100(x_1^2 - x_2)^2 + (x_1 - 1)^2 \\ & + (x_3 - 1)^2 + 90(x_3^2 - x_1)^2 \\ & + 10{,}2\left[(x_2 - 1)^2 + (x_4 - 1)^2\right] \\ & + 19{,}8(x_2 - 1)(x_4 - 1) \end{aligned}$$

Para este caso se consideró el punto inicial $x_0 = [-3, -1, -3, -1]$. Además se propuso el punto inicial $x_0 = [1{,}2, 1{,}2, 1{,}2, 1{,}2]$.

Para encontrar α_k con un método de búsqueda en linea se consideró $c_2 = 0{,}44$ y $c_1 = 0{,}2$ ya que con base en las notas las constantes c_1 y c_2 deben de satisfacer las desigualdades $0 < c_1 < c_2 < \frac{1}{2}$ para el caso del algoritmo de Fletcher-Reeves.

III. Análisis de resultados

Se generaron las tablas correspondientes por cada uno de los algoritmos para cada una de las funciones que se encuentran en el apéndice A, cada tabla contiene el número de iteraciones de cada algoritmo, la norma del gradiente y el valor de la función en ese caso en comparación al valor real del óptimo.

IV. Conclusión

Con base en los resultados obtenidos en este trabajo podemos concluir que el método de gradiente conjugado no lineal da buenos resultados en las modificaciones de Polak-Ribiere y Hestenes-Stiefel ya que en ambas modificaciones se llega al óptimo global en la mayoría de las elecciones de x_0 y las iteraciones son menores que en el método de Fletcher-Reeves. Este último método llega al óptimo en un número considerablemente mayor de iteraciones que los otros dos métodos, en este caso se utilizó una tolerancia de 1×10^{-4} para el gradiente.

Apéndices

Apéndice A

Función Rosenbrock para n=100

Fletcher-Reeves: Se generaron las siguientes tablas para el punto inicial $x_0 = [-1{,}2, 1, 1, ..., 1, -1{,}2, 1]$.

Fletcher-Reeves		
Iteraciones	$\|g_k\|$	$\|f(x_k) - f(x^*)\|$
1	309.743409087	182.218793162
2	161.295100863	114.66688225
3	122.762700002	70.591739898
1111	0.0001036355	3.986623854
1112	0.000101218381	3.9866238546
1113	9.912326681e-05	3.9866238546

En este caso vemos que el algoritmo encontró un óptimo local ya que el gradiente vale 0 en ese punto sin embargo la función evaluada en ese punto no corresponde al óptimo global de la función.

Polak-Ribiere:

Polak-Ribiere		
Iteraciones	$\|g_k\|$	$\|f(x_k) - f(x^*)\|$
1	309.7434090	182.2187931
2	160.2131560	118.584318
3	166.59416102	63.329037
254	0.000113470	3.98662385965
255	0.000145847751	3.986623859
256	8.9691108e-05	3.986623859

Hestenes-Stiefel:

Hestenes-Stiefel		
Iteraciones	$\|g_k\|$	$\|f(x_k) - f(x^*)\|$
1	309.74340908	182.2187931
2	160.2131560	118.5843183
3	166.594161	63.32903733
153	0.000182905	3.986623854
154	0.00017936378	3.98662385
155	6.85922160e-05	3.98662385

Para este caso de punto inicial podemos notar que ninguno de los algoritmos encontró el óptimo global, pero encontraron un óptimo local.

Se generaron las siguientes tablas para el punto inicial $x_0 = [1{,}2, 1, 1, ..., 1, -1{,}2, 1]$.

Fletcher-Reeves		
Iteraciones	$\|g_k\|$	$\|f(x_k) - f(x^*)\|$
1	309.2692259	177.975167
2	161.1434655	110.626150
3	122.75466286	66.589451
1147	0.0001066933864	2.721851e-11
1148	0.00010308757	2.69750813e-11
1149	9.8036493e-05	2.6511388e-11

Para este caso notamos que el algoritmo de gradiente conjugado no lineal de Fletcher-Reeves encontró el óptimo global de la función.

Polak-Ribiere:

Polak-Ribiere		
Iteraciones	$\|g_k\|$	$\|f(x_k) - f(x^*)\|$
1	309.2692259	177.97516
2	160.060843	114.530339
3	166.7577956	59.340718
208	0.0002094630602	2.538136503e-09
209	0.00014284503	2.51067021e-09
210	6.4551477e-05	2.48185508e-09

Hestenes-Stiefel:

Hestenes-Stiefel		
Iteraciones	$\|g_k\|$	$\|f(x_k) - f(x^*)\|$
1	309.26922	177.975167
2	160.060843	114.5303392
3	166.75779	59.340718
188	0.000169902006	1.04434907e-09
189	0.0001351762	1.00722595e-09
190	5.11158151e-05	9.955138137e-10

Para este caso de punto inicial podemos notar que en todas las variantes del algoritmo de gradiente conjugado fue posible llegar al óptimo global de esta función.

A partir de esta tabla se consideró el formato de los números flotantes a una expansión con 4 decimales para estandarizar el tamaño de las tablas.

V. Función Wood

Fletcher-Reeves: Se generaron las siguientes tablas para el punto inicial $x_0 = [-3, -1, -3, -1]$.

Fletcher-Reeves		
Iteraciones	$\|g_k\|$	$\|f(x_k) - f(x^*)\|$
1	258.7141	160.2761
2	8.2509	35.1420
3	8.1599	35.0125
6408	0.0001	0.0000
6409	0.0001	0.0000
6410	0.0001	0.0000

Polak-Ribiere:

Polak-Ribiere		
Iteraciones	$\|g_k\|$	$\|f(x_k) - f(x^*)\|$
1	3063.3315	2261.4021
2	1229.3338	734.8549
3	493.2486	245.6652
297	0.0003	0.0000
298	0.0002	0.0000
299	0.0001	0.0000

Hestenes-Stiefel:

Hestenes-Stiefel		
Iteraciones	$\|g_k\|$	$\|f(x_k) - f(x^*)\|$
1	3063.3315	2261.4021
2	1229.3338	734.8549
3	493.2486	245.6652
169	0.0001	0.0000
170	0.0002	0.0000
171	0.0001	0.0000

Se generaron las siguientes tablas para el punto inicial $x_0 = [1{,}2, 1{,}2, 1{,}2, 1{,}2]$.

Fletcher-Reeves		
Iteraciones	$\|g_k\|$	$\|f(x_k) - f(x^*)\|$
1	258.7141	160.2761
2	8.2509	35.1420
3	8.1599	35.0125
6408	0.0001	0.0000
6409	0.0001	0.0000
6410	0.0001	0.0000

Polak-Ribiere:

Polak-Ribiere		
Iteraciones	$\|g_k\|$	$\|f(x_k) - f(x^*)\|$
1	3063.3315	2261.4021
2	1229.3338	734.8549
3	493.2486	245.6652
297	0.0003	0.0000
298	0.0002	0.0000
299	0.0001	0.0000

Hestenes-Stiefel:

Hestenes-Stiefel		
Iteraciones	$\|g_k\|$	$\|f(x_k) - f(x^*)\|$
1	3063.3315	2261.4021
2	1229.3338	734.8549
3	493.2486	245.6652
169	0.0001	0.0000
170	0.0002	0.0000
171	0.0001	0.0000

Referencias

[1] Nocedal Jorge, *Numerical Optimization*, Segunda edición, Springer, 2006