Tarea 4-Optimización 1

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

***Resumen*—El propósito del presente trabajo es implementar los métodos de Barzilai Borwein, Zhang Hager e interpolación cúbica para minimizar las funciones objetivo de Rosenbrock para $n = 100$ y la función Wood descritas en la tarea 2, así mismo se usarán estos algoritmos para maximizar la función de log verosimilitud para la clasificación de dígitos.**

***Index Terms*—Barzilai Borwein, Zhang Hager, Interpolación cuadrática, Interpolación cúbica.**

I. INTRODUCCIÓN

A lo largo del curso se han podido comparar diversos métodos de búsqueda en linea, desde el método de gradiente descendente hasta el método de Newton, una mejora a encontrar el tamaño de paso es utilizar el algoritmo de interpolación cuadrática e intepolación cúbica las cuales dados valores α_1, α_0 que no satisfacen las condiciones de Armijo se regresa una tamaño de paso que garantiza un apropiado descenso, por otra parte Barzilai Borwein representa una alternativa más rápida al método de máximo descenso mientras que Zhang Hager, hace una modificación del método de backtracking para poder encontrar un tamaño de paso que satisfaga una una nueva forma de la condición de Armijo. En este trabajo se implementarán estos algoritmos y se compararán sus resultados en número de iteraciones, norma del gradiente en el punto óptimo y valor de la función en el punto encontrado

NOTA: La parte de clasificación de dígitos no se pudo realizar, ya que el código adjunto presentaba error en el dominio de la función log, si embargo no se encontró explicación aparente ya que las cuentas del gradiente son correctas para la estimación de parámetros, así mismo su implementación en Python corresponde a los resultados obtenidos en las cuentas.

II. MÉTODO/ALGORITMO

II-A. *Interpolación cúbica*

El método de interpolación cúbica calcula el tamaño de paso óptimo basada en un valor inicial α_0 y α_1 que no satisfacen la condición de armijo, luego encuentra un punto α_2 que satisface la condición $\phi(\alpha_2) \leq \phi^{'}(0) + c2 * \alpha_2\phi(0)$. Luego se aplica este tamaño de paso al método de Newton de búsqueda en linea implementado en la tarea anterior.

II-B. *Método de Barzilai Borwein*

Para el método de Barzilai Borwein se siguió el siguiente algoritmo.

Input: x_0 $0 < \epsilon << 1$
Output: x^*
$k = 0$,
while $\|\nabla f(x_k)\| \neq 0$ **do**
  if $k == 0$
    $\alpha_k = argmin_{\alpha>0} f(x_k + \alpha_k g_k)$ usando búsqueda en linea
  else
    $\alpha_k = BB(\alpha_k)$
  $x_{k+1} = x_k - \alpha_k g_k$
  k=k+1
end while

II-C. *Método de Zhang Hager*

Para el método de Zhang Hager se siguió el siguiente algoritmo.

Input: x_0 $0 < \epsilon << 1$, c_1 $\nu_1 \in (0, 1)$, $C_0 = f(\mathbf{x}_0)$ $Q_0 = 1$
Output: x^*
$k = 0$,
while $\|\nabla f(x_k)\| \neq 0$ **do**
  Calcular α_k tal que $f(x_k+\alpha_k d_k) \leq C_k+\alpha_k c_1 d_k$
  $x_{k+1} = x_k + \alpha_k d_k$
  $Q_{k+1} =_k +1$
  $C_{k+1} = \frac{(\nu Q_k C_k + f(x_{k+1}))}{Q_{k+1}}$
end while

III. ANÁLISIS DE RESULTADOS

Se generaron las tablas contenidas en el apéndice A, donde se presentan los resultados de los resultados de los métodos a probar, comparando el número de iteraciones, el gradiente evaluado en el punto encontrado y el valor del error entre la función evaluada en el óptimo y la función evaluada en el punto obtenido.

IV. CONCLUSIONES

Con base en los resultados obtenidos al realizar el presente trabajo se puede concluir que el método de Barzilai Borwein fue el método que más rápido converge al óptimo de las funciones, a probar, el método de Zhang Hager fue el método que más iteraciones se llevó y el que más lento convergió, el método de interpolación cúbica es muy sensible a condiciones iniciales sin embargo converge en un número pequeño de iteraciones.

Respecto al problema 3, los datos arrojaban math domain error al momento de evaluarlo en la función gradiente, aún restándoles la media y dividiéndolos entre su desviación estándar, se adjunta código del problema 3.

APÉNDICES

Apéndice A

Tablas donde se muestran los resultados de los métodos de Interpolación cúbica, Barzilai Borwein y Zhang Hager

Caso 1: Función Rosenbruck para n=100

Punto inicial $[1{,}2, 0, 0, 0, ..., -1{,}2, 0]$

Interpolación cúbica:

Iteraciones	$\|\nabla f(x_k)\|$	$\|f(x_k - f(x^*)\|$
2.3000e+01	1.90124877e-05	5.39760782e-11

Barzilai Borwein:

Iteraciones	$\|\nabla f(x_k)\|$	$\|f(x_k - f(x^*)\|$
8.85000000e+02	9.87641780e-05	7.30916061e-09

Zhang Hager:

Iteraciones	$\|\nabla f(x_k)\|$	$\|f(x_k - f(x^*)\|$
7.19750000e+04	9.99927783e-05	1.00220659e-08

Caso 2: Función Wood

Punto inicial $[-3, -1, -3, -1]$

Interpolación cúbica:

Iteraciones	$\|\nabla f(x_k)\|$	$\|f(x_k - f(x^*)\|$
2.00000000e+03	4.97870664e-03	6.98602452e-06

Barzilai Borwein:

Iteraciones	$\|\nabla f(x_k)\|$	$\|f(x_k - f(x^*)\|$
1.01000000e+02	2.08468685e-05	7.23780008e-13

Zhang Hager:

Iteraciones	$\|\nabla f(x_k)\|$	$\|f(x_k - f(x^*)\|$
2.00000000e+04	4.23805020e-03	1.24803929e-05

Apéndice B

Problema 1

Problema 1. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = (x-a)^4$, where a is a constant. Suppose that we apply Newtons Method to the problem of minimizing f.

1. Write down the update equation for Nwetons method applied to the problem.
2. Let $y_k = |x_k - a|$ where x_k is the k-th iterate in Newtons method.
   Show that the sequence $\{y_k\}$ satisfies $y_{k+1} = \frac{2}{3} y_k$
3. Show that $x_k \to a$ for any initial guess x_0
4. Which is the order of convergence of the sequence x_k
5. Does the previous convergence order contradict the theorem about the convergence order of Newtons method? Explain

Solución

1. Calculamos primero la primera y la segunda derivada de la función f, las cuales están dadas de la siguiente forma $f'(x) = 4(x-a)^3$ y $f^2(x) = 12(x-a)^2$, de la ecuación de Newton se tiene que la dirección de descenso d_k es de la forma $12(x-a)^2 d_k = 4(x-a)^3$ de modo que $d_k = \frac{-1}{3}(x-a)$ por lo que la dirección de Newton está dada por

$$x_{k+1} = x_k - \frac{1}{3}(x_k - a)$$

   la cual es la ecuación de Newton para tamaño de paso 1.
2. Dem:

   Consideremos lo siguiente

$$\begin{aligned} y_{k+1} &= |x_{k+1} - a| \\ &= |x_k - \frac{1}{3}(x_k - a) - a| \text{ Por (1)} \\ &= |\frac{2}{3}(x_k - a)| \\ &= \frac{2}{3}|x_k - a| \\ &= \frac{2}{3} y_k \end{aligned}$$

   ■
3. Dem:

Para demostrar que $x_0 \to a$ para cualquier punto inicial x_0 implica demostrar que $y_k \to 0$ para cualquier $y_0 = |x_0 - a|$. Podemos notar entonces lo siguiente.

$$\begin{aligned}
\lim_{k\to\infty} y_k &= \lim_{k\to\infty} \frac{2^k}{3^k} y_0 \\
&= y_0 \lim_{k\to\infty} \frac{2^k}{3^k} \\
&= 0
\end{aligned}$$

Lo que implica que $x_0 \to a$ para cualquier punto inicial x_0■.

4. Notemos que $y_k = \frac{2^k}{3^k} y_0$ por lo que $y_k \leq y_0$ por lo que $y_k = O(|x_0 - a|)$, lo que implica que el orden de convergencia es lineal, es decir es de orden 1.
5. Lo anterior no contradice el método de Newton para búsqueda en linea ya que de acuerdo al teorema 2.1, para que el orden de convergencia de Newton sea de orden al menos 2 se ocupa que el hessiano evaluado en el ótimo sea positivo definido en este caso $f^2(a) = 0$ por lo que esa condición no se cumple.

REFERENCIAS

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