

**Tarea 10 - Optimización**  
**Giovanni Gamaliel López Padilla**

## Índice

1. Introducción	1
2. Métodos	2
2.1. Gradiente Conjugado . . . . .	2
2.2. Fletcher-Reeves . . . . .	2
2.3. Polak-Ribiere . . . . .	3
2.4. Hestenes-Stiefel . . . . .	3
2.5. Fletcher-Reeves Polak-Ribiere . . . . .	3
2.6. Función de suavizado . . . . .	3
3. Resultados	4
3.1. Evolución temporal de la solución . . . . .	4
3.2. Suavizado de imágenes . . . . .	4
4. Conclusiones	5
5. Referencias	6

## 1. Introducción

El método del CG fue presentado por Hestenes y Stiefel<sup>1</sup> como un método directo, mas tarde Reid<sup>2</sup> y Concus<sup>3</sup> le descubrieron su verdadero potencial al mirarlo como un método iterativo bastante adecuado para resolver sistemas de la forma  $Ax = b$ , donde  $A$  es simétrica y definida positiva. El método de CG está basado en un principio de conjugación teniendo un almacenamiento modesto y siempre converge.

El problema de minimización a tratar será el caso de una función cuadrática, la tiene asociada una matriz simétrica y definida positiva. Esta definición se encuentra escrita en la ecuación 1.

$$f(x) = \frac{1}{2}x^T Ax - x^T b \quad A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n \quad (1)$$

La solución analítica de la ecuación 1 es  $Ax = b$ , el cual se trata de un sistema de  $n$  ecuaciones. El método de GC compromete a una clase de funciones sin restricciones. El método se caracteriza por la baja memoria que requiere para adquirir propiedades de una convergencia local o global. El método de GC genera una secuencia de pasos  $x_k$  para  $k > 1$  a partir de la ecuación 2. El primer paso ( $x_0$ ) siempre es dado.

$$x_{k+1} = x_k + \alpha_k d_k \quad (2)$$

donde  $\alpha_k$  es un parámetro positivo determinado por una búsqueda en linea y  $d_k$  son las direcciones de descenso. La secuencia  $d_k$  es obtenida con la ecuación 3.

$$d_{k+1} = \beta_k d_k - g_{k+1} \quad d_0 = -g_0 \quad (3)$$

En la ecuación 3, el término  $\beta_k$  es un parámetro que es obtenido por el método GC. En este trabajo nos centraremos en optimizar una función usando el método GC con  $\beta_k$  obtenidas a partir de los algoritmos de Fletcher-Reeves (FR), Polak-Ribiere (PR), Hestenes-Stiefel (HS) y una combinación de Fletcher-Reeves con Polak-Ribiere (FR-PR).

## 2. Métodos

### 2.1. Gradiente Conjugado

De manera general, el algoritmo para el método de Gradiente Conjugado se encuentra descrito en .

---

**Input:**  $x_0, d_0$   
**Output:**  $x^*$

```

1  $d_0 \leftarrow -\nabla f(x_0)$ 
2 while  $\|g(x_k)\| < \tau$  do
3    $\alpha_k \leftarrow$  line search
4   Calcular  $\nabla f(x_{k+1})$ 
5   Calcular  $\beta_{k+1}$ 
6    $d_{k+1} = \beta_{k+1} d_k - \nabla f(x_{k+1})$ 

```

---

La razón por la que  $\alpha_k$  es calculado usando una búsqueda en linea es debido a que así se asegura que la dirección  $d_{k+1}$  es de descenso. Esto es debido a que  $d_{k+1}^T d_k = 0$ , entonces

$$\begin{aligned} d_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \beta_{k+1}^{FR} g_{k+1}^T d_k \\ &= -\|g_{k+1}\|^2 \\ &< 0 \end{aligned}$$

Si  $\alpha_{k+1}$  no es un tamaño de paso exacto. El término  $\beta_{k+1}^{FR} g_{k+1}^T d_k$  podría dominar provocando que no se garantice el descenso. Para el caso cuando  $\alpha_{k+1}$  no es un tamaño de paso exacto, se puede garantizar el descenso si se usan las condiciones fuertes de Wolfe. Esto es

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + c_1 \alpha_k g_k^T d_k \\ |g_{k+1}^T d_k| &\leq -c_2 g_k^T d_k \end{aligned}$$

con  $0 < c_1 < c_2 < 1$ . Para el caso del algoritmo de gradiente Conjugado, el valor de  $c_2 = 0.1$  garantiza la convergencia.

### 2.2. Fletcher-Reeves

El algoritmo de Fletcher-Reeves<sup>4</sup> se basa en encontrar una secuencia de  $\beta_k$  descritas en la ecuación 4.

$$\beta_{k+1}^{FR} = \frac{\nabla^T f(x_{k+1}) \nabla f(x_{k+1})}{\nabla^T f(x_k) \nabla f(x_k)} \quad (4)$$

## 2.3. Polak-Ribiere

El algoritmo de Polak-Ribiere<sup>5</sup> se basa en encontrar una secuencia de  $\beta_k$  descritas en la ecuación 5.

$$\beta_{k+1}^{PR} = \frac{\nabla^T f(x_{k+1})(\nabla f(x_{k+1}) - \nabla f(x_k))}{\nabla^T f(x_k) \nabla f(x_k)} \quad (5)$$

Es conveniente elegir el valor máximo entre  $\beta_{k+1}^{PR}$  y 0. Esto es debido a que existen funciones las cuales pueden ciclarse de forma indefinida.

## 2.4. Hestenes-Stiefel

El algoritmo de Hestenes-Stiefel<sup>6</sup> se basa en encontrar una secuencia de  $\beta_k$  descritas en la ecuación 6.

$$\beta_{k+1}^{HS} = \frac{\nabla^T f(x_{k+1})(\nabla f(x_{k+1}) - \nabla f(x_k))}{(\nabla f(x_{k+1}) - \nabla f(x_k))^T d_k} \quad (6)$$

## 2.5. Fletcher-Reeves Polak-Ribiere

El algoritmo de Fletcher-Reeves en combinación con Polak-Ribiere<sup>7</sup> se basa en encontrar una secuencia de  $\beta_k$  descritas en la ecuación 7.

$$\beta_{k+1}^{FR-PR} = \begin{cases} -\beta_k^{FR} & \text{si } \beta_k^{PF} < -\beta_k^{FR} \\ \beta_k^{PR} & \text{si } |\beta_k^{PF}| \leq \beta_k^{FR} \\ \beta_k^{FR} & \text{si } \beta_k^{PF} > \beta_k^{FR} \end{cases} \quad (7)$$

La secuencia  $\beta_{k+1}^{FR-PR}$  garantiza la convergencia, debido a que  $\beta_k \leq \beta_k^{FR}$  para  $k > 2$ .

## 2.6. Función de suavizado

En este trabajo utilizaremos como función de costo la ecuación 8.

$$f(x) = \sum_{ij} \left[ (x_{i,j} - g_{i,j})^2 + \lambda \sum_{(l,m) \in \Omega_{i,j}} \sqrt{(x_{i,j} - x_{l,m})^2 + \mu} \right] \quad (8)$$

donde  $g$  es la función a suavizar, en este caso será la imagen [lena.png](#),  $\lambda$  es un parámetro positivo a elegir,  $\mu$  es un parámetro fijo con valor de 0.01 y el conjunto  $\Omega_{i,j}$  esta definido de la siguiente manera:

$$\Omega_{i,j} = \{(i+1, j), (i-1, j), (i, j+1), (i, j-1)\}$$

El gradiente de la ecuación 8 se encuentra descrito en la ecuación 9.

$$\frac{\partial f(x)}{\partial x_{i,j}} = 2(x_{i,j} - g_{i,j}) + \lambda \sum_{(l,m) \in \Omega_{i,j}} \frac{x_{i,j} - x_{l,m}}{\sqrt{(x_{i,j} - x_{l,m})^2 + \mu}} \quad (9)$$

### 3. Resultados

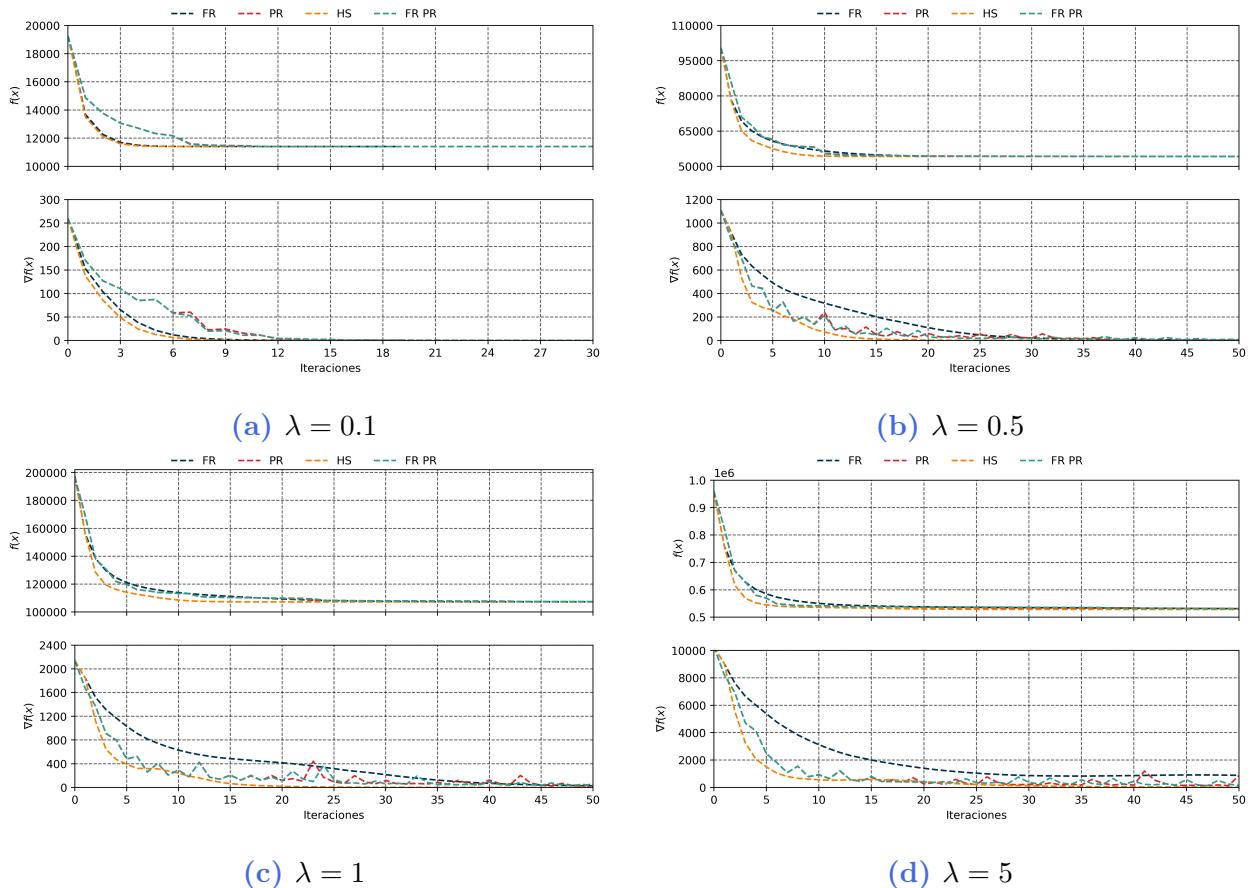
Se implementaron los algoritmos de gradiente conjugado con los cálculos de  $\beta$  para FR, PR, HS y FR-PR. En todos los casos se partió desde un punto inicial aleatorio dado por la ecuación

$$X_0 = \{x_{i,j} \in U(0, 1) \text{ para } 1 < i < n \ 1 < j < m\}$$

Se usaron los valores de 0.1, 0.5, 1 y 5 para la ecuación 8.

#### 3.1. Evolución temporal de la solución

Realizando comparaciones de los valores de la función y la norma del gradiente en cada iteración se obtuvieron las figuras 1.



**Figura 1:** Valor de la función y norma del gradiente en cada iteración para el método de gradiente conjugado con pasos FR, PR, HS y FR-PR para  $\lambda = \{0.1, 0.5, 1, 2\}$

#### 3.2. Suavizado de imágenes

En la figura 2 se visualizan los resultados al introducir una imagen en el parámetro  $g$  de la ecuación 8 para los métodos FR, PR, HS y FR-PR.



**Figura 2:** Resultados de las imágenes para cada método y  $\lambda$ .

## 4. Conclusiones

A partir de la figura 1 se obtienen las siguientes conclusiones:

- Conforme aumenta el valor del parámetro  $\lambda$  el número de iteraciones crece, por ende tarda más tiempo en llegar al mínimo global.
- El mejor método para esta imagen es usando el tamaño de paso de Hestenes-Stiefel, debido a que llega más rápido al mínimo local.
- El peor método para esta imagen es el tamaño de paso de Fletcher-Reeves debido a que tarda más en llegar al punto óptimo.
- El método con tamaño de paso Polak-Ribiere y el tamaño de paso Fletcher-Reeves con

Polak-Ribiere tiene una dinámica muy semejante. Esto es debido a que se toma más seguido el paso PR que el FR en FR-PR.

A partir de la figura 2 se obtiene que independientemente del tamaño de paso, se obtiene una imagen muy semejante para cada valor de  $\lambda$ . Sin embargo, conforme aumenta el valor de  $\lambda$  la imagen se ve cada vez más difuminada.

## 5. Referencias

- [1] Hestenes MR, Stiefel E. Methods of conjugate gradients for solving linear systems. J RES NATL BUR STAN. 1952 dec;49(6):409. Available from: <https://doi.org/10.6028%2Fjres.049.044>.
- [2] Reid JK. The Use of Conjugate Gradients for Systems of Linear Equations Possessing "Property A". SIAM Journal on Numerical Analysis. 1972;9(2):325–332. Available from: <http://www.jstor.org/stable/2156406>.
- [3] Concus P, Golub GH, O'Leary DP. A generalized conjugate gradient method for the numerical solution for elliptic partial differential equations. In: Sparse Matrix Computations. Elsevier; 1976. p. 309–332. Available from: <https://doi.org/10.1016%2Fb978-0-12-141050-6.50023-4>.
- [4] Fletcher R, Reeves CM. Function minimization by conjugate gradients. The computer journal. 1964;7(2):149–154. Available from: <https://doi.org/10.1093/comjnl/7.2.149>.
- [5] Polak E, Ribiere G. Note sur la convergence de méthodes de directions conjuguées. ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique. 1969;3(R1):35–43. Available from: [http://www.numdam.org/item/M2AN\\_1969\\_3\\_1\\_35\\_0/](http://www.numdam.org/item/M2AN_1969_3_1_35_0/).
- [6] Hestenes MR. Conjugate direction methods in optimization. In: Optimization Techniques Part 1. Springer; 1978. p. 8–27. Available from: <https://doi.org/10.1007/978-1-4612-6048-6>.
- [7] Babaie-Kafaki S, Ghanbari R. A hybridization of the Polak-Ribièr-Polyak and Fletcher-Reeves conjugate gradient methods. Numer Algor. 2014 may;68(3):481–495. Available from: <https://doi.org/10.1007%2Fs11075-014-9856-6>.