Optimización

Método de Máximo Descenso, Newton y Newton Modificado

Tarea 6

9 de marzo de 2021

Erika Rivadeneira Pérez
erika.rivadeneira@cimat.mx
Matemáticas Aplicadas - CIMAT

1. Resumen

En el presente reporte se comparan los promedios de iteraciones y de tiempo en 30 corridas al utilizar los algoritmos de máximo descenso, Newton y Newton modificado en las funciones de Rosembrock y de Wood.

2. Introducción

Como se ha mencionado en reportes anteriores el algoritmo de descenso por gradiente hace uso de una estrategia por búsqueda lineal, donde elige una dirección d_k y busca a lo largo de esta direccón desde la iteración actual x_k para una nueva iteración con un valor de función más bajo. Así hasta localizar un mínimo de la función.

En este reporte también se considerará la iteración de Newton para la búsqueda de mínimos, esta búsqueda está dada por

$$p_k^{\mathrm{N}} = -\nabla^2 f_k^{-1} \nabla f_k.$$

Como la matriz Hessiana $\nabla^2 f_k$ no siempre es positiva definida p_k^{N} no siempre será una dirección de descenso. Para solucionar este caso se modifica el método de Newton. Se discutirán los métodos de Newton y Newton modificados en la siguiente sección ya que en reportes previos se ha hablado sobre el método de máximo descenso.

3. Metodología

3.1. Método de Newton

Algoritmo

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

1. Resuelva $\mathrm{Dg}\left(\boldsymbol{x}_k\right) \boldsymbol{d}_k = -\boldsymbol{g}\left(\boldsymbol{x}_k\right)$
2. Actualice $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}_k$.

La derivada de $g = \nabla f(\cdot)$ en x es la matriz Jacobiana de $g = \nabla f(\cdot)$ en x, denotado como $J(x_k) = Dg(x_k)$, o el Hessiano de f, i.e., $H(x_k) = Df(x_k) = \nabla^2 f(x_k)$.

Ventajas

- Convergencia cuadrática de buenas suposiciones iniciales si $H(x^*) = \nabla^2 f(x^*)$ es no singular.
- Solución exacta en una iteración para un ∇f afine (exacta en cada iteración para cualquier funcion componente afin de ∇f), i.e., para un problema cuadrático.

Desventajas

- En general, no converge globalmente.
- Requiere el cómputo de $H(x_k) = \nabla^2 f(x_k)$ en cada iteración.
- En cada iteración requiere del cálculo de la solución de un sistema de ecuaciones lineales que podrían ser singulares o mal condicionadas.
- $d_k = -H(x_k)^{-1} g(x_k)$ puede no ser una dirección de descenso.

Para evitar estos inconvenientes se puede modificar el Hessiano, i.e.,

$$\mathbf{B}_k = \nabla^2 f(\boldsymbol{x}_k) + \mathbf{E}_k$$

tal que $\mathbf{B}_k \succ 0$ y la nueva dirección

$$\mathbf{B}_k \boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k)$$

es una dirección de descenso.

3.2. Búsqueda lineal de Newton con Modificación

Algoritmo

1. Factorizar la matriz $B_k = \nabla^2 f(x_k) + E_k$
2. Resolver $\mathbf{B}_k \boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k)$
3. Calcular α
4. Actualizar $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$

Notemos que $E_k = 0$ cuando $\nabla^2 f(x_k)$ es suficientemente definida positiva. De otra forma E_k fue elegida como aplicando la descomposición de Cholesky a la matriz resultante $B_k = \nabla^2 f(\boldsymbol{x}_k) + \tau_k \mathbf{I}$, donde $\tau_k \mathbf{I}$ es un múltiplo de la matriz identidad.

4. Resultados

Se realizaron 30 corridas de los algoritmos de Máximo descenso, Newton y Newton Modificado a las funciones

- Función de Rosembrock para $n = 100$

$$\begin{aligned}
f(\boldsymbol{x}) &= \sum_{i=1}^{n-1} \left[100\left(x_{i+1} - x_i^2\right)^2 + (1 - x_i)^2\right] \\
\nabla f(x) &= \left(-400 x_1 \left(x_2 - x_1^2\right) - 2(1 - x_1), \cdots, 200\left(x_k - x_{k-1}^2\right)\right. \\
&\quad \left. - 400 x_k \left(x_{k+1} - x_k^2\right) - 2(1 - x_k), \cdots, 200\left(x_n - x_{n-1}^2\right)\right) \\
\boldsymbol{x}^* &= [1, 1, \cdots, 1, 1]^T \\
f(\boldsymbol{x}^*) &= 0
\end{aligned}$$

con matriz Hessiana

$$\nabla^2 f(x) = \begin{bmatrix}
2 - 400\left(x_2 - 3x_1^2\right) & -400x_1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & & \ddots & & \vdots & \vdots & \vdots & \\
0 & 0 & \cdots & -400x_{k-1} & 202 - 400x_{k+1} + 1200x_k^2 & -400x_k & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & & \ddots & 0 & 0 & \cdots & -400x_{n-1} & 200
\end{bmatrix}$$

- Función de Wood

$$
\begin{aligned}
f(\boldsymbol{x}) =& 100\left(x_1^2 - x_2\right)^2 + (x_1 - 1)^2 + (x_3 - 1)^2 + 90\left(x_3^2 - x_4\right)^2 \\
&+ 10{,}1\left[(x_2 - 1)^2 + (x_4 - 1)^2\right] + 19{,}8\,(x_2 - 1)\,(x_4 - 1) \\
\nabla f(x) =& \big(400(x_1^2 - x_2)x_1 + 2(x_1 - 1), -200(x_1^2 - x_2) + 19{,}8(x_4 - 1) + 20{,}2(x_2 - 1), \\
& 2(x_3 - 1) + 360x_3(x_3^2 - x_4), -180(x_3^2 - x_4) + 20{,}2(x_4 - 1) + 19{,}8(x_2 - 1)\big) \\
\boldsymbol{x}^* =& [1, 1, 1, 1]^T \\
f\left(\boldsymbol{x}^*\right) =& 0,
\end{aligned}
$$

con matriz Hessiana

$$
\nabla^2 f(x) = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 & 0 & 0 \\ -400x_1 & 220{,}2 & 0 & 19{,}8 \\ 0 & 0 & 2 + 1080x_3^2 - 360x_4 & -360x_3 \\ 0 & 19{,}8 & -360x_3 & 200{,}2 \end{bmatrix}.
$$

La selección los puntos iniciales fue de modo que cada entrada se obtuviera de la siguiente forma

$$
x_i^0 = x_i^* + \eta
$$

donde x_i^* es la $i-$ésima entrada del óptimo x^* proporcionado para cada función, y $\eta \sim \mathcal{U}(-1, 1)$. Se hizo uso de estos datos para la implementación de los algoritmos mencionados y se calculó el promedio de iteraciones y de tiempo para cada uno, obteniendo

	Máximo Descenso	**Newton**	**Newton Modificado**
Función de Rosembrock	135.829	209.065	0.212
Función de Wood	1.236	0.0017	0.00386

Tabla 1: Promedio de tiempo en segundos de 30 corridas seleccionando puntos iniciales de forma aleatoria.

	Máximo Descenso	**Newton**	**Newton Modificado**
Función de Rosembrock	14219.8	399.7	15.6
Función de Wood	4514.3	8.6	13.6

Tabla 2: Promedio de iteraciones de 30 corridas seleccionando puntos iniciales de forma aleatoria.

5. Conclusiones y Discusión

Una observación importante es que se logró que el método de Newton modificado convergiera en un promedio de 15 iteraciones, para ambas funciones, en las 30 corridas sumando un múltiplo de la matriz identidad al Hessiano de las funciones con el algoritmo establecido en la diapositiva 19 de la clase 9. Se intentó sumarle el autovalor más pequeño del Hessiano pero fue muy costoso computacionalmente.

La implementación del método de Newton fue más sencilla que el de Newton modificado, pero este último fue el método más eficiente, en comparación al de máximo descenso y Newton, ya que el número de iteraciones y el tiempo de ejecución promedio de las corridas realizadas fue menor. Esto es evidente especialmente al usar el algoritmo con la función de Rosembrock ya que de 14000 iteraciones (con máximo descenso) desciende a aproximadamente 15 con el método de Newton modificado.

Referencias

[1] J. Nocedal and S. J. Wright. Numerical Optimization. Springer Series in Operation Research, 2000.