Tarea 6: Método de Newton y Newton Modificado

Optimización I

Maestría en Computación
Centro de Investigación en Matemáticas

Esteban Reyes Saldaña
esteban.reyes@citmat.mx

9 de marzo de 2021

Resumen

En esta tarea se utilizó el método de Newton y el método de Newton modificado para la función de Rosembrock y la función de Wood. El método de Newton busca una dirección de descenso haciendo uso de un sistema lineal entre el Hessiano el gradiente. El método de Newton modificado intenta encontrar una mejor aproximación de dicho sistema lineal haciend una modificación para que la matriz Hessiana sea definida positiva. Se presenta a continuación una descripción general, así como el pseudocódigo de los métodos implementados. Finalmente se incluyen conclusiones observadas a partir de la experimentación.

---

1 Introducción

Definición 1.1. Una **dirección de descenso** $d \in \mathbb{R}^n$ para $f \in \mathcal{C}^1$ es un vector tal que

$$f(x + td) < f(x)$$

para $t \in (0, T)$. Es decir, permite que el punto x más cerca al mínimo local x^* de la función objetivo $f : \mathbb{R}^n \to \mathbb{R}$.

Teorema 1.1. Si $g(x)^T d < 0$ entonces d es una dirección de descenso.

Observación. La dirección

$$d_k = -g(x_k)$$

En descenso Gradiente con búsqueda exacta se vio que

- Con búsqueda en línea dos direcciones consecutivas son ortogonales, esto es

$$g_k \perp g_{k+1}$$

- La trayectoria solución con búsqueda exacta sigue un patrón zig-zag.

2 Método

2.1 Método de Newton

Dada $g : \mathbb{R}^n \to \mathbb{R}^n$ contiuamente diferenciable y $x_0 \in \mathbb{R}^n$: en cada iteración k, (en nuestro caso $g = \nabla f$ y $Dg = \nabla^2 f = H$)

1. Resolver $Dg(x_k)d_k = -g(x_k)$.
2. Actualizar $x_{k+1} = x_k + d_k$

La derivada de $g = \nabla f(\cdot)$ en x es la matriz Jacobiana de $g = \nabla f(\cdot)$ y la denotamos por $J(x_k) = Dg(x_k)$ o el Hessiano de f, esto es, $H(x_k) = Dg(x_k) = \nabla^2 f(x_k)$.
Algunas de las **ventajas** de este método son

1. Convergencia cuadrática para un buen punto inicial si $H(x^*) = \nabla^2 f(x^*)$ es no singular.
2. Solución exacta en una iteración para un ∇f afín (exacto en cada iteración para cualquiera de las funciones componentes de ∇f).

Algunas desventajas son

1. En general no es convergencia global.
2. Requiere calcular el Hessiano $H(x_k) = \nabla^2 f(x_k)$ en acada iteración.
3. En cada iteración se requiere solucionar un sistema lineal de ecuaciones que podría tener una matriz singular asociada o mal condicionada.
4. $d_k = -H(x_k)^{-1} g(x_k)$ podrpia no ser una dirección de descenso.

2.2 Método de Newton Modificado

Teorema 2.1. Suponga que $f \in \mathcal{C}^3$ y $x^* \in \mathbb{R}^n$ es un punto tal que $\nabla f(x^*) = 0$ y $H(x^*)$ es invertible. Entonces para todo x_0 suficientemente cercano a x^*, el método de Newton está bien definido para todo k y converge a x^* con orden de convergencia al menos dos.

Teorema 2.2. Sea $\{x_k\}$ una sucesión generada por el método de Newton para minimizar la función $f(x)$. Si el Hessiano $H(x_k)$ es definido positivo y si $g_k = \nabla f(x_k) \neq 0$ entonces la dirección

$$d_k = -H(x_k)^{-1} g(x_k) = x_{k+1} - x_k$$

es una direccióon de descenso.

Notemos que aunque el método de Newton tiene propiedades de convergencia superiores si el punto inicial está cerca del punto solución, no se garantiza la convergencia si dicho punto inicial está lejos de él. Además, podría no ser un método de descenso, es decir,

$$f(x_{k+1}) > f(x_k).$$

Si la matriz Hessiana $\nabla^2 f(x)$ no está definida possitiva, la dirección de Newton d_k^N

$$\nabla^2 f(x_k) d_k^N = -\nabla f(x_k)$$

podría no ser una dirección de descenso.
Una alternativa para solucionar este problema es modificar el Hessiano,

$$B_k = \nabla^2 f(x_k) + E_k$$

tal que B_k sea definida positiva y la nueva dirección de búsqueda

$$B_k d_k = -\nabla f(x_k)$$

sea una dirección de descenso.
Se puede seleccionar simplemente,

$$E_k = \tau_k I \tag{1}$$

entonces

$$B_k = \nabla^2 f(x_k) + \tau_k I \tag{2}$$

con $\tau_k \geq 0$ asegura que B_k es lo suficientemente definida positiva.

Se podría seleccionar τ_k basado en el eigenvalor más pequeño de la matriz Hessiana $\nabla^2 f(x_k)$ pero no siempre es posible y es computacionalmente costo. Existen distintas maneras de elegir dicho parámetro, se presenta en el pesudocódigo una elección que se basa en la factorización de Cholesky.

2.3 Pseudocódigo

2.3.1 Método de Newton

Input: punto inicial x_0
Haga $alpha = \hat{\alpha}$
$inum = 0$
while $||\nabla f(x_k)|| \neq 0$ **do**
  Resolver $Dg(x_k)d_k = -g_k$
  Actualizar $x_{k+1} = x_k + d_k$
end while
Regresa $\alpha_k = \alpha$

2.3.2 Cholesky con Múltiplo de la Identidad

Input: Matriz A y tolerancia $\beta = 1e - 3$
Output: Factorización $A = LL^T$
if $min_i(a_{ii}) > 0$ **then**
  $\tau_0 = 0.$
else
  $\tau_0 = -\min_i(a_{ii}) + \beta$
end if
for $k = 0, 1, \ldots$ **do**
  Intenta factorizar $LL^T = A + \tau_k I$
  if La descomposición fue exitosa **then**
    Regresa L
  else
    $\tau_{k+1} = \max(2\tau_k, \beta)$
  end if
end for

2.3.3 Método de Newton Modificado

Input: Matriz A y tolerancia $\beta = 1e - 3$
Output: Mínimo x_k
k = 0
while $||\nabla f_k|| > \tau$ **do**
  Factorice $B_k = \nabla^2 f(x_k) + \tau I$ usando Cholesky para obtener L
  Resuelva hacia atrás $LL^T d_k = B_k d_k = -\nabla f(x_k)$
  Haga $x_{k+1} = x_k + \alpha_k d_k$
  $k = k + 1$
end while

2.4 Función de Rosembrock

$$f(x) = \sum_{i=1}^{N-1} [100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2]$$

donde $x = [x_1, \ldots, x_N]^T \in \mathbb{R}^N$.

Gradiente. Primero notemos que para $2 \leq j \leq N - 1$ se puede reescribir $f(x)$ como

$$\begin{aligned} f(x) &= \sum_{i=1}^{j-2}[100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2] \\ &\quad +[100(x_j - x_{j-1}^2)^2 + (1 - x_{j-1})^2] \\ &\quad [100(x_{j+1} - x_j^2)^2 + (1 - x_j)^2] \\ &\quad + \sum_{i=j+1}^{N-1}[100(x_{i-1} - x_i^2)^2 - (1 - x_i)^2]. \end{aligned}$$

Notemos que el primer y útlimo término son constantes con respecto a x_j, entonces se anularán al calcular la correspondiente derivada parcial.
Para D_1 se tiene que

$$\begin{aligned} \frac{\partial}{\partial x_1} f(x) &= 100(2)(x_2 - x_1^2)(-2x_1) - 2(1 - x_1) \\ &= -400(x_2 - x_1^2)x_1 - 2(1 - x_1). \end{aligned}$$

Para $D_j \in \{2, \ldots, N - 1\}$,

$$\begin{aligned} \frac{\partial}{\partial x_j} f(x) &= [200(x_j - x_{j-1}^2)] + [200(x_{j+1} - x_j^2)(-2x_j)] \\ &\quad -[2(1 - x_j)] \\ &= 200(x_j - x_{j-1}^2) - 400(x_{j+1} - x_j^2)x_j \\ &\quad -2(1 - x_j) \end{aligned}$$

Para D_N,

$$\frac{\partial}{\partial x_N} f(x) = 200(x_N - x_{N-1}^2).$$

De lo anterior tenemos que

$$\nabla f(x) = \begin{bmatrix} -400(x_2 - x_1^2)x_1 - 2(1 - x_1) \\ \vdots \\ 200(x_j - x_{j-1}^2) - 400(x_{j+1} - x_j^2)x_j - 2(1 - x_j) \\ \vdots \\ 200(x_N - x_{N-1}^2). \end{bmatrix}$$

Hessiano. Para calcular el Hessiano, se debe considerar como casos particulares, aquellos que se apliquen a $\dfrac{\partial f(x)(x)}{\partial x_1}$ y $\dfrac{\partial f(x)(x)}{\partial x_N}$ (como se observó en el gradiente).

Para $D_{x_k x_1}$,

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_1} = \begin{cases} 1200x_1^2 - 400x_2 + 2 & si \quad k = 1 \\ -400x_1 & si \quad k = 2 \\ 0 & e.o.c \end{cases}$$

Para $D_{x_k x_j}$ con $j \in \{2, \ldots, N-1\}$,

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_j} = \begin{cases} 1200x_j^2 - 400x_2 + 2 & si \quad k = j \\ -400x_j & si \quad k = j+1 \\ -400x_{j-1} & si \quad k = j-1 \\ 0 & e.o.c \end{cases}$$

Para $D_{x_k x_N}$,

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_1} = \begin{cases} -400x_{N-1} & si \quad k = N-1 \\ 200 & si \quad k = N \\ 0 & e.o.c \end{cases}$$

Así que el Hessiano tiene la forma

$$H(x) = \begin{pmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 & \ldots & 0 \\ -400x_1 & 1200x_2^2 - 400x_3 + 202 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & -400x_{N-1} & 200 \end{pmatrix}.$$

2.5 Función de Wood

Para $n = 4$ esta función está dada por

$$\begin{aligned} f(x) &= 100(x_1^2 - x_2)^2 + (x_1 - 1)^2 \\ &\quad +(x_3 - 1)^2 + 90(x_3^2 - x_4)^2 \\ &\quad +10{,}1[(x_2 - 1)^2 + (x_4 - 1)^2] \\ &\quad +19{,}8(x_2 - 1)(x_4 - 1). \end{aligned}$$

Gradiente. Derivando respecto a cada entrada obtenemos

$$\begin{aligned} \frac{\partial f(x)}{\partial x_1} &= 400(x_1^2 - x_2)x_1 + 2(x_1 - 1) \\ \frac{\partial f(x)}{\partial x_2} &= -200(x_1^2 - x_2) + 20{,}2(x_2 - 1) \\ &\quad +19{,}8(x_4 - 1) \\ \frac{\partial f(x)}{\partial x_3} &= 2(x_3 - 1) + 360(x_3^2 - x_4)x_3 \\ \frac{\partial f(x)}{\partial x_4} &= -180(x_3^2 - x_4) + 20{,}2(x_4 - 1) \\ &\quad +19{,}8(x_2 - 1). \end{aligned}$$

Hessiano

$$H_f(x) = \begin{bmatrix} 1200x_1^2 - 400x_1 + 2 & -400x_1 & 0 & 0 \\ -400x_1 & 220{,}2 & 0 & 19{,}8 \\ 0 & 0 & 1080x_3^2 - 360x_4 + 2 & -360x_3 \\ 0 & 19{,}8 & -360x_3 & 2020{,}2. \end{bmatrix}$$

3 Resultados

Para comparar los resultados de los métodos de Newton se utilizó el método de descenso gradiente con backtracking.

3.1 Función de Rosembrock

Se eligieron los parámetros

Parámetro	Valor
α	0.1
ρ	0.1
c_1	10^{-4}

3.2 Función de Wood

De los resultados de tareas previas se eligieron los parámetros

Parámetro	Valor
α	0.9
ρ	0.5
c_1	10^{-4}

El experimento se repitió 30 veces usando puntos iniciales aleatorios de la forma

$$x_0 = [x_0^1 + \eta_1, x_0^2 + \eta_2, \ldots, x_0^n + \eta_n]$$

donde $\eta_k \sim \mathcal{U}(-1, 1)$.

Los resultados para el promedio de tiempo fueron

Algoritmo	Máximo Descenso	Newton	Newton Modificado
Rosembrock	123.3166 segundos	12.3198 segundos	28.5644 segundos
Wood	6.7318 segundos	0.0160 segundos	0.0220 segundos

Los resultados para el promedio de iteraciones fueron

Algoritmo	Máximo Descenso	Newton	Newton Modificado
Rosembrock	10003	3599.7	7468.8
Wood	9479.46	102.03	98.13

4 Conclusiones

4.1 Máximo Descenso

Como se observó en tareas anteriores, aunque el algoritmo converge, la solución no siempre es el vector de $1's$. Para la función de Wood se observó un comportamiento similar.

4.2 Newton

En este método se mejoró considerablemente el desempeño del algoritmo tanto para tiempo como para número de iteraciones en ambas funciones a pesar que el cada iteración se debe resolver un sistema de ecuaciones.
Este método no garantiza la convergencia global, esto se observó en la función de Rosembrock, ya qu no siempre se llegaba al óptimo dado.

4.3 Newton Modificado

Este método intenta solucionar algunos de los problemas del método de Newton forzando a la matriz Hessiana a ser lo suficientemente definda positiva agregándole un múltiplo de la identidad. Se observó que la variación entre Newton y Newton Modificado es pequeña, aunque más grande. Esto es de esperarse debido a la robustés del método.