Optimización

Método de Gradiente Conjugado, Gradiente de Barzilai-Borwein y Máximo Descenso

Tarea 9

28 de abril de 2021

Erika Rivadeneira Pérez
erika.rivadeneira@cimat.mx
Matemáticas Aplicadas - CIMAT

1. Resumen

En el presente reporte se comparan los promedios de iteraciones, de tiempo y de la norma del gradiente en 30 corridas al utilizar los métodos de gradiente conjugado, gradiente de Barzilai-Borwein y de máximo descenso. Como resultados se presentan las gráficas de las primeras y últimas iteraciones de las normas de los gradientes y se concluye que el método de gradiente conjugado es el que realiza menos iteraciones en un menor tiempo.

2. Introducción

El método del gradiente conjugado debido a Fletcher y Reeves (1964) combina las características de la convergencia cuadrática del método de las direcciones conjugadas con las del método del gradiente. El método supone una importante mejora del método del gradiente con sólo un pequeño incremento en el esfuerzo de cálculo. El método del gradiente conjugado, esencialmente, combina la información obtenida del vector gradiente con la información acerca del vector gradiente de iteraciones previas. Lo que hace el método es calcular la nueva dirección de búsqueda utilizando una combinación lineal del gradiente en la etapa considerada y el de la etapa anterior. La principal ventaja del método es que necesita almacenar muy poca cantidad de información con lo que puede ser programado fácilmente incluso en calculadoras.
Además se consideran los métodos de Barzilai-Borwein y máximo descenso con tamaño de paso exacto (el cual ha sido discutido en reportes pasados) para la búsqueda del mínimo de una función cuadrática.
En la siguiente sección se mencionan los algoritmos a seguir para implementar cada método y así obtener los resultados deseados.

3. Metodología

Se minimizará la siguiente función cuadrática

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} \tag{1}$$

Donde

$$\mathbf{Q} = \mathbf{P}\mathbf{D}\mathbf{P}^T$$

$$\mathbf{P} = \prod_{j=1}^{m}\mathbf{H}_j$$

$$\mathbf{H}_j = \mathbf{I} - 2\frac{\boldsymbol{u}_j\boldsymbol{u}_j^T}{\boldsymbol{u}_j^T\boldsymbol{u}_j}$$

I es la matriz identidad, $\mathbf{H}_j$ son matrices Householder ortogonales, $\boldsymbol{u}_j$ son vectores con entradas generadas aleatoriamente en $(-1,1)$.$\mathbf{D}$ es una matriz diagonal cuya $i-$ésima componente está definida por

$$\log d_i = \left(\frac{i-1}{n-1}\right) n \text{ cond } , i = 1, 2, \cdots , n$$

Notemos que el parámetro *ncond* está relacionado con el número de condición de Q. b es generada a partir de x^* con componentes aleatorios en $(-1, 1)$, entonces

$$\boldsymbol{b} = \mathrm{Q}\boldsymbol{x}^*$$

La minimización de esta función se la realiza mediante el uso de los siguientes métodos.

3.1. Método Gradiente conjugado

A continuación se presenta el algoritmo a seguir para la posterior implementación del método.

Gradiente conjugado

Input: x_0
1: Hacer $\boldsymbol{g}_0 = \mathbf{Q}\boldsymbol{x}_0 - \boldsymbol{b}, \boldsymbol{d}_0 = -\boldsymbol{g}_0, k = 0$
2: **while** $||g_k|| \neq 0$ no converja **do**
3: $\quad \alpha_k = -\frac{\boldsymbol{g}_k^T \boldsymbol{d}_k}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}$
4: $\quad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$
5: $\quad \boldsymbol{g}_{k+1} = \mathbf{Q}\boldsymbol{x}_{k+1} - \boldsymbol{b} = \nabla f(\boldsymbol{x}_{k+1})$
6: $\quad \beta_{k+1} = \frac{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{g}_{k+1}}{\boldsymbol{d}_k^T \mathbf{Q} \boldsymbol{d}_k}$
7: $\quad \boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_{k+1} \boldsymbol{d}_k$
8: $\quad k = k + 1$
9: **end while**
Output: x^*

Para la implementación del método se consideraron como criterios de paro una tolerancia de 1×10^{-4} y un número máximo de iteraciones de 10000.

3.2. Método del Gradiente de Barzilai-Borwein

Para la implementación de este método se consideró el siguiente algoritmo

Gradiente de Barzilai-Borwein

Input: x_0, τ_g
1: Hacer $\boldsymbol{y}_0 = \boldsymbol{x}_0, k = 0$
2: **while** $||g_k|| > \tau_g$ **do**
3: $\quad \alpha_k = \frac{\boldsymbol{s}_{k-1}^T \boldsymbol{y}_{k-1}}{\boldsymbol{y}_{k-1}^T \boldsymbol{y}_{k-1}}$
4: $\quad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k$
5: $\quad k = k + 1$
6: **end while**
Output: x^*

donde para $k > 0$ se tiene que $\boldsymbol{s}_{k-1} = \boldsymbol{x}_k - \boldsymbol{x}_{k-1}$ y $\boldsymbol{y}_{k-1} = \boldsymbol{g}_k - \boldsymbol{g}_{k-1}$. Para el caso $k = 0$ tomamos el paso exacto $\alpha = 0{,}01$. Como criterios de paro se consideró una tolerancia de 1×10^{-4} y un número máximo de iteraciones de 10000.

4. Resultados

Se realizaron 30 corridas de los métodos del Gradiente conjugado, gradiente de Barzilai-Borwein y de máximo descenso en la función (1). Notemos que el gradiente de (1) es $\nabla f(x) = Qx - b$. Considerando $m = 3, n = 5000$, $ncond \in \{2, 4, 6\}$ y $x_0 = [1, 1, ..., 1]$ se obtuvieron los siguientes resultados

- Para número de condición $ncond = 2$

	GC	**G B-B**	**Steepest Desc.**
Prom. Iteraciones	19.63	26.67	9510.5
Prom. Tiempo	1.82	6.75	135.829

Tabla 1: Promedio de iteraciones y de tiempo en segundos de 30 corridas para $ncond = 2$

- Para número de condición $ncond = 4$

	GC	**G B-B**	**Steepest Desc.**
Prom. Iteraciones	56.63	72.67	9188.23
Prom. Tiempo	2.02	6.33	124.33

Tabla 2: Promedio de iteraciones y de tiempo en segundos de 30 corridas para $ncond = 4$

- Para número de condición $ncond = 6$

	GC	**G B-B**	**Steepest Desc.**
Prom. Iteraciones	157.97	214.6	8956.97
Prom. Tiempo	4.43	3.95	119.67

Tabla 3: Promedio de iteraciones y de tiempo en segundos de 30 corridas para $ncond = 6$

Notemos que en las tablas mostradas previamente las iniciales **GC** corresponden al método del *Gradiente Conjugado*, **G B-B** al gradiente de Barzilai-Borwein y **Steepest Desc** al método del máximo descenso.

Por otro lado, se graficó el promedio de las normas del gradiente de las tres primeras y últimas iteraciones de cada método para cada número de condición. Las cuales se muestran a continuación

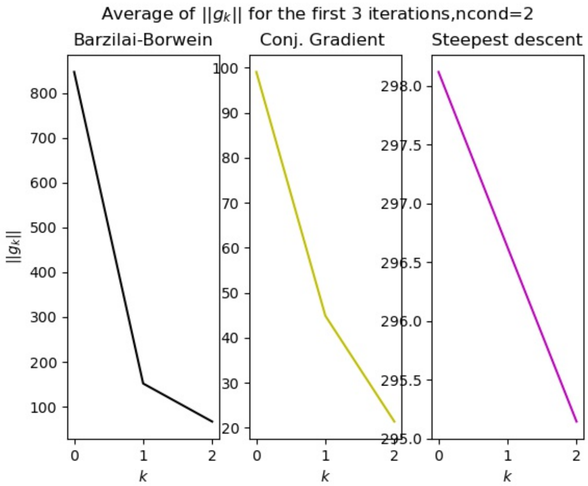

Figura 1: Promedio de $||g_k||$ para las primeras 3 iteraciones con cada método y número de condición 2

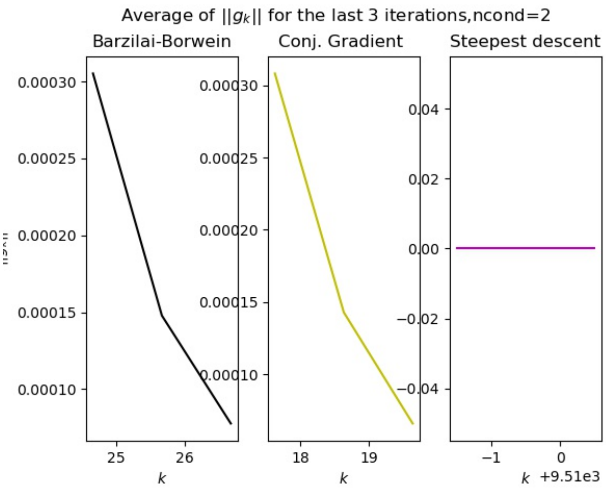

Figura 2: Promedio de $||g_k||$ para las últimas 3 iteraciones con cada método y número de condición 2

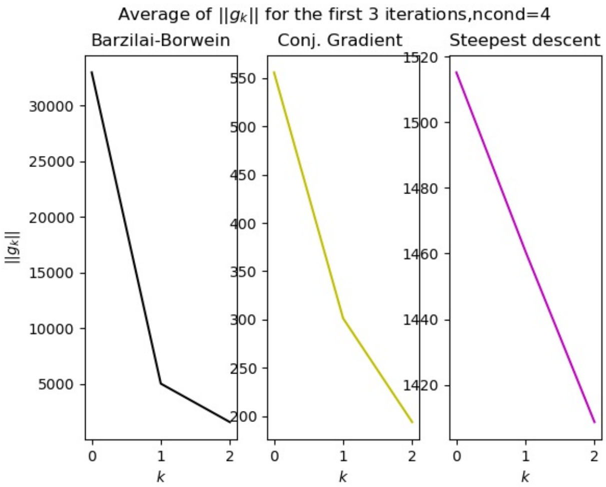

Figura 3: Promedio de $||g_k||$ para las primeras 3 iteraciones con cada método y número de condición 4

5. Conclusiones y Discusión

Es importante mencionar que se consideró un $n = 5000$ debido a que la memoria ram de la computadora en la que se realizó la ejecución del programa no toleró una longitud de $n = 10000$. A pesar de este percanse se obtuvieron los resultados deseados.
Por otro lado, las tablas de resultados mostradas en la sección anterior nos muestran que para cada número de condición $ncond \in \{2, 4, 6\}$ se observa que el método de gradiente conjugado tiene un menor promedio de iteraciones y de tiempo de ejecución en relación al método del gradiente de Barzilai-Borwein y de máximo descenso. El método que más tardó en converger fue el método de máximo descenso, es necesario mencionar que el número máximo de iteraciones aceptado para este método fue de 15000 y un error aceptado de 1×10^{-4}. Lo cual implica que el tiempo de ejecución sea mayor.
Además, las gráficas mostradas en la sección de resultados muestran que efectivamente la norma del gradiente es decrece para las 3 primeras y 3 últimas iteraciones en cada método considerado. Pero, en las gráficas correspondientes a las 3 últimas iteaciones del método de máximo descenso [Figs. (2), (4) y (6)] muestran que la norma del gradiente se mantiene cercano a 0 en estas iteraciones.

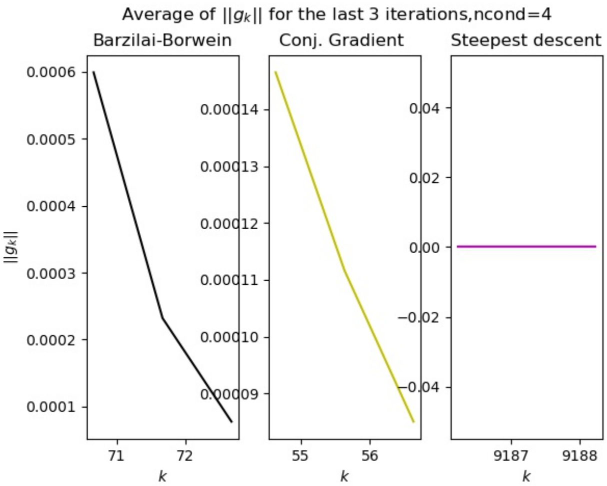

Figura 4: Promedio de $||g_k||$ para las últimas 3 iteraciones con cada método y número de condición 4

Referencias

[1] J. Nocedal and S. J. Wright. Numerical Optimization. Springer Series in Operation Research, 2000.

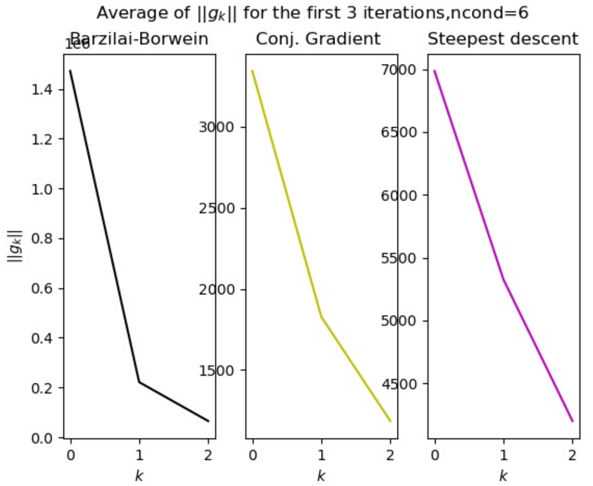

Figura 5: Promedio de $||g_k||$ para las primeras 3 iteraciones con cada método y número de condición 6

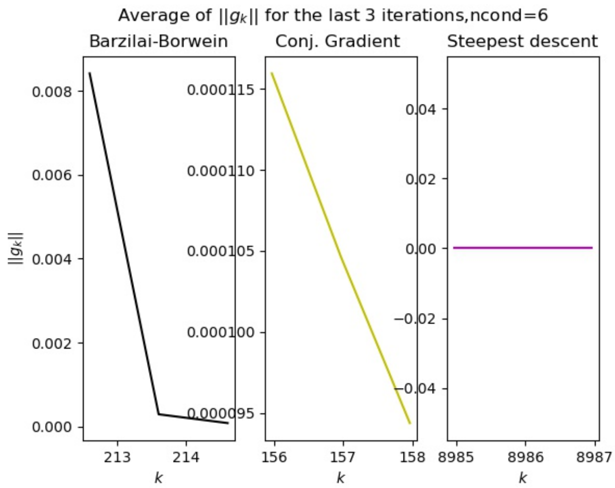

Figura 6: Promedio de $||g_k||$ para las últimas 3 iteraciones con cada método y número de condición 6