

Tarea 07 - Optimización

Giovanni Gamaliel López Padilla

Resumen

Los métodos del descenso del gradiente y de Newton son los métodos más simples en la rama de la optimización. Esto debido a que se basan en principios matemáticos los cuales indican que existe un mínimo global o local en una función con dominio convexo. En este trabajo se analizó la eficiencia de estos dos métodos usando las funciones de Rosembrock, de Wood y de suavizado. Llegando así a que el método del descenso del gradiente tiende a obtener mínimos locales cercanos al mínimo global en cuestión del valor de la función a analizar. Por otro lado el método de Newton obtiene un mínimo local cercano al punto inicial con un número menor de iteraciones a comparación del método del descenso del gradiente.

Índice

1. Introducción	1
2. Marco teórico	2
2.1. Región de confianza	2
2.2. Punto de Cauchy	3
2.3. Método Dogleg	4
2.4. Método de Newton con Cauchy	4
2.5. Método de Newton modificado	4
3. Métodos	5
3.1. Función de Rosembrock	5
3.2. Función de Wood	6
3.3. Función de Branin	7
3.4. Parámetros	7
4. Resultados	7

1. Introducción

La optimización es procedimiento que con sus resultados se toman de decisiones en el análisis de sistemas físicos. Para realizar es la optimización de un sistema o situación se debe de contemplar un objetivo, el cual debe ser caracterizado por una función cuantitativa.

El resultado de realizar la optimización a un sistema puede representarse como un ahorro de tiempo, energía o cualquier objeto que pueda ser reflejado en un número. El objetivo del proceso de una optimización es obtener un conjunto de números o características que representen un mínimo o máximo de objeto el cual esta siendo caracterizado. Esto puede ser representado como se encuentra en la ecuación 1.

$$\min_{x \in \mathbb{R}} f(x) \quad (1)$$

donde f es la caracterización cuantitativa del problema y x son los objetos que interactúan con el sistema.

Una optimización local es una solución al problema de optimización en una vecindad alrededor del valor de x encontrado. En cambio una optimización global es aquella solución que es menor o mayor con respecto a todas las demás. La solución de un proceso de optimización no siempre encontrará los valores en que el sistema se sitúe una optimización global.

2. Marco teórico

Una técnica para la optimización de funciones es definir la función de objetivo, generalmente para obtener alguna solución del problema se escogen pasos de búsqueda unidireccionales.

2.1. Región de confianza

Un método empleado para acotar las soluciones del problema es usar una región de confianza dado un punto. Para formalizar este método se define un modelo que aproxima a la función objetivo en un punto. El modelo se encuentra descrito en la ecuación 2.

$$m_k(p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p \quad (2)$$

donde su mínimo se encuentra dentro de la región de confianza.

Una medida para calcular el radio de la región de confianza Δ_k es la medida de ajuste que está definida en la ecuación 2.1.

$$\rho_k = \frac{f(x_k) - f(x + p_k)}{m_k(0) + m_k(p_k)}$$

Donde el numerador representa la reducción en la función y el denominador la reducción en el modelo. El denominador siempre es positivo debido a que p_k minimiza el modelo en cada iteración. En el caso en que ρ_k sea un valor negativo este representa un incremento en la función f . Por lo que el paso deberá rechazarse. Si ρ_k es muy cercano a uno, entonces el comportamiento de la función f y el modelo m_k tienen una semejanza, por lo que se optaría a incrementar la región de confianza. Si ρ_k se encuentra entre 0 y 1, entonces se opta por no realizar modificaciones en la región de confianza. Si ρ_k es cercano a cero o negativo se propone una reducción en la región de confianza.

El algoritmo 1 describe este método.

Algorithm 1: Región de confianza

Input: $\hat{\Delta}, \Delta_0 \in (0, \hat{\Delta})$ y $\eta \in [0, \frac{1}{4}]$

```

1 for  $k=0,1,2,3$  do
2    $p_k \leftarrow \arg \min$  ecuación 2
3    $\rho_k \leftarrow$  ecuación 2.1
4   if  $\rho_k < \eta_1$  then
5      $\Delta_{k+1} \leftarrow \hat{\eta}_1 \Delta_k$ 
6   else if  $\rho_k > \eta_2$  y  $\|p_k\| = \Delta_k$  then
7      $\Delta_{k+1} \leftarrow \min\{\hat{\eta}_2 \Delta_k, \hat{\Delta}\}$ 
8   else
9      $\Delta_{k+1} \leftarrow \Delta_k$ 
```

2.2. Punto de Cauchy

Se denomina al arco de Cauchy al segmento tal que cumple la ecuación 3.

$$x_k^C(t) = \{x : x = x_k + t\nabla f(x_k), t \leq 0, \|t\nabla f(x_k)\| < \Delta_k\} \quad (3)$$

El punto de Cauchy se encuentra definido en la ecuación 4.

$$p_k^S = \arg \min_{\|p\| < \Delta_k} m_k(p) \quad (4)$$

donde Δ_k es la región de confianza.

Teniendo el punto p_k^S , se buscará un parámetro τ_k , el cual minimice el modelo de la ecuación 2 en la región de confianza (ecuación 5).

$$\tau_k = \arg \min_{\tau_k \geq 0} m_k(\tau p_k^S) \leq \Delta_k \quad (5)$$

por lo que el punto de Cauchy, lo estaremos calculando de tal manera que $p_k^C = \tau_k p_k^S$. Tomando a p_k como $-\lambda_k g_k$, sujeto a la región de confianza, se tiene que

$$\|p\| = \|-\lambda_k g_k\| \Rightarrow \lambda_k \leq \frac{\Delta_k}{\|g_k\|}$$

Por lo que la ecuación 5 se puede escribir como en la ecuación .

$$\lambda_k = \arg \min_{\lambda \in [0, \hat{\lambda}]} m(-\lambda g_k) \quad (6)$$

La solución al problema planteado en la ecuación 6 se encuentra descrita en la ecuación 7.

$$\lambda_k = \hat{\lambda} \begin{cases} 1 & \text{si } g_k^T B_k g_k \\ \min \left(1, \frac{\|g_k\|^3}{\Delta_k g_k^T B_k g_k} \right) & \text{en cualquier otro caso} \end{cases} \quad (7)$$

Por lo tanto, el punto de Cauchy se obtiene como en la ecuación 8.

$$p_k^C = -\tau_k \frac{\Delta_k}{\|g_k\|} g_k \quad (8)$$

El algoritmo 2 describe este método.

Algorithm 2: Método de Dogleg con región de confianza

Input: $\hat{\Delta} > 0$, $\Delta_0 \in (0, \hat{\Delta})$ y $\eta \in [0, \frac{1}{4}]$

```

1 for  $k = 0, 1, 2, 3, \dots$  do
2    $p_k \leftarrow$  ecuación 9
3    $\rho_k \leftarrow$  ecuación 2.1
4   if  $\rho_k > \eta$  then
5      $x_{k+1} \leftarrow x_k + p_k$  else
6      $x_{k+1} = x_k$ 
7    $\Delta_{k+1} \leftarrow$  algoritmo 1
```

2.3. Método Dogleg

La idea del método Dogleg es minimizar el modelo cuadrático sin restricciones a lo largo del gradiente. La trayectoria Dogleg toma como primera instancia el punto p_k^U y como segunda línea va desde p_k^U hasta p_k^B . Lo anterior puede resumirse en la ecuación 9.

$$\check{p}(\tau) = \begin{cases} \tau p_k^U & \text{si } 0 \leq \tau \leq 1 \\ p_k^U + (\tau - 1)(p_k^B - p_k^U) & \text{si } 1 \leq \tau \leq 2 \end{cases} \quad (9)$$

Si se toma a p_k^U en dirección al gradiente, de tal manera que $p_k^U = \alpha \nabla f_k$, entonces el problema será encontrar α tal que p_k^U minimiza al modelo cuadrático (ecuación 2). Llegando a que α debe tener el valor descrito en la ecuación .

$$\alpha^* = -\frac{\|\nabla f_k\|}{\nabla^T f_k B_k \nabla f_k} \quad (10)$$

Por lo tanto, los puntos p_k^U y p_k^B son:

$$p_k^U = -\frac{\|\nabla f_k\|}{\nabla^T f_k B_k \nabla f_k} \nabla f_k \quad p_k^B = -B_k^{-1} \nabla f_k$$

2.4. Método de Newton con Cauchy

El método de Newton para sistemas no lineales supone que existe una función f definida en \mathbb{R}^n y de clase C^2 tal que tiene un mínimo relativo en x^* . Suponemos que el eigenvalor menor del hessiano $H(x^*)$ de f es mayor a cero. Esto provoca que $H(x^*)$ sea positiva definida. Entonces el método de Newton se puede obtener resolviendo el sistema matricial descrito en la ecuación 11.

$$H(x_k)d_k = -g(x_k) \quad (11)$$

Donde $g(x_k)$ y $H(x_k)$ son el hessiano y el gradiente de f en el paso x_k . Con esta dirección obtenida se puede calcular el siguiente paso del método (ecuación 12).

$$x_{k+1} = x_k + d_k \quad (12)$$

Las desventajas de usar el método de Newton es que no se suele obtener una soluciones globales. El costo computacional de calcular el hessiano y el calculo de la solución de la ecuación 11 en cada iteración es grande.

Al incorporar una región de confianza al método de Newton, se estimaría si el paso obtenido en la solución de la ecuación 11 sale de la región de confianza Si este es mayor, entonces se toma el paso de Cauchy descrito en la ecuación 8, si no se toma el paso de Newton calculado. A esto se le conoce como el método de Newton-Cauchy.

2.5. Método de Newton modificado

Una manera de mejorar el método de Newton es el uso de una búsqueda en línea de un parámetro α que regule el paso d_k para que se garantice una convergencia.

Si se llega a obtener un hessiano que sea no singular, entonces existe una β y η tal que

$$\|H(x_0)\| \leq \beta \quad \|H(x_0)^{-1}g(x_0)\| \leq \eta$$

Si definimos a α como $\alpha = \beta\gamma\eta$ y $r \geq r_0 \equiv (1 - 2\sqrt{1 - 2\alpha})/(2\beta\gamma)$, entonces la secuencia x_k obtenida del método de Newton está bien definida y converge a x^* , el único cero de f en $B(x_0, r_0)$. Si $\beta\gamma\eta < \frac{1}{2}$, entonces x^* es el único cero en $B(x_0, r_1)$ donde

$$r_1 = \min \left(r, \frac{1 + \sqrt{1 - 2\alpha}}{\beta\gamma} \right) \quad ||x_k x^*|| \leq (2\alpha)^{2k} \quad k = 1, 2, \dots$$

Si se obtiene un hessiano que no sea positivo definido en la dirección de newton, entonces su paso puede no ser de un descenso. Una alternativa es modificar el hessiano tal que

$$B_k = H(x_k) + E_k$$

tal que $B_k \succ 0$ y la dirección usando esta matriz sea de descenso. Para simplificar la elección de E_k , se propone el siguiente calculo:

$$E_k = \tau_k \mathbf{I} \quad \tau_k > 0$$

En el algoritmo 3 se encuentra el calculo de B_k .

Algorithm 3: Calculo de la matriz B_k usando la factorización de Cholesky

Input: H

```

1  $\beta \leftarrow 10^{-3}$ 
2  $\lambda \leftarrow \text{eigenvalues}(H)$ 
3 if  $\min(\lambda > 0)$  then
4    $\tau_0 = 0$ 
5 else
6    $\tau_0 \leftarrow \beta - \min(\lambda)$ 
7 for  $k = 1, 2, \dots$  do
8    $LL^T \leftarrow A + \tau_k \mathbf{I}$  if  $L \exists$  then
9     return  $L$ 
10  else
11     $\tau_{k+1} \leftarrow \max(2\tau_k, \beta)$ 
```

El uso de la matriz B_k , la búsqueda de linea y el método de Newton se le llama método de Newton modificado (algoritmo 4).

Algorithm 4: Método de Newton modificado

Input: x_0
Output: x^*

```

1 while  $||\nabla f_k|| > \tau_g$  do
2    $B_k \leftarrow \nabla^2 f(x_k) + E_k$ 
3   Solve  $B_k d_k = -g_k$ 
4   Compute  $\alpha_k$  with line search
5    $x_{k+1} \leftarrow x_k + \alpha_k d_k$ 
```

3. Métodos

3.1. Función de Rosembrock

La función de Rosembrock se define en la ecuación 13.

$$f(x) = \sum_{i=1}^{n-1} 100(x_{i+1} - x_i^2)^2 + (1 - x_i) \quad (13)$$

donde $x \in \mathbb{R}^n$

Con la función de Rosembrock definida, se tiene que su gradiente es calculado con en la ecuación 14.

$$\nabla f(x) = \begin{cases} -400x_i(x_{i+1} - x_i^2) & \text{para } i = 1 \\ 200(x_i - x_{i-1}^2 - 400x_i(x_{i+1} - x_i^2) - 2(1 - x_i)) & \text{para } 1 < i < n \\ 200(x_i - x_{i-1}) & \text{para } i = n \end{cases} \quad (14)$$

Con la función de Rosembrock definida, se tiene que su hessiano es calculado con la ecuación 15.

$$\nabla^2 f(x) = \begin{cases} \frac{\partial^2 f}{\partial x_i^2} = -200 & \text{para } i = 1 \\ \frac{\partial^2 f}{\partial x_i^2} = 1200x_i^2 - 400x_{i+1} + 202 & \text{para } 1 < i < n \\ \frac{\partial^2 f}{\partial x_i \partial x_{i+1}} = -400x_i & \text{para } 0 \leq i < n \\ \frac{\partial^2 f}{\partial x_i^2} = 200 & \text{para } i = n \end{cases} \quad (15)$$

3.2. Función de Wood

La función de Wood se define en la ecuación 16.

$$f(x) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2 + (x_3 - 1)^2 + 90(x_3 - x_4)^2 + 10.1((x_2 - 1)^2 + (x_4 - 1)^2) + 19.8(x_2 - 1)(x_4 - 1) \quad (16)$$

Donde $x \in \mathbb{R}^4$.

Con la función de Wood definida, podemos obtener le gradiente de la función de Wood. El resultado del gradiente de la función de Wood se encuentra en la ecuación 17.

$$\nabla f(x) = \begin{cases} \frac{\partial f}{\partial x_1} = 400(x_1^2 - x_2)x_1 + 2(x_1 - 1) \\ \frac{\partial f}{\partial x_2} = -200(x_1^2 - x_2) + 20.2(x_2 - 1) + 19.8(x_4 - 1) \\ \frac{\partial f}{\partial x_3} = 2(x_3 - 1) + 360(x_3^2 - x_4)x_3 \\ \frac{\partial f}{\partial x_4} = -180(x_3^2 - x_4) + 20.2(x_4 - 1) + 19.8(x_2 - 1) \end{cases} \quad (17)$$

De igual forma, se puede obtener el hessiano de la función de Wood. El resultado del Hessiano se encuentra en la ecuación 18.

$$\nabla^2 f(x) = \begin{cases} \frac{\partial^2 f}{\partial x_1^2} = 400(x_1^2 - x_2) + 800x_1^2 + 2 \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = -400x_1 \\ \frac{\partial^2 f}{\partial x_2^2} = 220.2 \\ \frac{\partial^2 f}{\partial x_4 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_4} = 19.8 \\ \frac{\partial^2 f}{\partial x_3^2} = 720x_3^2 + 360(x_3^2 - x_4) + 2 \\ \frac{\partial^2 f}{\partial x_4 \partial x_3} = \frac{\partial^2 f}{\partial x_3 \partial x_4} = -360x_3 \\ \frac{\partial^2 f}{\partial x_4^2} = 200.2 \end{cases} \quad (18)$$

3.3. Función de Branin

La función de Branin se define en la ecuación 19.

$$f(x) = a(x_2 - bx_1^2 + cx_1 - r)^2 + s(1 - t)\cos(x_1) + s \quad (19)$$

Donde $a = 1$, $b = 5.1/(4\pi^2)$, $c = 5/\pi$, $r = 6$, $s = 10$ y $t = 1/(8\pi)$.

Con la función de Branin definida, podemos obtener el gradiente de la función de Branin. El resultado del gradiente de la función de Branin se encuentra en la ecuación 20.

$$\nabla f(x) = \begin{cases} \frac{\partial f}{\partial x_1} = 2a(x_2 - bx_1^2 + cx_1 - r)(c - 2bx_1) - s(1 - t)\sin(x_1) \\ \frac{\partial f}{\partial x_2} = 2a(x_2 - bx_1^2 + cx_1 - r) \end{cases} \quad (20)$$

De igual forma, se puede obtener el hessiano de la función de Branin. El resultado del Hessiano se encuentra en la ecuación 21.

$$\nabla^2 f(x) = \begin{cases} \frac{\partial^2 f}{\partial x_1^2} = 2a(c - 2bx_1)^2 - 4ab(x_2 - bx_1^2 + cx_1 - r) - s(1 - t)\cos(x_1) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} = 2a(c - 2bx_1) \\ \frac{\partial^2 f}{\partial x_2^2} = 2a \end{cases} \quad (21)$$

3.4. Parámetros

Los parámetros que se usaron para cada metodo se encuentran en la tabla 1.

Función	Δ_k	Δ_{max}	η	η_{min}	η_{max}	$\hat{\eta}_1$	$\hat{\eta}_2$	c_1	c_2
Wood	0.06	0.1	0.05						
Rosembrock	0.05	0.1	0.01	0.25	0.75	0.25	2.0	10^{-4}	0.9
Branin	0.1	0.2	0.1						

Tabla 1: Parámetros para los metodos de Newton-Cauchy, Newton modificado, Dogleg y Cauchy.

Las condiciones de paro usadas fueron la diferencia de la posicion actual y la anterior, la diferencia del valor de la función anterior y actual y la cercania del gradiente a cero. Todas con una tolerancia igual a 10^{-12} .

4. Resultados

Se ejecutaron 30 veces cada función con diferente método. Cada ejecución partio desde el punto

$$x_0 = x^* + \mathcal{U}(-2, 2)$$

Donde el punto óptimo para era dado para cada función. El promedio del tiempo de ejecución, valor de la función en la última iteración, el valor del gradiente en la última y el número de iteraciones se encuentran en las tablas 2, 3, 4 y 5 respectivamente.

Funcion	Newton-Cauchy	Newton modificado	Dogleg	Cauchy
wood	1.28845	0.027438	0.330035	1.85432
rosenbrock	74.3634	1.58818	41.6005	44.9421
branin	0.0171616	0.00662657	0.0159199	0.0343389

Tabla 2: Tiempo promedio en segundos de las ejecuciones realizadas con cada función y método implementado.

Funcion	Newton-Cauchy	Newton modificado	Dogleg	Cauchy
wood	7.00204e-08	0.201103	2.65275	0.00010695
rosenbrock	1.19599	19090.9	32.0271	0.39945
branin	0.397888	0.648666	0.397887	0.397889

Tabla 3: Promeio del valor de la función en la última iteración de las ejecuciones realizadas con cada función y método implementado.

Funcion	Newton-Cauchy	Newton modificado	Dogleg	Cauchy
wood	0.00190684	1.08682	0.343427	0.0198407
rosenbrock	0.00176435	6112.75	1.9312	0.0430613
branin	0.000997418	0.278091	0.000877879	0.00262148

Tabla 4: Promedio del valor del gradiente en la última iteración de las ejecuciones realizadas con cada función y método implementado.

Funcion	Newton-Cauchy	Newton modificado	Dogleg	Cauchy
wood	1044	15.4667	258.467	1538.7
rosenbrock	6623.5	63.3	4912.67	6845.4
branin	11.1	3.66667	10.2	22.5

Tabla 5: Iteraciones promedio de las ejecuciones realizadas con cada función y método implementado.