

Un método para el cálculo de derivadas segundas

Hugo D. Scolnik
e-mail: scolnik@fd.com.ar

M. Juliana Gambini
e-mail: jgambini@dc.uba.ar

Dto. de Computación FCEN, UBA.

September 19, 2000

Abstract

En este trabajo consideramos el problema de calcular aproximaciones para la derivada segunda de funciones de varias variables usando la técnica de diferencias finitas. Mostramos de donde surgen las diferentes fórmulas y calculamos los errores que se cometen con esas aproximaciones en función del incremento h , para derivadas primeras y segundas. Basándonos en estas observaciones describimos dos métodos, el de Gill y Murray y el de diferencias de gradientes. También introducimos un nuevo algoritmo que calcula derivadas segundas. Presentamos también la correspondiente comparación con los otros dos métodos. Finalmente incluimos algunas experiencias numéricas y aportamos conclusiones sobre los resultados obtenidos.

Palabras clave: algoritmo, diferencias finitas, diferenciación numérica.

1 Introducción

El objetivo de este trabajo es desarrollar un método para calcular derivadas segundas de una función dada, $f(x)$, $x \in R^n$ usando aproximaciones por diferencias finitas. Si las derivadas de f son muy difíciles de calcular o muy costosas de evaluar, los métodos para minimizar $f(x)$, quasi Newton o Newton, utilizan aproximaciones por diferencias finitas del gradiente y del hessiano de la función, y el éxito de encontrar el óptimo depende directamente de obtener buenas aproximaciones para las derivadas. Las fórmulas de diferencias finitas se deducen del desarrollo de Taylor de f , como veremos mas adelante, resultará

$$\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j} \cong \frac{f(x_0 + h_i e_i + h_j e_j) - f(x_0 + h_i e_i) - f(x_0 + h_j e_j) + f(x_0)}{h_i h_j}$$

donde lo más importante de considerar es la elección de los incrementos h_i y h_j . Si ellos son demasiado pequeños, se producen errores aritméticos en los cálculos que se traducen en aproximaciones no satisfactorias para las derivadas. La idea de este trabajo es encontrar un incremento que sea lo mejor posible, en el sentido de que minimice una función de error. Para esto se supone que h depende de ciertos parámetros, es decir $h = h(\alpha_1, \dots, \alpha_k)$. Se toman f_1, f_2, \dots, f_m funciones test, de las cuales se conocen los hessianos en determinados puntos, Hv_1, Hv_2, \dots, Hv_m son los hessianos verdaderos correspondientes, Hc_1, Hc_2, \dots, Hc_m son los hessianos calculados con el incremento h , y se toma una función objetivo del tipo :

$$F(\alpha_1, \dots, \alpha_k) = \sum_{i=1}^m \frac{\|Hv_i - Hc_i(\alpha_1, \dots, \alpha_k)\|_F}{\|Hv_i\|_F}$$

y luego

$$\min_{\alpha} F(\alpha)$$

para calcular este mínimo se usaron métodos de minimización sin derivadas.

2 Métodos clásicos para aproximar derivadas

2.1 Fórmulas para derivadas primeras

Dada $f : R^n \rightarrow R$, $x_0 \in R^n$, $h \in R$, podemos calcular numéricamente $\nabla f(x_0)$ usando solo valores de la función f . Llamaremos $g(x_0) = \nabla f(x_0)$, es decir que $g(x_0) = (g_1(x_0), \dots, g_n(x_0))$ con $g_i(x_0) = \frac{\partial f}{\partial x_i}(x_0)$

Fórmula de diferencias hacia adelante

$$g_i(x_0) \cong \frac{f(x_0 + h e_i) - f(x_0)}{h}$$

Esto se deduce de la siguiente manera: Por el desarrollo en serie de Taylor se obtiene,

$$f(x_0 + h e_i) = f(x_0) + h g^t(x_0) \cdot e_i + \frac{1}{2} h^2 e_i^t H(x_0) e_i + O(h^3)$$

luego

$$g^t(x_0) \cdot e_i = \frac{f(x_0 + h e_i) - f(x_0)}{h} - \frac{1}{2} h e_i^t H(x_0) e_i + O(h^2)$$

En esta aproximación se comete un error de truncamiento que proviene de despreciar los términos de mayor orden de la serie

$$E_t = \frac{1}{2} h e_i^t H(x_0) e_i = \frac{1}{2} h H_{ii}(x_0)$$

donde $H(x_0)$ es el hessiano de la función, evaluado en x_0 . Este no es el único error que se comete al tratar de aproximar una derivada, existe también el error aritmético que tiene varias contribuciones. Podría suponerse que cuanto más pequeño sea el incremento h , mejor será la aproximación para la derivada, pero esto no es cierto.

Para mostrarlo consideremos la tolerancia de la máquina con la que se trabaja, es el menor número positivo tol tal que $1 + tol \neq 1$ y multiplicando por $|x_0|$

$$|x_0| + |x_0| tol \neq |x_0|$$

por lo tanto debe ser $h \geq |x_0| tol$ pues si se tomara $h < |x_0| tol$ resultaría $f(x_0 + h) = f(x_0)$ en cuyo caso se obtendría siempre una derivada nula. Supongamos ahora que h cumple con la condición anterior y que $f(x_0)$ se calcula con un error relativo menor que E_f , así que si $\Psi(x_0)$ es el valor representado en punto flotante, de $f(x_0)$

$$\frac{|f(x_0) - \Psi(x_0)|}{|f(x_0)|} < E_f$$

entonces, si despreciamos otros errores que provienen de los cálculos

$$\left| \frac{f(x_0 + he_i) - f(x_0)}{h} - \frac{\Psi(x_0 + he_i) - \Psi(x_0)}{h} \right| < \frac{2E_f |f(x_0)|}{h}$$

y por lo tanto, si h es muy pequeño el error total puede aumentar considerablemente. Esto muestra que h demasiado cercano a 0 no es quizás la mejor elección, por ejemplo en este caso el h óptimo se encuentra minimizando la función

$$\Upsilon(h) = \frac{E_f |f(x_0)|}{h} + h |H_{ii}(x_0)|$$

de donde resulta

$$h \sim \sqrt{\frac{E_f |f(x_0)|}{|H_{ii}(x_0)|}}$$

entonces, en el mejor de los casos, el error relativo total es

$$ert \sim \frac{\sqrt{E_f} \sqrt{|f(x_0)| |H_{ii}(x_0)|}}{|g_i(x_0)|} \sim \sqrt{E_f}$$

si se asume que f, H_{ii}, g_i tienen aproximadamente el mismo orden. Además, observando $\Upsilon(h)$, vemos que si h es muy grande, el error de truncamiento de la serie será grande, y dominará el error total, en cambio si h es muy pequeño, el error de redondeo será el dominante. Luego, una buena elección del incremento será aquella en que los dos errores sean del mismo orden, que es, naturalmente, el h encontrado anteriormente, ya que si minimiza la función de error total debe compensar los errores de truncamiento y redondeo. Esta idea fue desarrollada en [8] por G.W. Stewart.

Fórmula de diferencias centrales Si es posible utilizar más evaluaciones de la función, es mejor usar la fórmula

$$g_i(x_0) \cong \frac{f(x_0 + he_i) - f(x_0 - he_i)}{2h}$$

Para deducirla se usa el desarrollo de Taylor de orden 3, obteniendo

$$f(x_0 + he_i) = f(x_0) + hg^t(x_0).e_i + \frac{1}{2}h^2 e_i^t H(x_0) e_i + \frac{h^3}{6} \frac{\partial^3 f(x_0)}{\partial^3 x_i} + O(h^4) \quad (1)$$

y

$$f(x_0 - he_i) = f(x_0) - hg^t(x_0).e_i + \frac{1}{2}h^2 e_i^t H(x_0) e_i - \frac{h^3}{6} \frac{\partial^3 f(x_0)}{\partial^3 x_i} + O(h^4) \quad (2)$$

luego, restando las fórmulas (1) y (2), se tiene

$$f(x_0 + he_i) - f(x_0 - he_i) \cong 2hg^t(x_0).e_i + 2 \frac{h^3}{6} \frac{\partial^3 f(x_0)}{\partial^3 x_i}$$

y resulta una fórmula de orden 2 de truncamiento, ya que

$$E_t \sim h^2 \left| \frac{\partial^3 f(x_0)}{\partial^3 x_i} \right|$$

El error de redondeo es, en este caso del mismo orden que el caso anterior, así resulta que el error total cometido para calcular $g_i(x_0)$ es

$$\Upsilon(h) = \frac{E_f |f(x_0)|}{h} + h^2 \left| \frac{\partial^3 f(x_0)}{\partial^3 x_i} \right|$$

y el óptimo resulta

$$h \sim \sqrt[3]{E_f} \sqrt[3]{\frac{f(x_0)}{\frac{\partial^3 f(x_0)}{\partial^3 x_i}}}$$

lo cual hace un error relativo total

$$ert \sim \frac{(E_f)^{\frac{2}{3}} (f(x_0))^{\frac{2}{3}} \left(\frac{\partial^3 f(x_0)}{\partial^3 x_i}\right)^{\frac{1}{3}}}{|g_i(x_0)|} \sim (E_f)^{\frac{2}{3}}$$

que tiene mayor orden que el caso anterior. Nuevamente se supuso aquí que $f(x_0)$, $\frac{\partial^3 f(x_0)}{\partial^3 x_i}$ y $g_i(x_0)$ tienen aproximadamente el mismo orden.

2.2 Fórmulas para derivadas segundas

Dada $f : R^n \rightarrow R$, $x_0 \in R^n$, $h \in R$, podemos calcular numéricamente las derivadas de orden 2 de f usando solo valores de la función.

2.2.1 Fórmula de diferencias hacia adelante

para $i \neq j$

$$\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i} (f_{x_j})$$

y usando las fórmulas de la sección anterior, resulta

$$\begin{aligned} \frac{\partial}{\partial x_i} (f_{x_j}(x_0)) &= \frac{f_{x_j}(x_0 + he_i) - f_{x_j}(x_0)}{h} \\ &= \frac{1}{h} \left[\frac{f(x_0 + he_i + he_j) - f(x_0 + he_i)}{h} - \frac{f(x_0 + he_j) - f(x_0)}{h} \right] \end{aligned}$$

luego

$$\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j} \cong \frac{f(x_0 + he_i + he_j) - f(x_0 + he_i) - f(x_0 + he_j) + f(x_0)}{h^2} \quad (3)$$

para $i = j$

$$\frac{\partial^2 f(x_0)}{\partial^2 x_j} \cong \frac{f(x_0 + 2he_j) - 2f(x_0 + he_j) + f(x_0)}{h^2}$$

Llamaremos H_{ij} a la aproximación para $\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j}$. Para calcular el error de truncamiento en esta aproximación, utilizaremos el desarrollo de Taylor de f , con la siguiente notación: Para $t \in R^n$

$$f(x_0 + ht) = f(x_0) + h \sum_{i=1}^n \frac{\partial f(x_0)}{\partial x_i} t_i + \frac{h^2}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(x_0)}{\partial x_i \partial x_j} t_i t_j +$$

$$+ \frac{h^3}{6} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{\partial^3 f(x_0)}{\partial x_i \partial x_j \partial x_k} t_i t_j t_k + O(h^4)$$

Usando esta fórmula tenemos lo siguiente:

$$f(x_0 + he_i) = f(x_0) + hg_i(x_0) + \frac{h^2}{2} \frac{\partial^2 f(x_0)}{\partial^2 x_i} + \frac{h^3}{6} \frac{\partial^3 f(x_0)}{\partial^3 x_i} + O(h^4) \quad (4)$$

$$f(x_0 + he_j) = f(x_0) + hg_j(x_0) + \frac{h^2}{2} \frac{\partial^2 f(x_0)}{\partial^2 x_j} + \frac{h^3}{6} \frac{\partial^3 f(x_0)}{\partial^3 x_j} + O(h^4) \quad (5)$$

$$f(x_0 + he_i + he_j) = f(x_0) + hg_i(x_0) + hg_j(x_0) + \frac{h^2}{2} \frac{\partial^2 f(x_0)}{\partial^2 x_i} + \frac{h^2}{2} \frac{\partial^2 f(x_0)}{\partial^2 x_j} + h^2 \frac{\partial^2 f(x_0)}{\partial x_i \partial x_j} + \frac{h^3}{6} \left(\frac{\partial^3 f(x_0)}{\partial^3 x_i} + \frac{\partial^3 f(x_0)}{\partial^3 x_j} + 3 \frac{\partial^3 f(x_0)}{\partial^2 x_i \partial x_j} + 3 \frac{\partial^3 f(x_0)}{\partial x_i \partial^2 x_j} \right) + O(h^4) \quad (6)$$

Reemplazando (4), (5) y (6) en (3) se obtiene

$$H_{ij} \cong \frac{\partial^2 f(x_0)}{\partial x_i \partial x_j} + \frac{h}{2} \left(\frac{\partial^3 f(x_0)}{\partial^2 x_i \partial x_j} + \frac{\partial^3 f(x_0)}{\partial x_i \partial^2 x_j} \right)$$

de donde

$$E_t = \frac{h}{2} \underbrace{\left(\frac{\partial^3 f(x_0)}{\partial^2 x_i \partial x_j} + \frac{\partial^3 f(x_0)}{\partial x_i \partial^2 x_j} \right)}_K$$

y por lo tanto resulta una aproximación de orden 1. El error de redondeo, se calcula aquí como en el caso de derivadas primeras, así que

$$Er \sim \frac{E_f |f(x_0)|}{h^2}$$

Luego si se minimiza la suma de estos errores resulta que el h óptimo es

$$h \sim \sqrt[3]{\frac{E_f |f(x_0)|}{K}}$$

Fórmula de diferencias centrales para $i \neq j$ usando la fórmula de diferencias centradas de la sección anterior

$$\begin{aligned} \frac{\partial}{\partial x_i} (f_{x_j}(x_0)) &= \frac{f_{x_j}(x_0 + he_i) - f_{x_j}(x_0 - he_i)}{2h} = \\ &= \frac{1}{2h} \left[\frac{f(x_0 + he_i + he_j) - f(x_0 + he_i - he_j)}{2h} - \frac{f(x_0 - he_i + he_j) - f(x_0 - he_i - he_j)}{2h} \right] \end{aligned}$$

luego

$$\begin{aligned} \frac{\partial^2 f(x_0)}{\partial x_i \partial x_j} &\cong \frac{f(x_0 + he_i + he_j) - f(x_0 + he_i - he_j)}{4h^2} + \\ &\quad + \frac{f(x_0 - he_i - he_j) - f(x_0 - he_i + he_j)}{4h^2} \end{aligned}$$

para $i = j$

$$\frac{\partial^2 f(x_0)}{\partial^2 x_j} \cong \frac{f(x_0 + he_j) - 2f(x_0) + f(x_0 - he_j)}{h^2}$$

Error de truncamiento para esta aproximación: Usando las fórmulas (4),(5) y (6), del desarrollo de Taylor para f obtenemos

$$H_{ij} \cong \frac{\partial^2 f(x_0)}{\partial x_i \partial x_j} + \frac{O(h^4)}{4h^2}$$

lo que muestra que la fórmula de diferencias centrales tiene un error de truncamiento de orden 2.

3 Otros métodos

El método propuesto en este trabajo ha sido comparado con los siguientes:

3.1 Método de Gill-Murray

Dada $f : R^n \rightarrow R$ este método aproxima los elementos del hessiano de la función usando la fórmula de diferencias finitas

$$\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j} \cong \frac{f(x_0 + h_i e_i + h_j e_j) - f(x_0 + h_i e_i) - f(x_0 + h_j e_j) + f(x_0)}{h_i h_j}$$

con

$$h_i = \sqrt[3]{tol} \max\{0.1, |x_{0i}|\}$$

$$h_j = \sqrt[3]{tol} \max\{0.1, |x_{0j}|\}$$

Esta elección del incremento tiene la siguiente explicación, en la sección.2 se vio que una buena elección del incremento para la técnica de diferencias finitas es aquella en que se compensen el error de redondeo y el error de truncamiento de la serie de Taylor, haciendo algunas cuentas, el resultado fue

$$h \sim \sqrt[3]{\frac{E_f |f(x_0)|}{K}}$$

con $K = \frac{\partial^3 f(x_0)}{\partial^2 x_i \partial x_j} + \frac{\partial^3 f(x_0)}{\partial x_i \partial^2 x_j}$ como $E_f \simeq E_m$ donde E_m es la precisión de la máquina con la que se está trabajando, es decir, *tol*. Además

$$x_c = \sqrt[3]{\frac{f(x_0)}{K}}$$

se llama la escala de curvatura de la función f , que por supuesto no está disponible, así que en ausencia de otra información, se utiliza $x_c = x_{0i}$ donde i es la coordenada que se desea incrementar, excepto que sea $x_{0i} = 0$ pues en ese caso, el incremento se anularía y la aproximación sería errónea. Es por esto que se toma $h = \max\{0.1, |x_{0i}|\}$ para evitar este problema.

3.2 Método por diferencia de gradientes y simetrizando

La subrutina NUGRAD calcula aproximadamente el gradiente de una función $f : R^n \rightarrow R$, en un punto x_0 , luego se obtiene $g = (g_1(x_0), \dots, g_n(x_0))$ entonces utiliza la fórmula de diferencias finitas para calcular una matriz M , siendo

$$M_{ij} = \frac{g_i(x_0 + h_j e_j) - g_i(x_0)}{h_j}$$

donde el incremento h_j es el mismo que usó para calcular g_j . Sin embargo, esta no es una buena aproximación para el hessiano de una función $f \in C^2$, ya que probablemente M no sea simétrica, entonces toma H la matriz simétrica más cercana a M , en norma de Frobenius, es decir $H = \frac{M+M^t}{2}$ y H resulta la matriz que aproxima al hessiano de la función. Para demostrar esto, tenemos el siguiente lema:

Lema:

Sea $A \in R^{n \times n}$ no singular, y

$$B = \frac{A + A^t}{2}$$

entonces

$$\|A - B\|_F = \min_X \|X - A\|_F, \forall X \text{ simétrica}$$

Dem: B es una matriz simétrica, pues

$$B^t = \left(\frac{A + A^t}{2}\right)^t = \frac{A^t + A}{2} = B$$

Además sea

$$F(x_{11}, \dots, x_{1n}, x_{21}, \dots, x_{nn}) = \|X - A\|_F$$

es decir $F(x_{11}, \dots, x_{1n}, x_{21}, \dots, x_{nn}) = \sum_{i=1}^n \sum_{j=1}^n (x_{ij} - a_{ij})^2$ entonces, como $x_{ij} = x_{ji} \frac{\partial F}{\partial x_{ij}} = 2(x_{ij} - a_{ij}) + 2(x_{ij} - a_{ji})$ luego $\frac{\partial F}{\partial x_{ij}} = 0 \Leftrightarrow 2x_{ij} - a_{ij} - a_{ji} = 0$ por lo tanto debe ser $x_{ij} = \frac{a_{ij} + a_{ji}}{2}$. Luego el punto crítico es $X = B$ que es un mínimo, ya que toda norma es una función convexa.

4 Método propuesto en este trabajo

Este algoritmo fue realizado con el objeto de calcular derivadas segundas de funciones $f : R^n \rightarrow R$ usando solo valores de la función. Está basado en la idea del Dr. Hugo Scolnik, implementada en la subrutina NUGRAD que calcula numéricamente derivadas primeras de este tipo de funciones. La subrutina que implementa el método se llama NUMHES. Dada una función que se supone continua y dos veces derivable, calcula los elementos del hessiano de dicha función. Las fórmulas de diferencias finitas usadas en el algoritmo, surgen de las consideraciones hechas en la sección 2:

$$\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j} \cong \frac{f(x_0 + h_i e_i + h_j e_j) - f(x_0 + h_i e_i) - f(x_0 + h_j e_j) + f(x_0)}{h_i h_j}$$

Lo más importante en la implementación de esta técnica es la elección de los incrementos h_i, h_j . El algoritmo propuesto en este trabajo proporciona una nueva forma de elegirlos. El procedimiento es el siguiente: se eligen una cantidad arbitraria de funciones test, que este caso fueron 13 $\{f_1, f_2, \dots, f_{13}\}$ de las cuales se conoce los hessianos verdaderos $\{Hv^1, Hv^2, \dots, Hv^{13}\}$ representativas de las que se utilizan en la práctica (polinomiales, sinusoidales, exponenciales, etc.) y se construye una función de error de la siguiente manera: Se supone que h_i depende de ciertos parámetros $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ a través de una ecuación, es decir que

$$h_i = h(\alpha_1, \alpha_2, \alpha_3, \alpha_4, x_i, f(x))$$

para elegir la fórmula a utilizar se han probado distintas alternativas, procurando conseguir una invariancia ante cambios de escala de las variables; algunas de ellas fueron las siguientes:

$$\begin{aligned} x_i &= a10^b \\ f(x) &= c10^d \end{aligned}$$

- 1) $h_i = \alpha_1^b \cdot \alpha_2^d \cdot \alpha_3 + \alpha_4 \cdot |x_i|$
 - 2) $h_i = (\alpha_1^b + \alpha_2^d)^{\alpha_3} + \alpha_4 \cdot |x_i|$
 - 3) $h_i = (\alpha_1^b - \alpha_2^d) \cdot \alpha_3 + \alpha_4 \cdot \max(|x_i|, 0.1)$
 - 4) $h_i = (\alpha_1 \cdot b \pm \alpha_2 \cdot d) \alpha_3^d + \alpha_4 \cdot \max(|x_i|, 0.1)$
 - 5) $h_i = (\alpha_1 |x_i|)^b + (\alpha_2 |f(x)|)^d \cdot \alpha_3 + \alpha_4 \cdot |x_i|$
 - 6) $h_i = \text{sen}(\alpha_1^b \cdot \alpha_2^d \cdot \alpha_3) + \alpha_4 \cdot \max(|x_i|, 0.1)$
- siendo la que arrojó mejores resultados

$$h_i = (\alpha_1^b + \alpha_2^d) \cdot \alpha_3 + \alpha_4 \cdot \max(|x_i|, 0.1)$$

por lo tanto es la que finalmente se usó. Luego se calculan los hessianos por diferencias finitas con este incremento y quedan los hessianos "calculados", dependiendo de los mismos parámetros

$$\begin{aligned} &Hc^1(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \\ &Hc^2(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \\ &\dots \\ &Hc^{13}(\alpha_1, \alpha_2, \alpha_3, \alpha_4) \end{aligned}$$

Construimos entonces, una función objetivo de la siguiente manera: Dadas 10 matrices $\{X_1, \dots, X_{10}\}$ conteniendo en la columna m , el punto en el que se quiere evaluar el hessiano de la función m . Es importante destacar que el hessiano de cada una de las funciones es evaluado en 10 puntos distintos, lo cual proporciona un amplio

espectro de valores para los puntos iniciales y para las derivadas. En las columnas de las matrices X_i hay todo tipo de valores (ceranos a cero, cercanos a 1000, etc.) que dan distintos valores para las derivadas. Tomamos

$$F_k(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = \sum_{m=1}^{13} \sum_i \sum_j \left| \frac{Hv_{ij}^m(x_k) - Hc_{ij}^m(\alpha_1, \alpha_2, \alpha_3, \alpha_4, x_k)}{Hv_{ij}^m(x_k)} \right|$$

Aquí $Hv_{ij}^m(x_k)$ significa que los hessianos se evalúan en los puntos que contiene la matriz X_k . entonces la función objetivo queda

$$F(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = \sum_{k=1}^{10} F_k(\alpha_1, \alpha_2, \alpha_3, \alpha_4) =$$

$$\sum_{k=1}^{10} \sum_{m=1}^{13} \sum_i \sum_j \left| \frac{Hv_{ij}^m(x_k) - Hc_{ij}^m(\alpha_1, \alpha_2, \alpha_3, \alpha_4, x_k)}{Hv_{ij}^m(x_k)} \right|$$

que luego se minimiza con el método de minimización de Powell. Esto da como resultado un α^* óptimo, que se usa para calcular el incremento de la fórmula de diferencias finitas, es decir

$$h_i = ((\alpha_1^*)^b + (\alpha_2^*)^d) \cdot \alpha_3^* + \alpha_4^* \cdot \max(|x_i|, 0.1)$$

fue el tamaño de paso usado finalmente para aproximar derivadas segundas con la técnica de diferencias finitas. Cuando el método se implementa, en la subrutina NUMHES, se pone como condición que

$$h_i \geq tol \cdot |x_i|$$

si el h_i calculado no cumple con esta condición entonces se toma

$$h_i \approx \sqrt[3]{tol} \max(0.1, |x_i|)$$

de esta manera se asegura la primer condición que debe cumplir el incremento en la sección 3. Este método ha sido probado para muchas funciones y los resultados están expuestos en la sección de Resultados numéricos.

5 Conclusiones

En este trabajo, se ha desarrollado un método nuevo para aproximar derivadas segundas, se ha probado y comparado con otros dos algoritmos, el de Gill-Murray y el de diferencia de gradientes y simetrizando.

M1-Gill-Murray

M2-diferencia de gradientes y simetrizando

M3- método nuevo

Como estos métodos solo difieren en la forma de elegir el incremento, han sido probados de dos maneras:

1-Con diferencias finitas hacia adelante

2-Con diferencias finitas centrales.

Los resultados fueron los siguientes: Con la primer forma, diferencias hacia adelante, se han hecho treinta y una pruebas, entre las cuales el M3 ha resultado con menor error, en veinticinco casos, el M1, en seis casos y el M2 en ninguno. Esto significa que M3 fue mejor en el 80,6% de las pruebas. Sin embargo, en aquellas funciones en que el M1 resultó mejor lo hizo con errores muy pequeños, del orden de 10^{-14} mientras los otros métodos arrojan errores del orden de 10^{-2} . El número de evaluaciones de función de M1 y M3 es el mismo, mientras que M2 es más costoso. Con la segunda forma, diferencias centrales, M3 resultó mejor en todos los casos, salvo para las funciones cuadráticas en las que el M1 resultó con menor error. Pudo observarse también que en todos los casos en que M3 fue el mejor, M2 fue el segundo, es decir mejor que M1, aunque utiliza muchas más evaluaciones de función. Los errores con diferencias centrales, son mucho menores que con diferencias comunes ya que, como se vio en el capítulo 2, el orden de error de truncamiento de la serie es más grande, pero es más costoso en cuanto a evaluaciones de función. A medida que la dimensión de la matriz a calcular crece, los errores en el cálculo de la derivada segunda también crecen, esto se debe a que aumentan los errores aritméticos. Se ha observado también que el M1 no tiene grandes diferencias en los errores si se usa con diferencias hacia adelante o con diferencias centradas, esto se debe a que el error predominante es el aritmético, y no el de truncamiento. La novedad del

método "nuevo" que se ha presentado, es que en la mayoría de los casos probados se ha obtenido una buena aproximación para la derivada segunda, sin aumentar el número de evaluaciones de función. En este trabajo, se logró aproximar las derivadas segundas con mejor precisión que los métodos usuales y eso es esencial para las aplicaciones en Optimización No Lineal porque la matriz hessiana define la métrica que se utiliza y por lo tanto incide sobre la velocidad de convergencia. Más aún, las fórmulas utilizadas son invariantes ante cambios de escala, propiedad que no poseen los algoritmos clásicos.

6 Resultados Numéricos

Por razones de espacio se presentan aquí solo una parte de las pruebas realizadas. El error es la suma de los errores relativos de cada término del hessiano.

Función cuadrática

Con diferencias hacia adelante

	$M1$	$M2$	$M3$
error	$0.2294d - 14$	$0.5163d - 01$	$0.1874d - 07$
num. de eval	7	9	7

Con diferencias centrales

	$M1$	$M2$	$M3$
error	$0.2294d - 14$	$0.8695d - 05$	$0.1264d - 08$
num. de eval	9	15	9

Función Gaussiana(n=3)

Con diferencias hacia adelante

	$M1$	$M2$	$M3$
error	$0.4367d - 01$	$0.15105d + 01$	$0.3288d - 03$
num. de eval	13	16	13

Con diferencias centrales

	$M1$	$M2$	$M3$
error	$0.9412d - 02$	$0.1224d - 03$	$0.9118d - 06$
num. de eval	19	15	19

Función de Biggs(n=6)

Con diferencias hacia adelante

	$M1$	$M2$	$M3$
error	$0.8514d - 02$	$0.3362d - 00$	$0.2580d - 02$
num. de eval	43	49	43

Función extendida de Rosenbrock(n=12)

Con diferencias centrales

	$M1$	$M2$	$M3$
error	$0.1313d + 03$	$0.22003d + 01$	$0.3959d - 01$
num. de eval	289	325	289

References

- [1] Fletcher, R. (1987), Practical Methods of Optimization. Wiley, New York
- [2] Gill, P.E, Murray, W. and Pitfield, R.A. (1972). The implementation of two modified Newton algorithms for unconstrained optimization. NPL Report NAC11.
- [3] Gill, P.E, Murray, W., Saunders, M.A. and Wright, M.H (1980), Computing the Finite-Difference Aproximations to Derivatives for Numerical Optimization. Technical Report, Departament of Operations Research-Stanford University.
- [4] Gill, P.E, Murray, W., Saunders, M.A. and Wright, M.H (1991) Numerical Linear Algebra and Optimization, vol.1. Addison Wesley.
- [5] Gill, P.E, Murray, W., Saunders, M.A. and Wright, M.H (1993) Practical Optimization. Academic Press.
- [6] Powell, M.J.D. (1964) An efficient method for finding the minimum of a function of several variables without calculating derivatives. Computer Journal, vol.7, pp 155-162.
- [7] Press, (1992), Numerical Recipes, Academic Press.

- [8] G.W.Stewart(1967) A modification of Davidon's minimization method to accept difference approximations to derivatives,J.Ass.Comput.Mach.,14,72-83.
- [9] G.R.Walsh(1975), Methods of Optimization, Wiley, London.