

Tarea 7. Optimización, Métodos de Gradiente Conjugado No Lineal

Oscar Esaú Peralta Rosales
Maestría en Computación
Centro de Investigación en Matemáticas

Resumen—En esta tarea se presenta una introducción e implementación a tres variantes del algoritmo de Gradiente Conjugado No Lineal. La primera de ellas es a través del método de Fletcher–Reeves (FR), la segunda con el método de Polak-Ribière en combinación con el método de Fletcher–Reeves (FR-PR) y la tercera con el método de Hestenes-Stiefel (HS). Se presentan los resultados obtenidos sobre dos funciones conocidas la de Rosembrock y la de Wood.

Index Terms—Gradiente Conjugado No lineal

I. INTRODUCTION

El método de Gradiente Conjugado resuelve el problema de minimización sobre la función cuadrática convexa $\phi(x) = \frac{1}{2}x^T A x - b^T x$. Podemos adaptar este algoritmo para resolver problemas de minimización sobre funciones convexas más generales o incluso en funciones no lineales.

Fletcher y Reeves mostraron que se podía extender el método de Gradiente Conjugado a funciones no lineales realizando dos cambios sobre el algoritmo. El primero de ellos es realizar el cálculo del tamaño de paso α_k a través de una búsqueda en línea sobre aproxime el mínimo de la función no lineal sobre la dirección d_k , el cálculo del gradiente g_k (o el residuo) del algoritmo de Gradiente Conjugado debe ser remplazado por el valor del gradiente de la función no lineal.

Algorithm 1: Algoritmo Gradiente Conjugado No Lineal

Result: x^*
 $x_0 \leftarrow$ Inicializar
 $g_0 = \nabla f(x_0)$
 $d_0 = -g_0$
while $\|g_k\| > tol$ **do**
 $\alpha_k =$ Calcular usando un método de búsqueda en línea
 $x_{k+1} = x_k + \alpha_k d_k$
 $g_{k+1} = \nabla f(x_{k+1})$
 $\beta_{k+1}^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_{k+1}}$
 $d_{k+1} = -g_{k+1} + \beta_{k+1}^{FR} d_k$
end

La otra variante del método de Fletcher–Reeves es el método de Polak-Ribière, el cual remplace el cálculo de β_{k+1}^{FR} por

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

Sin embargo con las condiciones de Wolf no se garantiza que d_k siempre sea una dirección de descenso, una solución es tomar el tomar $\beta_{k+1}^+ = \max\{\beta_{k+1}^{PR}, 0\}$. Por otro lado, se puede garantizar convergencia para cualquier parámetro β_k que satisfaga $|\beta_k| \leq \beta_k^{FR}$ para todo $k \geq 2$, así se puede modificar el método de Polak-Ribière de la forma:

$$b_k = \begin{cases} -\beta_k^{FR} & \text{si } \beta_k^{PR} < -\beta_k^{FR} \\ \beta_k^{PR} & \text{if } |\beta_k^{PR}| \leq \beta_k^{FR} \\ \beta_k^{FR} & \text{if } \beta_k^{PR} > \beta_k^{FR} \end{cases}$$

El tercer método es el de Hestenes-Stiefel el cual define:

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

Estos dos métodos coinciden cuando el modelo es cuadrático convexo y se elige el paso exacto que mimiza la función sobre la dirección d_k reduciéndose así al método de Gradiente Conjugado Lineal.

II. METODOLOGÍA

En previas tareas hemos trabajado con las funciones de Rosembrock con $n = 100$ y la de Wood, así omitiremos lo relacionado con el cálculo de su gradiente. En este caso para la obtención del parámetro α_k se debe obtener mediante algún proceso de búsqueda en línea. Se optó por usar el método de backtracking tal que α_k satisfaga las condiciones de Wolf.

II-A. Algoritmo para obtención de tamaño de paso Backtracking

Algorithm 2: Método de Backtracking para tamaño de paso

Result: α
Inicializar α , ρ y c_1
while $f(x_k + \alpha d_k) > f(x_k) + c_1 \alpha g_k^T d_k$ **do**
 $\alpha = \alpha \rho$
end

La evaluación del algoritmo sobre las funciones se realizaron sobre dos puntos cada una. Para

la función de Rosembrock se usaron los puntos $x^0 = [-1, 2, 1, 1, \dots, 1, -1, 2, 1]$ y $x^0 = [1, \dots, x_4 = 0, 5, 1, \dots, x_4 = 0, 5, 1, \dots, 1]$. Para la función de Wood se usaron los puntos $x^0 = [-3, -1, -3, -1]$ y $x^0 = [0, 7, 0, 5, 0, 1, 0, 8]$

Los resultados obtenidos se muestran en la siguiente sección.

III. RESULTADOS

Las tablas I y II muestran los resultados para dos puntos iniciales para la función de Rosembrock. Con el primer punto solo se alcanzó la convergencia a un mínimo local ya conocido mientras que con el otro punto se alcanzó el mínimo global.

Cuadro I

TABLA COMPARATIVA DE RESULTADOS Y PARÁMETROS PARA EL PUNTO INICIAL $x^0 = [-1, 2, 1, 1, \dots, 1, -1, 2, 1]$ EN LA FUNCIÓN DE ROSEMBROCK

<i>Método</i>	<i>Iters</i>	$\ \nabla g(x^*)\ $	$f(x^*)$	α_0	c_1	ρ
FR	640	9.86e-9	3.99	0.01	0.5	0.8
FRPR	328	8.56e-9	3.99	0.01	0.4	0.8
HS	268	6.38e-9	3.99	0.01	0.4	0.8

Cuadro II

TABLA COMPARATIVA DE RESULTADOS Y PARÁMETROS PARA EL PUNTO INICIAL $x^0 = x^0 = [1, \dots, x_4 = 0, 5, 1, \dots, x_4 = 0, 5, 1, \dots, 1]$ EN LA FUNCIÓN DE ROSEMBROCK

<i>Método</i>	<i>Iters</i>	$\ \nabla g(x^*)\ $	$f(x^*)$	α_0	c_1	ρ
FR	41	9.86e-9	9.81e-20	0.01	0.5	0.8
FRPR	69	7.15e-9	3.98e-20	0.01	0.4	0.8
HS	39	4.90e-9	1.51e-20	0.01	0.4	0.8

Las tablas III y IV muestran los resultados para dos puntos iniciales para la función de Wood. Con ambos puntos se alcanzó la convergencia al mínimo global.

Cuadro III

TABLA COMPARATIVA DE RESULTADOS Y PARÁMETROS PARA EL PUNTO INICIAL $x^0 = [-3, -1, -3, -1]$ EN LA FUNCIÓN DE WOOD

<i>Método</i>	<i>Iters</i>	$\ \nabla g(x^*)\ $	$f(x^*)$	α_0	c_1	ρ
FR	280	6.57e-9	2.22e-17	0.005	0.3	0.8
FRPR	157	5.32e-9	1.61e-18	0.01	0.3	0.8
HS	209	8.25e-9	2.16e-18	0.01	0.4	0.8

Cuadro IV

TABLA COMPARATIVA DE RESULTADOS Y PARÁMETROS PARA EL PUNTO INICIAL $x^0 = [0, 7, 0, 5, 0, 1, 0, 8]$ EN LA FUNCIÓN DE WOOD

<i>Método</i>	<i>Iters</i>	$\ \nabla g(x^*)\ $	$f(x^*)$	α_0	c_1	ρ
FR	217	7.50e-9	4.83e-17	0.01	0.5	0.8
FRPR	176	6.79e-9	1.30e-18	0.01	0.3	0.8
HS	165	5.635e-9	3.82e-20	0.01	0.4	0.8

IV. CONCLUSIONES

Los métodos HS y FR-PR se comportaron mucho mejor que el metodo de FR. Para poder obtener resultados buenos con FR se tuvo que afinar mucho más los parámetros en el algoritmo de backtracking para la obtención de tamaño de paso. Las convergencias fueron en un número de iteraciones pequeños, logrando salir del mínimo local en la función de Wood, en la función se Rosembrock solo se alcanzó el minimo global usando otro punto inicial distinto al propuesto en la definición de la tarea. Además, los puntos a elegir deben estar cercanos a la solución, se observó que en varias de las pruebas que al ser estos elegidos aleatoriamente la solución divergia.

V. APÉNDICE

V-A. Gráficos complementarios

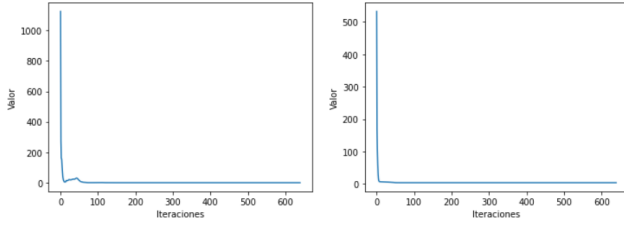


Figura 1. Gráficos de $\|g_k\|$ (izquierda) y $f(x_k)$ (derecha) a través de las iteraciones de la función de Rosenbrock con GC-FR no lineal

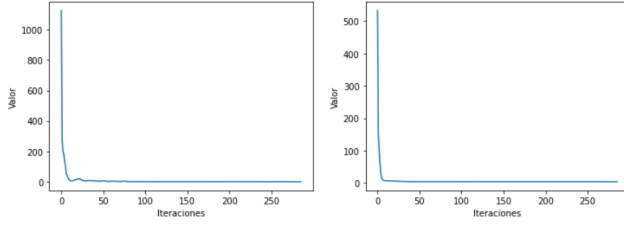


Figura 2. Gráficos de $\|g_k\|$ (izquierda) y $f(x_k)$ (derecha) a través de las iteraciones de la función de Rosenbrock con GC-FRPR no lineal

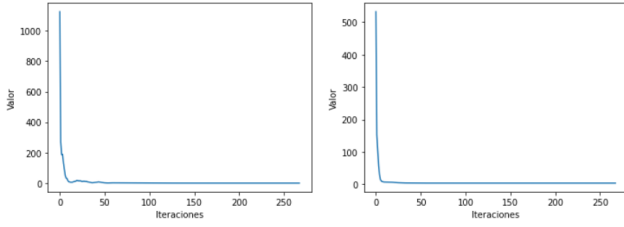


Figura 3. Gráficos de $\|g_k\|$ (izquierda) y $f(x_k)$ (derecha) a través de las iteraciones de la función de Rosenbrock con GC-HS no lineal

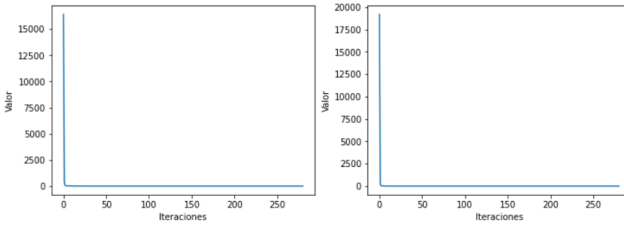


Figura 4. Gráficos de $\|g_k\|$ (izquierda) y $f(x_k)$ (derecha) a través de las iteraciones de la función de Wood con GC-FR no lineal

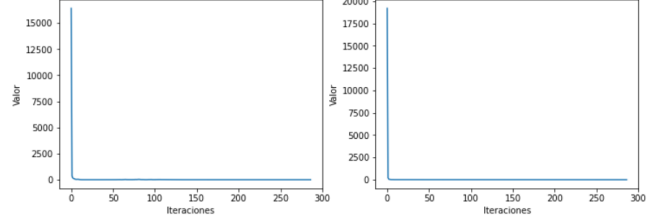


Figura 5. Gráficos de $\|g_k\|$ (izquierda) y $f(x_k)$ (derecha) a través de las iteraciones de la función de Wood con GC-FRPR no lineal

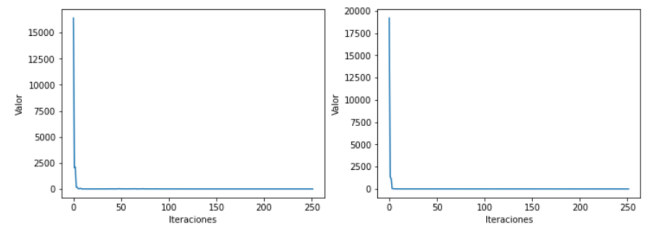


Figura 6. Gráficos de $\|g_k\|$ (izquierda) y $f(x_k)$ (derecha) a través de las iteraciones de la función de Wood con GC-HS no lineal

REFERENCIAS

- [1] Jorge Nocedal, Stephen J. Wright, “Numerical Optimization,” Second Edition, Springer.