Optimización
Método de Máximo Descenso, Newton y Newton Modificado
Tarea 6

12 de abril 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 métodos de Región de Confianza basados en Dogleg y en el paso alternado entre Newton y Cauchy y Newton modificado en las funciones de Rosembrock, Wood y Branin.

2. Introducción

Se ha visto anteriormente métodos que utilizan estrategia por búsqueda lineal para localizar un mínimo de la función. En esta ocasión se considera el método de región de confianza para lograr este cometido, se restringe la minimización de la aproximación cuadrática a una vecindad de x^*.

Suponiendo que a partir de un punto x en el espacio n-dimensional, se quiere mejorar moviéndose a un punto donde la función sea menor, la idea es aproximar f mediante una función q, que razonablemente refleja el comportamiento de la función f en las cercanías, Δ, del punto x. Esta zona es la región de confianza. Por tanto, para un algoritmo basado en la región de confianza se necesita elegir una aproximación q definida en el punto actual x, elegir y modificar la región de confianza Δ y la precisión del subproblema de la región de confianza.

En este reporte también se considerará el método de Newton modificada para la búsqueda de mínimos, la cual se ha visto en reportes pasados y se promediará el tiempo e iteraciones con 30 corridas en cada función a considerar con los métodos de región de confianza y de Newton modificado.

3. Metodología

3.1. Método de Región de Confianza

El método de Región de Confianza define una región alrededor de un punto dado y luego lo ajusta y minimiza a partir de un modelo que representa bien el comportamiento de la función que pretendemos minimizar en el entorno del punto, por ejemplo, el desarrollo en serie de Taylor hasta segundas derivadas o una aproximación. Los métodos de Región de Confianza eligen simultáneamente la dirección de descenso y el tamaño de paso. Sea nuestro modelo a minimizar:

$$\min_{p\in\mathbb{R}^n} m_k(\mathrm{~d}) = f_k + \nabla f_k^T d + \frac{1}{2} d^T B_k \mathrm{~d} \operatorname{con} \|d\| \leq \Delta_k$$

Siendo Δ_k el radio de la región de confianza, d el paso de la región de confianza y $B_k \approx \nabla^2 f_k$, con B_k una matriz definida positiva. La calidad del modelo elegido se evalúa mediante la ganacia:

$$\rho_k = \frac{f(x_k) - f(x_k + d_k)}{m_k(0) - \mathrm{m}_k(d_k)}$$

Es decir, el numerador es la reducción real y el denominador la reducción esperada $(> 0) d_k$ el minimizador de m_k :

- Si ρ_k es negativo o cercano a 0 , el paso debe ser rechazado y es necesario reducir el radio de la región.
- Si ρ_k es cercano a 1 existe acuerdo entre el modelo y la función objetivo por lo que puede aumentarser el radio de la región de confianza para una próxima iteración.
- Si $0 < \rho_k < 1$ no altero el radio de la región.
- Si $\rho_k > 1$ acepto el radio d la región, con posibilidad de aumentarlo.

3.1.1. Basado en Dogleg

La determinación del subproblema del paso de ensayo puede resultar dificultosa en los casos en los que: $d_k^B = -B_k^{-1} g_k$ con $g_k \nabla f$ sea $\left\| d_k^B \right\| > \Delta_k$, donde B denota a la matriz Hessiana de f. Por tanto, debemos hallar una variable $\lambda \geq 0$ que imponga la restricción de nuestra región de confianza, esto es:

- $(B + \lambda I) d = -g$
- $\lambda(\Delta - \|d\|) = 0$

- $(B + \lambda I)$ sea semidefinida positiva, algo que siempre ocurre en problemas de minimos cuadrados. Es decir, debemos hallar λ tal que:

$$\|d_k(\lambda)\| = \left\|(B + \lambda I)^{-1} g_k\right\| = \Delta_k$$

Lo que resulta una ecuación no lineal no trivial. Para solucionar dicho problema, el algoritmo de dogleg aproxima la función $d_k(\lambda)$ por un poligono lineal de la siguiente forma:

$$d(\tau) = \begin{cases} \tau d^U & 0 \leq \tau \leq 1, \\ d^U + (\tau - 1)\left(d^B - d^U\right) & 1 \leq \tau \leq 2. \end{cases}$$

Siendo $d^U = -\frac{g^T g}{g^T B g} g$. El algoritmo para región de confianza basado en Dogleg es el siguiente

Región de confianza basado en Dogleg

Input: $x_0, G = gradiente, H = Hessiano, \eta_1, \eta_2, \hat{\eta_1}, \hat{\eta_2}, \Delta, \hat{\Delta}$

$x = x_0$
while $||G(x)|| \geq error$ **do**
  $p_k^b =$ Solve $(-H, G)$
  $p_k^u = (G^2/GHG)G$
  if $\|p_k^b\| <= \delta$
    $p = p_k^b$ paso de Newton
  else :
    if $\|p_k^u\| >= \delta$
      if $GHG <= 0$
        $\lambda = \delta/|G\|$
      else
        $\lambda = min(\|G\|^2/GHG, \delta/\|G\|)$
      $p = -\lambda G$
    else
      $\lambda =$ solución de $a\lambda^2 + b\lambda + c = 0$
      $\tau = \lambda + 1$
      if $\tau >= 0$ and $\tau <= 1$
        $p = \tau p_k^u$
      else
        $p = p_k^u + (\tau - 1)(p_k^b - p_k^u)$
  $\rho = (F(x) - F(x + p))/(m(0) - m(p))$
  if $\rho < \eta_1$
    $\Delta = \hat{\eta_1}\Delta$
  else if $\rho > \eta_2$ and $\|p\| = \Delta$:
    $\Delta = min(\hat{\eta_2}\Delta, \hat{\Delta})$
  if $\rho > \eta_1$
    $x = x + p$
  $i = i + 1$
Se actualizan los valores

Output: x^*, iteraciones

3.1.2. Newton-Cauchy alterno

El método de Newton-Cauchy Alterno, es muy similar que el método de Dogleg con la diferencia que para la toma del paso en cada iteración va alternando entre el método de paso de Newton y el método de paso de Cauchy.

Newton-Cauchy Alterno

Input: $x_0, G = gradiente, H = Hessiano,$ $\eta_1, \eta_2, \hat{\eta_1}, \hat{\eta_2}, \Delta, \hat{\Delta}$
$x = x_0$
while $||G(x)|| \geq error$ **do**
  $p_k^b =$ Solve $(H, -G)$
  if $GHG \leq 0$: Calcular τ
    $\tau = 1$
  else :
    $\tau = min(\|G\|^3/\Delta GHG, 1)$
  $p = \tau\Delta/\|G\|G$ (paso de Cauchy)
  if $\|p_k^b\| \leq \Delta$
    $p = p_k^b$ (paso de Newton)
  $\rho = (F(x) - F(x+p))/(m(0) - m(p))$
  if $\rho < \eta_1$
    $\Delta = \hat{\eta_1}\Delta$
  else if $\rho > \eta_2 and \|p\| = \Delta$
    $\Delta = min(\hat{\eta_2}\Delta, \hat{\Delta})$
  if $\rho > \eta_1$
    $x = x + p$
  $i = i + 1$
Se actualizan los valores
Output: x^*, iteraciones

3.2. Newton Modificado

Este método ya fue discutido en el reporte pasado. Su pseudocódigo se encuentra a continuación.

Newton modificado

Input: $x_0, G = gradiente, H = Hessiano$
$x = x_0$
while $||G(x)|| \leq error$ **do**
  $Hn =$ usar el método de Cholesky para volver H una matriz definidaa positiva
  $d =$ Solve $(H, -G)$
  $x = xd$
  Calcular $||G(x)||$
  $k = k + 1$
end while
Output: x^*, k

4. Resultados

Se realizaron 30 corridas de los métodos de Región de Confianza utilizando Dogleg y Newton-Cauchy alternado y Newton Modificado con las funciones

- Función de Rosembrock para $n = 100$

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

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

con matriz Hessiana

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

- Función de Wood

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

con matriz Hessiana

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

- Función de Branin

$$\begin{aligned}f(\boldsymbol{x}) &= a\left(x_2 - bx_1^2 + cx_1 - r\right)^2\\ &+ s(1 - t)\cos\left(x_1\right) + s\\ \nabla f(x) &= (2a(x_2 - bx_1^2 + cx_1 - r)(-2bx_1 + c)\\ &- s(1 - t)\sin(x_1), 2a(x_2 - bx_1^2 2 + cx_1 - r))\\ \boldsymbol{x}^* &= [\pi, 2{,}275]^T\\ f\left(\boldsymbol{x}^*\right) &= 0{,}397887\end{aligned}$$

con $a = 1, b = \dfrac{5{,}1}{4\pi^2}, c = \dfrac{5}{\pi}, r = 6, s = 10$ y $t = \dfrac{1}{8\pi}$

con matriz Hessiana

$$\nabla^2 f(x) = \begin{bmatrix} a_{11} & -a_{12} \\ a_{12} & a_{22} \end{bmatrix}.$$

donde $a_{11} = -4abx_2 + 12ab^2x_1^2 - 12abcx_1 + 2ac^2 + 4abr - s(1-t)\cos(x_1)$, $a_{12} = 4abx_1 + 2ac$ y $a_{22} = 2a$

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 $\eta \sim \mathcal{U}(-2, 2)$. Se consideraron dos criterios de paro, el error y el número máximo de iteraciones. Se usó una *toleracia* de 1×10^{-4} y un *número máximo de iteraciones* de 10000 en cada método. 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. Estos resultados se pueden visualizar en las tablas (1) y (2).

	Dogleg	**Newton-Cauchy Alterno**	**Newton Modificado**
Función de Rosembrock	1.62	2.07	1.04
Función de Wood	0.0057	0.0093	0.0081
Función de Branin	0.0026	0.0034	0.0036

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

	Máximo Descenso	**Newton**	**Newton Modificado**
Función de Rosembrock	746	885	80
Función de Wood	18.6	45.5	16.7
Función de Branin	3.7	4	4.87

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

5. Conclusiones y Discusión

El método de Dogleg parece ser rápido pero puede tener tres desventajas dependiendo del problema: la primera es que B^{-1} podría no existir, al igual que es posible que la matriz no sea definida positiva, lo que se requiere para garantizar convergencia. La última desventajas es que aún cumpliendo las condiciones anteriores, si nuestro problema es dimensiones altas, es posible que el tiempo de cálculo para p_k^B disminuya de forma considerable la velocidad del método, aún si sigue convergiendo en pocos pasos.

Finalmente, observando los resultados obtenidos de las tablas (1) y (2) se puede decir que el método de Newton modificado fue el más rápido en converger y con menos iteraciones que los métodos de Dogleg y Newton-Cauchy alternado. Por otro lado, el método de Dogleg parece ser más eficiente que el de Newton-Cauchy al probarlo con la función de Rosembrock, pero con funciones más simples como la de Wood o Branin el método de Dogleg parece ser el más eficiente de todos en cuanto a tiempo e iteraciones.

Referencias

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