Tarea 9 - Optimización 1

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

***Resumen*—En el presente trabajo se presentarán los algoritmos cuasi Newton DFP y BFGS, se compararán sus resultados en cuanto a tiempo promedio de ejecución, número promedio de iteraciones por corrida y la norma promedio del gradiente para cada una de las funciones.**

***Index Terms*—Métodos Cuasi-Newton, DFP, BFGS**

I. INTRODUCCIÓN

El método clásico de Newton tiene la problemática del cálculo del hessiano $\nabla^2 f$, el cual puede ser muy costoso computacionalmente. Además se requiere el cálculo de la inversa del hessiano y puede que la dirección $\mathbf{d}_k$ que encontremos no sea de descenso.

Una alternativa a este problema son los métodos cuasi-newton, los cuales construyen un modelo que se basa en medir los cambios del gradiente, solo se ocupa el cálculo del gradiente similar al algoritmo de máximo descenso. En lugar de calcular el Hessiano en cada iteración se propone un método que permite calcular una secuencia $\mathbf{B}_k$ usando la curvatura medida en el paso actual.

Los dos algoritmos a considerar en el presente trabajo son los algoritmos DFP (Davidson-Fletcher-Powell) y BFGS (Broyden-Fletcher-Goldfarb-Shannon)

En el algoritmo DFP se generan las aproximaciones de la inversa del Hessiano de mediante la siguiente fórmula de actualización

$$\mathbf{H}_{k+1} = \mathbf{H}_k - \frac{\mathbf{H}_k \mathbf{y}_k \mathbf{y}_k^t \mathbf{H}_k}{\mathbf{y}_k^t \mathbf{H}_k \mathbf{y}_k} + \frac{\mathbf{s}_k \mathbf{s}_k^t}{\mathbf{y}_k^t \mathbf{a}_k} \tag{1}$$

Mientras que en el algoritmo BFGS la fórmula de actualización para la inversa del Hessiano es de la forma

$$\mathbf{H}_{k+1} = \left(\mathbf{I} - \rho_k \mathbf{s}_k \mathbf{y}_k^t\right) \mathbf{H}_k \left(\mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k^t\right) \tag{2}$$

Para $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$ y $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$.

II. MÉTODO/ALGORITMO

Los algoritmos utilizados en esta tarea tienen el mismo esquema general, dado $\mathbf{x}_0$ un vector inicial y $\mathbf{H}_0$ una matriz definida positiva y simétrica (para esta tarea se toma la matriz identidad) se busca calcular $\mathbf{x}^*$.

El esquema general de estos dos algoritmos cuasi Newton es el siguiente

Input: $\mathbf{x}_0$ $\mathbf{H}_0$
Output: $\mathbf{x}^*$
$k = 0$
while $\|g_k\| \neq 0$
    $\mathbf{d}_k = -\mathbf{H}_k \mathbf{g}_k$
    Calcular α_k usando búsqueda en linea
    $\mathbf{x}_k = \mathbf{x}_k + \alpha_k \mathbf{d}_k$
    Actualizar $\mathbf{g}_{k+1}, \mathbf{y}_k, \mathbf{s}_k$
Actualizar $\mathbf{H}_k$ usando (1) para DFP y (2) para BFGS
    k=k+1
end while

Ambos algoritmos se probarán con la función Rosenbrock para $n = 100$ y la función Wood considerando treinta corridas en donde en cada una de ellas se genera un punto aleatorio.

III. ANÁLISIS DE RESULTADOS

Se generaron cuatro tablas, una por algoritmo probado con las dos diferentes funciones, en ellas se puede observar el número de iteraciones promedio que hizo cada algoritmo, la norma promedio del gradiente así como el tiempo promedio de ejecución de cada algoritmo, dichas tablas se presentan en el Apéndice A

IV. CONCLUSIÓN

Con base en lo obtenido al realizar los experimentos en este trabajo se concluye que ambos algoritmos tienen un desempeño eficaz ya que encuentran el óptimo para cada punto aleatorio inicial que se les proporcione sin embargo no muestran una eficiencia superior al método de máximo descenso como se había mencionado en las diapositivas.

En este trabajo notamos que el método DFP tuvo un rendimiento superior al BFGS ya que en promedio el número de iteraciones requerido para encontrar el óptimo con una tolerancia de 1×10^{-4} fue dos veces más en el algoritmo BFGS que en el DFP.

APÉNDICES

Apéndice A

Función Rosenbrock: Para el algoritmo DFP se obtuvieron los siguientes resultados

DFP		
Norma p	Tiempo p	Iteraciones p
5.20510097e-05	2.97823033	3214

Para el algoritmo BFGS se obtuvieron los siguientes resultados

DFP		
Norma p	Tiempo p	Iteraciones p
4.23259630e-01	1.87405304e+01	6.22400000e+03

Función Wood: Para el algoritmo DFP se obtuvieron los siguientes resultados

DFP		
Norma p	Tiempo p	Iteraciones p
5.89442622e-05	1.53116333e-02	5.80000000e+01

Para el algoritmo BFGS se obtuvieron los siguientes resultados

DFP		
Norma p	Tiempo p	Iteraciones p
8.30001995e-05	1.26689867e-01	1.62000000e+02

Apéndice B

Problema 1

Las matrices de corrección de rango 1 se escriben como

$$\mathbf{B}_{k+1}^{RS1} = \mathbf{B}_k + \frac{(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)^t}{(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)^t\mathbf{s}_k}$$
$$\mathbf{H}_{k+1}^{RS1} = \mathbf{H}_k + \frac{(\mathbf{s}_k - \mathbf{H}_k\mathbf{y}_k)(\mathbf{s}_k - \mathbf{H}_k\mathbf{y}_k)^t}{\mathbf{y}_k^t(\mathbf{s}_k - \mathbf{H}_k\mathbf{y}_k)}$$

donde $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$ y $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$.

- Derive la matriz $\mathbf{H}_{k+1}^{RS1}$ a partir de la matriz $\mathbf{B}_{k+1}^{RS1}$ usando la fórmula de Sherman-Morrison
- Si $\mathbf{H}_k$ es una matriz definida positiva y $\mathbf{y}_k^t(\mathbf{s}_k - \mathbf{H}_k\mathbf{y}_k) > 0$ demuestre que $\mathbf{H}_{k+1}$ es una matriz definida positiva

Solución:

Consideremos primero la ecuación de Sherman-Morrison dada de la siguiente forma

$$(\mathbf{A} + \mathbf{UV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{I} + \mathbf{VA}^{-1}\mathbf{U}\right)^{-1}\mathbf{VA}^{-1}$$

De la expresión de la matriz $\mathbf{B}_{k+1}^{RS1}$ consideremos $\mathbf{U} = a(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)$ y $\mathbf{V} = \mathbf{U}^t$ donde $a^2 = \frac{1}{(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)^t\mathbf{s}_k}$ entonces vemos que $\mathbf{B}_{k+1}^{RS1} = \mathbf{B}_k + \mathbf{UV}$. Aplicando la ecuación de Sherman-Morrison notamos que

$$\begin{aligned}\left(\mathbf{B}_{k+1}^{RS1}\right)^{-1} &= (\mathbf{B}_k + \mathbf{UV})^{-1}\\ &= \mathbf{B}_k^{-1} - \mathbf{B}_k^{-1}\mathbf{U}\left(\mathbf{I} + \mathbf{VB}_k^{-1}\mathbf{U}\right)^{-1}\mathbf{VB}_k^{-1}\end{aligned}$$

Sabemos que $\mathbf{B}_k = \mathbf{H}_K$, luego tenemos lo siguiente

$$\begin{aligned}\mathbf{I} + \mathbf{VB}_k^{-1}\mathbf{U} &= \mathbf{I} + \mathbf{VH}_k^{-1}\mathbf{U}\\ &= 1 + a^2\left[(\mathbf{y}_k - \mathbf{s}_k\mathbf{B}_k)^t(\mathbf{H}_k\mathbf{y}_k - \mathbf{s}_k)\right]\\ &= 1 + \frac{(\mathbf{y}_k - \mathbf{s}_k\mathbf{B}_k)^t(\mathbf{H}_k\mathbf{y}_k - \mathbf{s}_k)}{(\mathbf{y}_k - \mathbf{s}_k\mathbf{B}_k)^t\mathbf{s}_k}\\ &= \frac{(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)^t(\mathbf{H}_k\mathbf{y}_k)}{(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)^t\mathbf{s}_k}\\ &= \frac{\mathbf{y}_k^t(\mathbf{H}_k\mathbf{y}_k - \mathbf{s}_k)}{(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)^t\mathbf{s}_k}\end{aligned}$$

Por lo tanto $(\mathbf{I} + \mathbf{VB}_k^{-1}\mathbf{U})^{-1} = \frac{(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)^t\mathbf{s}_k}{\mathbf{y}_k^t(\mathbf{H}_k\mathbf{y}_k - \mathbf{s}_k)}$. Como esta cantidad es un escalar, entonces tenemos la relación

$$\mathbf{B}_k^{-1}\mathbf{U}\left(\mathbf{I} + \mathbf{VB}_k^{-1}\mathbf{U}\right)^{-1}\mathbf{VB}_k^{-1} = \left(\mathbf{I} + \mathbf{VB}_k^{-1}\mathbf{U}\right)^{-1}\mathbf{B}_k^{-1}\mathbf{UVB}_k^{-1}$$

Por lo que al calcular $\mathbf{B}_k^{-1}\mathbf{UVB}_k^{-1}$,tenemos

$$\begin{aligned}\mathbf{B}_k^{-1}\mathbf{UVB}_k^{-1} &= \mathbf{H}_k\mathbf{UVH}_k\\ &= a^2\left[(\mathbf{H}_k\mathbf{y}_k - \mathbf{s}_k)(\mathbf{H}_k\mathbf{y}_k - \mathbf{s}_k)^t\right]\\ &= \frac{(\mathbf{H}_k\mathbf{y}_k - \mathbf{s}_k)(\mathbf{H}_k\mathbf{y}_k - \mathbf{s}_k)^t}{(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)^t\mathbf{s}_k}\end{aligned}$$

Por lo que al sustituir las expresiones encontradas se tiene lo siguiente

$$\begin{aligned}\left(\mathbf{B}_{k+1}^{RS1}\right)^{-1} &= \mathbf{H}_k - \frac{(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)^t\mathbf{s}_k}{\mathbf{y}_k^t(\mathbf{H}_k\mathbf{y}_k - \mathbf{s}_k)}\frac{(\mathbf{H}_k\mathbf{y}_k - \mathbf{s}_k)(\mathbf{H}_k\mathbf{y}_k - \mathbf{s}_k)^t}{(\mathbf{y}_k - \mathbf{B}_k\mathbf{s}_k)^t\mathbf{s}_k}\\ &= \mathbf{H}_k - \frac{(\mathbf{H}_k\mathbf{y}_k - \mathbf{s}_k)(\mathbf{H}_k\mathbf{y}_k - \mathbf{s}_k)^t}{\mathbf{y}_k^t(\mathbf{H}_k\mathbf{y}_k - \mathbf{s}_k)}\\ &= \mathbf{H}_k + \frac{(-\mathbf{H}_k\mathbf{y}_k + \mathbf{s}_k)(-\mathbf{H}_k\mathbf{y}_k + \mathbf{s}_k)^t}{\mathbf{y}_k^t(-\mathbf{H}_k\mathbf{y}_k + \mathbf{s}_k)}\\ &= \mathbf{H}_{k+1}^{RS1}\end{aligned}$$

Multiplicando numerador y denominador por -1 y finalmente multiplicar toda esa fracción por -1. De esta forma se deriva la fórmula para $\mathbf{H}_{k+1}^{RS1}$ a partir de $\mathbf{B}_{k+1}^{RS1}$.

Demostración:

Procedemos por definición de matriz definida positiva, consideremos un vector arbitrario $\mathbf{x} \neq 0$ entonces consideremos la multiplicación

$$
\begin{aligned}
\mathbf{x}^t\mathbf{H}_{k+1}\mathbf{x} &= \mathbf{x}^t\left(\mathbf{H}_k + \frac{(-\mathbf{H}_k\mathbf{y}_k + \mathbf{s}_k)(-\mathbf{H}_k\mathbf{y}_k + \mathbf{s}_k)^t}{\mathbf{y}_k^t(-\mathbf{H}_k\mathbf{y}_k + \mathbf{s}_k)}\right)\mathbf{x} \\
&= \mathbf{x}^t\mathbf{H}_k\mathbf{x} + \frac{\mathbf{x}^t(-\mathbf{H}_k\mathbf{y}_k + \mathbf{s}_k)(-\mathbf{H}_k\mathbf{y}_k + \mathbf{s}_k)^t\mathbf{x}}{\mathbf{y}_k^t(-\mathbf{H}_k\mathbf{y}_k + \mathbf{s}_k)} \\
&= \mathbf{x}^t\mathbf{H}_k\mathbf{x} + \frac{(-\mathbf{x}^t\mathbf{H}_k\mathbf{y}_k + \mathbf{x}^t\mathbf{s}_k)(-\mathbf{x}\mathbf{H}_k\mathbf{y}_k + \mathbf{x}\mathbf{s}_k)^t}{\mathbf{y}_k^t(-\mathbf{H}_k\mathbf{y}_k + \mathbf{s}_k)} \\
&= \mathbf{x}^t\mathbf{H}_k\mathbf{x} + \frac{(-\mathbf{x}^t\mathbf{H}_k\mathbf{y}_k + \mathbf{x}^t\mathbf{s}_k)^2}{\mathbf{y}_k^t(-\mathbf{H}_k\mathbf{y}_k + \mathbf{s}_k)} \\
&> 0
\end{aligned}
$$

Ya que por hipótesis se tiene que $\mathbf{H}_k$ es definida positiva y $\mathbf{y}_k^t(-\mathbf{H}_k\mathbf{y}_k + \mathbf{s}_k) > 0$ de modo que $\mathbf{H}_{k+1}$ es definida positiva que es lo que se buscaba probar $\blacksquare$.

REFERENCIAS

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