

Centro de Investigación en Matemáticas

MAESTRÍA EN CIENCIAS COMPUTACIONALES Y MATEMÁTICAS
INDUSTRIALES

IMPLEMENTACIÓN DEL ARTÍCULO: “TWO EFFICIENT GRADIENT METHODS WITH APPROXIMATELY OPTIMAL STEPSIZES BASED ON REGULARIZATION MODELS FOR UNCONSTRAINED OPTIMIZATION”

Autor:

Antonio Daniel Reyes Ramírez
Ernesto Antonio Reyes Ramírez

Docente:
Dr. Oscar Dalmau

Mayo de 2023

1. Resumen

En el presente trabajo implementamos dos métodos de descenso de gradiente mediante la elección de un tamaño de paso aproximadamente óptimo. Si la función objetivo se puede aproximar mediante un modelo cuadrático en el segmento formado por dos puntos seguidos de la sucesión generada por el descenso de gradiente, entonces calculamos el tamaño de paso usando dicho modelo cuadrático. Por el contrario, si no se puede aproximar con un modelo cuadrático utilizamos un modelo de regularización, y con él calculamos el tamaño de paso. Para los modelos de regularización existen dos posibles elecciones, las cuales generan los dos métodos mencionados.

2. Introducción

Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función continuamente diferenciable. Consideremos el problema de optimización sin restricciones

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

El método de máximo descenso tiene la forma

$$x_{k+1} = x_k + \alpha_k d_k,$$

donde α_k es el tamaño de paso y d_k la dirección de descenso. A lo largo de este trabajo,

$$f_k := f(x_k), s_{k-1} := x_k - x_{k-1}, y_{k-1} := g_k - g_{k-1}.$$

En el algoritmo clásico se utiliza el negativo del gradiente $d_k = -g_k$, con $g_k := \nabla f(x_k)$, como dirección de descenso, y el tamaño de paso se escoge como

$$\alpha_k = \min_{\alpha} f(x_k + \alpha d_k).$$

A pesar de ser un algoritmo muy sencillo, su rendimiento es bajo debido a que converge linealmente y se ve muy afectado por un mal condicionamiento.

Para mejorar el rendimiento del método de descenso de gradiente se han realizado estudios sobre la elección del tamaño de paso α_k . Dos tamaños de paso muy utilizados son

$$\alpha_k^{BB_1} = \frac{\|s_{k-1}\|^2}{s_{k-1}^T y_{k-1}} \text{ y } \alpha_k^{BB_2} = \frac{s_{k-1}^T y_{k-1}}{\|y_{k-1}\|^2}.$$

Una de las estrategias ha sido calcular α_k mediante un modelo de aproximación de la función $f(x_k - \alpha g_k)$. De manera más precisa:

Definición 1. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función continuamente diferenciable, y sea $\phi(\alpha)$ un modelo de aproximación de la función $f(x_k - \alpha g_k)$. Una constante positiva α_k^{AOS} se conoce como tamaño de paso aproximadamente óptimo asociado a $\phi_k(\alpha)$ si α_k^{AOS} satisface

$$\alpha_k^{AOS} = \arg \min_{\alpha > 0} \phi_k(\alpha).$$

Para que el método de descenso de gradiente tenga un buen desempeño, se necesita escoger de manera adecuada el modelo de aproximación $\phi_k(\alpha)$, el cual debe capturar la información de la función f en una vecindad de x_k . En este reporte se implementan dos métodos de descenso de gradiente con tamaño de paso aproximadamente óptimo, los cuales son desarrollados en la sección 3.

3. Desarrollo matemático

La elección de un buen modelo de aproximación $\phi_k(\alpha)$ es de suma importancia ya que de este depende en su mayoría que obtengamos una buena aproximación al tamaño de paso α_k . Vamos a tomar en cuenta dos casos, cuando $f(x)$ se aproxima a un modelo cuadrático y cuando no. Claramente, si $f(x)$ se aproxima a un modelo cuadrático utilizaremos uno de este tipo para $\phi_k(\alpha)$, mientras que si no lo hace utilizaremos un modelo de regularización.

Dada la importancia de determinar el tipo de modelo que mejor aproxima a $f(x)$, utilizaremos los siguientes hechos útiles. Definimos,

$$\mu_k = \left| \frac{2(f_{k-1} - f_K + g_k^T s_{k-1})}{s_{k-1}^T y_{k-1}} - 1 \right|$$

Lo anterior, según (1), es una medida para determinar cuan cerca podría estar $f(x)$ de un modelo cuadrático en el segmento de línea entre x_{k-1} y x_k si se cumple la siguiente condición:

$$\mu_k \leq c_1 \quad \text{o} \quad \max\{\mu_k, \mu_{k-1}\} \leq c_2, \quad (2)$$

para c_1, c_2 números positivos pequeños y $c_1 < c_2$.

Con esto, procedamos por casos a determinar el tamaño de paso aproximadamente óptimo para métodos de gradiente.

Caso I

Supongamos que $s_{k-1}^T y_{k-1} > 0$ se satisface pero la condición (2) no lo hace. Dentro de este caso tenemos dos subcasos posibles:

p=3

Dado que la condición (2) no se cumple, puede que $f(x)$ no se aproxime bien con un modelo cuadrático en el segmento de línea de x_{k-1} y x_k , por lo que vamos a considerar el siguiente modelo de regularización:

$$\phi_{11}(\alpha) = f(x_k) - \alpha g_k^T g_k + \frac{1}{2} \alpha^2 g_k^T B_k g_k + \frac{1}{3} \alpha^3 \sigma_k(3) \|g_k\|^3. \quad (3)$$

Para generar la matriz B_K utilizaremos el método BFGS modificado,

$$B_k = D_k - \frac{D_k s_{k-1} s_{k-1}^T D_k}{s_{k-1}^T D_k s_{k-1}} + \frac{\overline{y_{k-1}} \cdot \overline{y_{k-1}}^T}{s_{k-1}^T \overline{y_{k-1}}} \quad (4)$$

donde $\overline{y_{k-1}} = y_{k-1} + \frac{\overline{r_k}}{\|s_{k-1}\|^2} s_{k-1}$, $\overline{r_k} = 3(g_k + g_{k-1})^T s_{k-1} + 6(f_{k-1} - f_k)$ y

$$D_k = \xi_0 \frac{y_{k-1}^T y_{k-1}}{s_{k-1}^T y_{k-1}}$$

una matriz escalar donde $\xi_0 \geq 1$. Para mejorar el cálculo numérico vamos a redefinir lo siguiente:

$$\overline{r_k} = \min\{\max\{\overline{r_k}, -\xi_1 s_{k-1}^T y_{k-1}\}, \xi_1 s_{k-1}^T y_{k-1}\},$$

con $0 < \xi_1 < 0.1$. Antes de proceder con la deducción del tamaño de paso aproximadamente óptimo enunciaremos el siguiente lema útil:

Lema 1. *Supongamos que $s_{k-1}^T y_{k-1} > 0$. Entonces, $s_{k-1}^T \overline{y_{k-1}} > 0$ y B_k es simétrica y positiva definida.*

La derivada de ϕ_{11} es la siguiente,

$$\frac{d\phi_{11}}{d\alpha} = -g_k^T g_k + \alpha g_k^T B_k g_k + \alpha^2 \sigma_k(3) \|g_k\|^3$$

Supongamos que $d\phi_{11}/d\alpha = 0$, de modo que igualando con la ecuación anterior:

$$-g_k^T g_k + \alpha g_k^T B_k g_k + \alpha^2 \sigma_k(3) \|g_k\|^3 = 0,$$

y

$$\alpha^2 \sigma_k(3) \|g_k\|^3 + \alpha g_k^T B_k g_k - g_k^T g_k = 0. \quad (5)$$

Notemos que la anterior es una ecuación cuadrática en α . Su discriminante es:

$$\begin{aligned}
\Delta &= (g_k^T B_k g_k)^2 - 4(\sigma_k(3)||g_k||^3)(-g_k^T g_k) \\
&= (g_k^T B_k g_k)^2 + 4(\sigma_k(3)||g_k||^3)(g_k^T g_k) \\
&= (g_k^T B_k g_k)^2 + 4(\sigma_k(3)||g_k||^3)||g_k||^2 \\
&= (g_k^T B_k g_k)^2 + 4\sigma_k(3)||g_k||^5
\end{aligned}$$

Por el lema (1) tenemos que B_k es positiva definida, así que $g_k^T B_k g_k > 0$. Además podemos suponer que $||g_k|| > 0$, ya que de lo contrario nuestro algoritmo habría convergido. De esta manera,

$$\Delta = (g_k^T B_k g_k)^2 + 12\sigma_k||g_k||^5 > 0$$

y así la ecuación (5) tiene dos soluciones reales y podemos resolverla mediante fórmula general. Por lo tanto, el tamaño de paso aproximadamente óptimo será,

$$\begin{aligned}
\alpha_k^{AOS(11)} &= \frac{-g_k^T B_k g_k + \sqrt{(g_k^T B_k g_k)^2 + 12\sigma_k||g_k||^5}}{2\sigma_k(3)||g_k||^3} \\
&= \left(\frac{-g_k^T B_k g_k + \sqrt{(g_k^T B_k g_k)^2 + 12\sigma_k||g_k||^5}}{2\sigma_k(3)||g_k||^3} \right) \left(\frac{\sqrt{(g_k^T B_k g_k)^2 + 12\sigma_k||g_k||^5} + g_k^T B_k g_k}{\sqrt{(g_k^T B_k g_k)^2 + 12\sigma_k||g_k||^5} + g_k^T B_k g_k} \right) \\
&= \frac{(g_k^T B_k g_k)^2 + 12\sigma_k||g_k||^5 - (g_k^T B_k g_k)^2}{2\sigma_k(3)||g_k||^3(\sqrt{(g_k^T B_k g_k)^2 + 12\sigma_k||g_k||^5} + g_k^T B_k g_k)} \\
&= \frac{12\sigma_k||g_k||^5}{2\sigma_k(3)||g_k||^3(\sqrt{(g_k^T B_k g_k)^2 + 12\sigma_k||g_k||^5} + g_k^T B_k g_k)} \\
&= \frac{2||g_k||^2}{\sqrt{(g_k^T B_k g_k)^2 + 12\sigma_k||g_k||^5} + g_k^T B_k g_k} \\
&= \frac{2||g_k||^2}{\sqrt{\Delta} + g_k^T B_k g_k}
\end{aligned}$$

Finalmente, tomamos la siguiente truncación:

$$\alpha_k^{AOS(11)} = \max\{\min\{\alpha_k^{AOS(11)}, \alpha_k^{BB_1}\}, \alpha_k^{BB_2}\} \quad (6)$$

para el método de gradiente.

p=4

Consideramos ahora el siguiente modelo de regularización,

$$\phi_{11}(\alpha) = f(x_k) - \alpha g_k^T g_k + \frac{1}{2}\alpha^2 g_k^T B_k g_k + \frac{1}{4}\alpha^4 \sigma_k(4)||g_k||^4. \quad (7)$$

Supongamos que $d\phi_{12}/d\alpha = 0$. Calculando la derivada obtenemos que,

$$-g_k^T g_k + \alpha g_k^T B_k g_k + \alpha^3 \sigma_k(4)||g_k||^4 = 0 \quad (8)$$

La anterior es una ecuación cúbica en α . Su discriminante viene dado por,

$$\Delta = \frac{1}{4\sigma_k(4)||g_k||^4} + \frac{(g_k^T B_k g_k)^3}{27\sigma_k(4)||g_k||^2} > 0$$

ya que B_k es positiva definida por el lema (1) y $\|g_k\| > 0$ ya que de lo contrario ya habríamos terminado el algoritmo. Por lo anterior, tenemos que la ecuación (13) tiene una solución real y dos complejas. De este modo el tamaño de paso aproximadamente óptimo será la solución real:

$$\alpha_k^{AOS(12)} = \sqrt[3]{\frac{1}{2\sigma_k(4)\|g_k\|^2} + \sqrt{\Delta}} + \sqrt[3]{\frac{1}{2\sigma_k(4)\|g_k\|^2} - \sqrt{\Delta}}$$

De manera similar al caso anterior, tenemos que una mejor aproximación numérica será la siguiente:

$$\alpha_k^{AOS(12)} = \max\{\min\{\alpha_k^{AOS(12)}, \alpha_k^{BB_1}\}, \alpha_k^{BB_2}\} \quad (9)$$

Parámetro de regularización

Para determinar el parámetro de regularización utilizamos la condición de interpolación la cual implica que,

$$\sigma_k(p) = \frac{p(f_{k-1} - f_k + g_k^T s_{k-1} - 1/2(s_{k-1}^T y_{k-1} + \bar{r}_k))}{\|s_{k-1}\|^p}$$

donde $p = 3$ o $p = 4$. Finalmente, para obtener mejoras numéricas realizamos la siguiente aproximación,

$$\sigma_k(p) = \max\{\min\{|\sigma_k(p)|, \sigma_{max}\}, \sigma_{min}\}, \quad (10)$$

donde $0 < \sigma_{min} < \sigma_{max}$ y $p = 3$ o $p = 4$.

Caso II

Supongamos que $s_{k-1}^T y_{k-1} > 0$ se satisface junto con la condición (2). Por las observaciones iniciales tenemos que $f(x)$ puede estar cerca de parecerse a un modelo cuadrático, por lo que vamos a aproximar por medio del siguiente:

$$\phi_2(\alpha) = f(x_k) - \alpha g_k^T g_k + \frac{1}{2} \alpha^2 g_k^T B_k g_k,$$

donde B_k se define como arriba con el algoritmo BFGS modificado. Calculemos la derivada de ϕ con respecto a α ,

$$\frac{d\phi_2}{d\alpha} = -g_k^T g_k + \alpha g_k^T B_k g_k.$$

Supongamos que $d\phi_2/d\alpha = 0$. Entonces,

$$-g_k^T g_k + \alpha g_k^T B_k g_k = 0$$

Despejando α obtenemos el tamaño de paso aproximadamente óptimo,

$$\alpha_k^{AOS(2)} = \frac{g_k^T g_k}{g_k^T B_k g_k}.$$

Finalmente, como la condición $s_{k-1}^T y_{k-1} > 0$ se satisface podemos definir la siguiente truncación,

$$\alpha_k^{AOS(2)} = \max\{\min\{\alpha_k^{AOS(2)}, \alpha_k^{BB_1}\}, \alpha_k^{BB_2}\}. \quad (11)$$

Caso III

Supongamos que $s_{k-1}^T y_{k-1} \leq 0$ y que se cumple la condición (12). Por definición de s_{k-1}, y_{k-1} tenemos que

$$\begin{aligned} 0 &\geq (x_k - x_{k-1})^T (g_k - g_{k-1}) \\ &= (x_{k-1} - \alpha_k g_{k-1} - x_{k-1})^T (g_k - g_{k-1}) \\ &= -\alpha_k g_{k-1}^T g_k + \alpha_k g_{k-1}^T g_{k-1}. \end{aligned}$$

Debido a que $\alpha_k > 0$ obtenemos que

$$\begin{aligned} \|g_{k-1}\|^2 &\leq g_{k-1}^T g_k \\ &\leq \|g_{k-1}\| \|g_k\| \\ \Rightarrow \|g_{k-1}\| &\leq \|g_k\|, \end{aligned}$$

donde la segunda desigualdad se sigue de la desigualdad de Cauchy-Schwarz. Por tanto,

$$0 < \frac{\|g_{k-1}\|}{\|g_k\|} \leq 1.$$

Ahora observemos que si se cumple la condición

$$\frac{\|g_{k-1}\|^2}{\|g_k\|^2} \geq \xi_2 \quad (12)$$

con $0 < \xi_2 < 1$ cercano a 1, entonces g_k y g_{k-1} tienen una tendencia a ser colineales, y los podemos considerar aproximadamente iguales. Usaremos la condición (12) para calcular modelos de regularización, y para ellos distinguimos dos casos: $p = 3, 4$.

(p=3)

Consideremos el modelo de regularización

$$\phi(\alpha) = f_k - \alpha g_k^T g_k + \frac{1}{2} \alpha^2 g_k^T \nabla^2 f(x_k) g_k + \frac{\sigma_k(3)}{3} \alpha^3 \|g_k\|^3.$$

Recordando que estamos suponiendo que se cumple la condición (12), tenemos que $g_k \approx g_{k-1}$ y entonces

$$\begin{aligned} g_{k-1}^T \nabla^2 f(x_k) g_{k-1} &\approx g_k^T \nabla^2 f(x_k) g_k \\ &\approx \frac{|(g(x_k + \alpha_{k-1} g_{k-1}) - g(x_k))^T g_{k-1}|}{\alpha_{k-1}} \\ &= \frac{|(g_{k-1} - g_k)^T g_{k-1}|}{\alpha_{k-1}}. \end{aligned}$$

Por otro lado, por definición de s_{k-1}, y_{k-1} se tiene que

$$\begin{aligned} s_{k-1}^T y_{k-1} &= (x_k - x_{k-1})^T (g_k - g_{k-1}) \\ &= (x_{k-1} - \alpha_k g_{k-1} - x_{k-1})^T (g_k - g_{k-1}) \\ &= \alpha_k g_{k-1}^T g_{k-1} - \alpha_k g_k^T g_{k-1} \\ &= \alpha_k (g_{k-1} - g_k)^T g_{k-1}. \end{aligned}$$

Sustituyendo esta igualdad en la expresión de arriba obtenemos que

$$g_{k-1}^T \nabla^2 f(x_k) g_{k-1} \approx \frac{|s_{k-1}^T y_{k-1}|}{\alpha_{k-1}^2}. \quad (13)$$

Usando la expresión anterior se tiene el siguiente modelo de aproximación:

$$\phi_{31}(\alpha) = f_k - \alpha g_k^T g_k + \frac{1}{2} \alpha^2 \frac{|s_{k-1}^T y_{k-1}|}{\alpha_{k-1}^2} + \frac{\sigma_k(3)}{3} \alpha^3 \|g_k\|^3. \quad (14)$$

Vamos a calcular un mínimo de $\phi_{31}(\alpha)$, y para ello calculamos su derivada y la igualamos a cero:

$$\frac{d\phi_{31}}{d\alpha} = -g_k^T g_k + \alpha \frac{|s_{k-1}^T y_{k-1}|}{\alpha_{k-1}^2} + \sigma_k(3) \alpha^2 \|g_k\|^3.$$

La única raíz real de la ecuación anterior está dada por

$$\alpha_k^{AOS(31)} = \frac{2\|g_k\|^2 \alpha_{k-1}^2}{\sqrt{|s_{k-1}^T y_{k-1}|^2 + 4\alpha_{k-1}^4 \sigma_k(3) \|g_k\|^5 + |s_{k-1}^T y_{k-1}|}}, \quad (15)$$

el cual es el tamaño de paso aproximadamente óptimo para este caso.

(p=4)

Ahora consideremos el modelo de regularización

$$\phi(\alpha) = f_k - \alpha g_k^T g_k + \frac{1}{2} \alpha^2 g_k^T \nabla^2 f(x_k) g_k + \frac{\sigma_k(4)}{4} \alpha^4 \|g_k\|^4.$$

Usando la identidad (13) obtenemos el modelo de aproximación

$$\phi_{32}(\alpha) = f_k - \alpha g_k^T g_k + \frac{1}{2} \alpha^2 \frac{|s_{k-1}^T y_{k-1}|}{\alpha_{k-1}^2} + \frac{\sigma_k(4)}{4} \alpha^4 \|g_k\|^4.$$

Calculemos el mínimo de la función ϕ_{32} , y para ello calculamos su derivada y la igualamos a cero:

$$\frac{d\phi_{32}}{d\alpha} = -\|g_k\|^2 + \alpha \frac{|s_{k-1}^T y_{k-1}|}{\alpha_{k-1}^2} + \sigma_k(4) \alpha^3 \|g_k\|^4 = 0.$$

El discriminante de la ecuación anterior es

$$\Delta_{32} = \frac{1}{4\sigma_k^2(4)\|g_k\|^4} + \frac{|s_{k-1}^T y_{k-1}|^3}{27\sigma_k^3(4)\alpha_{k-1}^6\|g_k\|^{12}}.$$

Mediante inspección y el hecho de que $\sigma_k(4) > 0$ podemos llegar a que $\Delta_{32} > 0$, y entonces la ecuación cúbica $d\phi_{32}/d\alpha = 0$ tiene una raíz real y dos imaginarias. El paso aproximadamente óptimo es la raíz real anterior, la cual se expresa como

$$\alpha_k^{AOS(32)} = \sqrt[3]{\frac{1}{2\sigma_k(4)\|g_k\|^2} + \sqrt{\Delta_{32}}} + \sqrt[3]{\frac{1}{2\sigma_k(4)\|g_k\|^2} - \sqrt{\Delta_{32}}}. \quad (16)$$

Parámetro de regularización

Ambos casos $p = 3, 4$ requieren de la elección de un parámetro de regularización $\sigma_k(p)$, y la condición de interpolación es

$$f_{k-1} = f_k - g_k^T s_{k-1} + \frac{1}{2} s_{k-1}^T y_{k-1} + \frac{\sigma_k(p)}{p} \|s_{k-1}\|^p,$$

y despejando $\sigma_k(p)$ llegamos a que

$$\sigma_k(p) = \frac{p(f_{k-1}f_k + g_k^T s_{k-1} - \frac{1}{2} s_{k-1}^T y_{k-1})}{\|s_{k-1}\|^p}.$$

Para evitar errores numéricos y hacer positivo este parámetro, realizamos la siguiente truncación:

$$\sigma_k(p) = \max\{\min\{|\sigma_k(p)|, \sigma_{max}\}, \sigma_{min}\}, \quad (17)$$

donde $0 < \sigma_{min} < \sigma_{max}$ son los mismos de la sección I.

Caso IV

Supongamos que $s_{k-1}^T y_{k-1} \leq 0$ y que no se satisface la condición (12). En este caso tomamos $\alpha_k = \xi_3 \alpha_{k-1}$, donde $\xi_3 > 0$. Luego, tomando $B_k = \frac{1}{\xi_3 \alpha_{k-1}} I$ obtenemos el modelo de aproximación cuadrático

$$\phi_4(\alpha) = f(x_k) - \alpha g_k^T g_k + \frac{1}{2} \alpha^2 g_k^T \left(\frac{1}{\xi_3 \alpha_{k-1}} I \right) g_k.$$

Para obtener el tamaño de paso aproximadamente óptimo calculamos la derivada e igualamos a cero:

$$\frac{d\phi_4}{d\alpha} = -\|g_k\|^2 + \alpha g_k^T \left(\frac{1}{\xi_3 \alpha_{k-1}} I \right) g_k = 0.$$

Despejando α llegamos a que

$$\alpha_k^{AOS(4)} = \xi_3 \alpha_{k-1} \quad (18)$$

3.1. Dos métodos de gradiente eficientes con tamaños de paso aproximadamente óptimos

Vamos a implementar el algoritmo de búsqueda en línea de Zhang-Hager junto con lo desarrollado en la sección anterior para obtener un algoritmo eficiente. El algoritmo se define con la ecuación:

$$f(x_k - \alpha g_k) \leq C_k - \delta \alpha \|g_k\|^2, \quad (19)$$

donde $0 < \delta < 1$,

$$Q_0 = 1, Q_{k+1} = \eta_k Q_k + 1, C_0 = f(x_0), C_{k+1} = (\eta_k Q_k C_k + f(x_{k+1}))/Q_{k+1}, 0 < \eta_k \leq 1. \quad (20)$$

Para obtener un mejor rendimiento numérico y obtener una buena convergencia tomamos η_k como:

$$\eta_k = \begin{cases} c, & \text{mod}(k, n) = n - 1 \\ 1 & \text{mod}(k, n) \neq n - 1 \end{cases} \quad (21)$$

donde $0 < c < 1$. Como resultado tenemos que el algoritmo sigue la siguiente estrategia:

$$\alpha = \begin{cases} \bar{\alpha}, & \alpha > 0, 1\alpha_k^0 \text{ y } \bar{\alpha} \in [0, 1\alpha_k^0, 0, 9\alpha] \\ 0,5\alpha & \text{otro caso} \end{cases} \quad (22)$$

donde α_0 es un tamaño de paso aproximadamente óptimo y $\bar{\alpha}$ es obtenido mediante una interpolación cuadrática en x_k y $x_k - \alpha g_k$. Finalmente a continuación presentamos los dos algoritmos completos.

Algorithm 1 GM_AOS (Reg p=3)

Step 0. Initialization. Given $x_0 \in R^n$, $\varepsilon > 0$, δ , c , c_1 , c_2 , α_{\max} , α_{\min} , α_0^0 , σ_{\min} , σ_{\max} , ξ_0 , ξ_1 , ξ_2 , ξ_3 . Set $Q_0 = 1$, $C_0 = f_0$ and $k = 0$.
Step 1. If $\|g_k\|_\infty \leq \varepsilon$, then stop.
Step 2. Compute approximately optimal stepsize.
 2.1 If $k = 0$, then set $\alpha = \alpha_0^{(0)}$ and go to Step 3.
 2.2 If $s_{k-1}^T y_{k-1} > 0$ holds and the condition (2) does not hold, then compute α_k by (6) and update σ_k by (10) with $p = 3$.
 2.3 If $s_{k-1}^T y_{k-1} > 0$ holds and the condition (2) holds, then compute α_k by (11).
 2.4 If $s_{k-1}^T y_{k-1} \leq 0$ holds and the condition (2) holds, then compute α_k by (15) and update σ_k by (17) with $p = 3$.
 2.5 If $s_{k-1}^T y_{k-1} \leq 0$ holds and the condition (2) does not hold, then compute α_k by (18).
 2.6 Set $\alpha_k^{(0)} = \max\{\min\{\alpha_k, \alpha_{\max}\}, \alpha_{\min}\}$ and $\alpha = \alpha_k^{(0)}$.
Step 3. Line search. If (19) holds, then go to Step 4, otherwise update α by (22) and go to Step 3.
Step 4. Update Q_{k+1} , C_{k+1} and η_k by (20) and (21).
Step 5. Set $\alpha_k = \alpha$, $x_{k+1} = x_k - \alpha_k g_k$, $k = k + 1$, and go to Step 1.

Algorithm 2 GM_AOS (Reg p=4)

Step 0. Initialization. Given $x_0 \in R^n$, $\varepsilon > 0$, δ , c , c_1 , c_2 , α_{\max} , α_{\min} , α_0^0 , σ_{\min} , σ_{\max} , ξ_0 , ξ_1 , ξ_2 , ξ_3 . Set $Q_0 = 1$, $C_0 = f_0$ and $k = 0$.
Step 1. If $\|g_k\|_\infty \leq \varepsilon$, then stop.
Step 2. Compute approximately optimal stepsize.
 2.1 If $k = 0$, then set $\alpha = \alpha_0^{(0)}$ and go to Step 3.
 2.2 If $s_{k-1}^T y_{k-1} > 0$ holds and the condition (2) does not hold, then compute α_k by (9) and update σ_k by (10) with $p = 4$.
 2.3 If $s_{k-1}^T y_{k-1} > 0$ holds and the condition (2) holds, then compute α_k by (11).
 2.4 If $s_{k-1}^T y_{k-1} \leq 0$ holds and the condition (2) holds, then compute α_k by (16) and update σ_k by (17) with $p = 4$.
 2.5 If $s_{k-1}^T y_{k-1} \leq 0$ holds and the condition (2) does not hold, then compute α_k by (18).
 2.6 Set $\alpha_k^{(0)} = \max\{\min\{\alpha_k, \alpha_{\max}\}, \alpha_{\min}\}$ and $\alpha = \alpha_k^{(0)}$.
Step 3. Line search. If (19) holds, then go to Step 4, otherwise update α by (22) and go to Step 3.
Step 4. Update Q_{k+1} , C_{k+1} and η_k by (20) and (21).
Step 5. Set $\alpha_k = \alpha$, $x_{k+1} = x_k - \alpha_k g_k$, $k = k + 1$, and go to Step 1.

4. Experimentos

Durante esta sección comparamos ambos métodos GM_{AOS} con $p = 3, 4$. Estos algoritmos fueron implementados en el lenguaje de programación Python.

Para probar ambos algoritmos utilizamos los siguientes parámetros: $\epsilon = 10^{-6}$, $\alpha_{min} = 10^{-30}$, $\alpha_{max} = 10^3$, $\xi_0 = 1,07$, $xi_1 = 5 \times 10^{-5}/3$, $\xi_2 = 0,8$, $\xi_3 = 5$, $\sigma_{min} = 10^{-30}$, $\sigma_{max} = 10^3$, $\delta = 10^{-4}$, $c_1 = 10^{-9}$, $c_2 = 10^{-7}$, $c = 0,99$, y

$$\alpha_0 = \begin{cases} 2 \frac{|f_0|}{\|g_0\|^2} & \text{si } \|x_0\|_\infty < 10^{-30} \text{ y } |f_0| \geq 10^{-30} \\ 1,0 & \text{si } \|x_0\|_\infty < 10^{-30} \text{ y } |f_0| < 10^{-30} \\ \min\{1,0, \max\{\frac{\|x_0\|_\infty}{\|g_0\|_\infty}, \frac{1}{\|g_0\|_\infty}\}\} & \text{si } \|x_0\|_\infty \geq 10^{-30} \text{ y } \|g_0\|_\infty \geq 10^7 \\ \min\{1,0, \frac{\|x_0\|_\infty}{\|g_0\|_\infty}\} & \text{si } \|x_0\|_\infty \geq 10^{-30} \text{ y } \|g_0\|_\infty < 10^7. \end{cases}$$

Se utilizó la interfaz PyCUTEst de CUTEst, el cual es un paquete de Fortran para realizar pruebas en software de optimización, y además fue el conjunto de funciones de prueba que se utilizó en el artículo original. Para poder reproducir los resultados de este proyecto se necesita instalar PyCUTEst, lo cual se puede hacer desde la página https://jfwkes.github.io/pycutest/_build/html/index.html. Para este proyecto no se utilizaron todas las funciones de prueba en PyCUTEst, sino que sólo tomamos tres de ellas pues con las que nuestros métodos pudieron converger. Estas funciones, en la nomenclatura de PyCUTEst son: ROSENBR, TOINTGSS, DENSCHNB. ROSENBR y TOINTGSS tienen dos variables, mientras que DENSCHNB tiene 5000. Los resultados con el método de $p = 3$ se muestran en la siguiente tabla.

Función	N. Iteraciones	Norma del gradiente
ROSENBR	8255	9.836662173690058e-06
TOINTGSS	1	9.836762893637634e-41
DENSCHNB	5383	5.153111352632671e-06

Los resultados con el método $p = 4$ aparecen a continuación:

Función	N. Iteraciones	Norma del gradiente
ROSENBR	8255	9.836662173690058e-06
TOINTGSS	1	9.836762893637634e-41
DENSCHNB	5383	5.153111352632671e-06

En las siguientes figuras podemos observar cómo va evolucionando la norma del gradiente en cada uno de los experimentos:

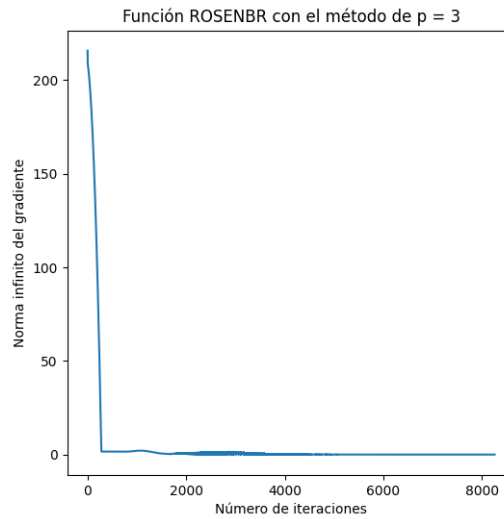


Figura 1: Función ROSENBR con el método $p = 3$

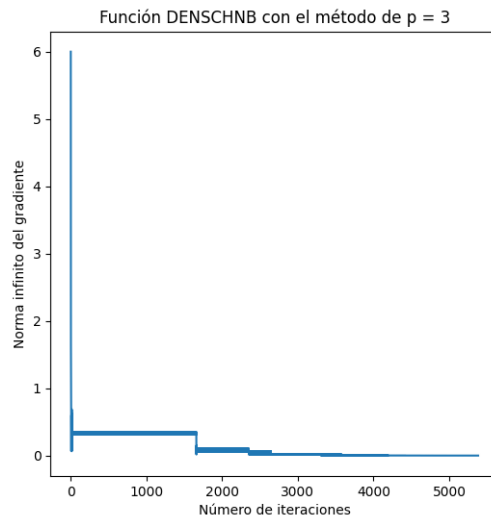


Figura 2: Función DENSCHNB con el método $p = 3$

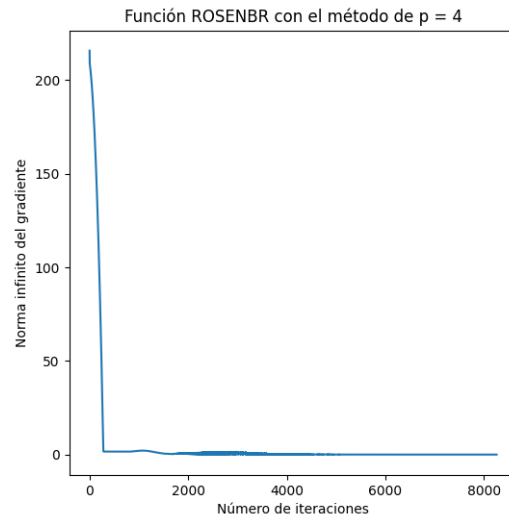


Figura 3: Función ROSENBR con el método $p = 4$

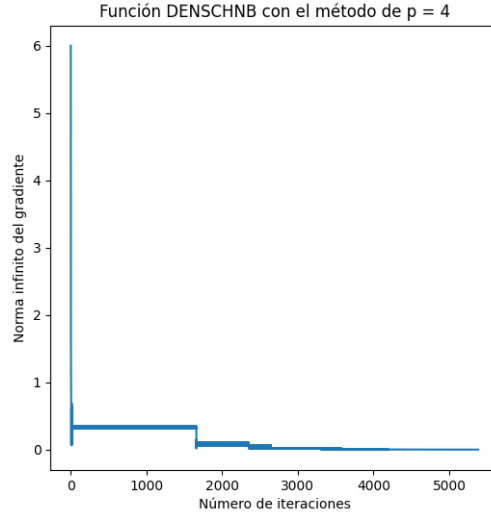


Figura 4: Función DENSCHNB con el método $p = 4$

5. Conclusiones

En este trabajo hemos estudiado e implementado los dos métodos de gradientes con tamaño de paso aproximadamente óptimo de manera satisfactoria y con resultados satisfactorios. Dichos algoritmos tienen las siguientes ventajas sobre otros como que: el tamaño de paso aproximadamente óptimo puede ser calculado fácilmente, la búsqueda de Armijo puede ser implementada de manera sencilla y que los métodos basados en dichos tamaños de paso funcionan muy bien para problemas donde las funciones tienen muchas variables.

Finalmente a la hora de realizar la implementación de los algoritmos no tuvimos que utilizar al algoritmo BFGS para calcular la aproximación del Hessiano ya que la librería de problemas de optimización que utilizamos ya tenía integrado el cálculo de la función objetivo, su gradiente y Hessiano.

Referencias

- [BB88] Jonathan Barzilai and Jonathan M Borwein. Two-point step size gradient methods. *IMA journal of numerical analysis*, 8(1):141–148, 1988.
- [DYY02] Yuhong Dai, Jinyun Yuan, and Ya-Xiang Yuan. Modified two-point stepsize gradient methods for unconstrained optimization. *Computational Optimization and Applications*, 22:103–109, 2002.
- [LCL19] Zexian Liu, Wangli Chu, and Hongwei Liu. Two efficient gradient methods with approximately optimal stepsizes based on regularization models for unconstrained optimization. *arXiv e-prints*, pages arXiv–1907, 2019.
- [LLD18] Zexian Liu, Hongwei Liu, and Xiaoliang Dong. An efficient gradient method with approximate optimal stepsize for the strictly convex quadratic minimization problem. *Optimization*, 67(3):427–440, 2018.
- [WN⁺99] Stephen Wright, Jorge Nocedal, et al. Numerical optimization. *Springer Science*, 35(67-68):7, 1999.
- [ZH04] Hongchao Zhang and William W Hager. A nonmonotone line search technique and its application to unconstrained optimization. *SIAM journal on Optimization*, 14(4):1043–1056, 2004.