Trabajo Investigativo de Optimización

Daniel Amaranto Mares Garcia
Grupo C-312

1. Modelo

$$\min_{(x,y)\in R^2} f(x,y) = -\tan(e^{-x^2}) - \tan(e^{-y^2})$$

2. Análisis Teórico de la función objetivo

Sea la función definida por:

$$f(x,y) = -\tan(e^{-x^2}) - \tan(e^{-y^2})$$

Las variables x y y pertenecen al conjunto de los números reales ($x, y \in R$). Dado que pueden tomar cualquier valor dentro de un intervalo continuo, ambas son variables continuas.

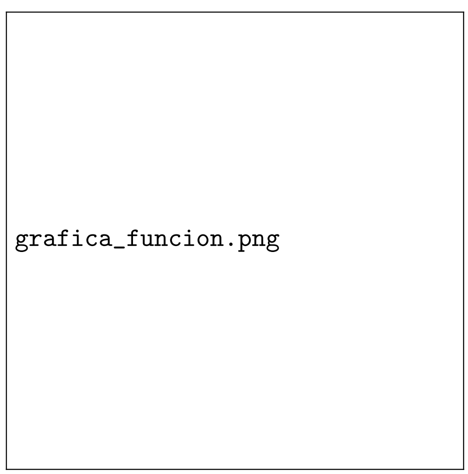

Figura 1: Gráfica de la función

La función $f(x,y)$ es continua en todo R^2 ya que las composiciones de funciones continuas como exp, tan (en su dominio $(0,1]$) son continuas.

Además, para todo $(x, y) \in R^2$ se tiene $e^{-x^2} \in (0, 1]$, $\tan(e^{-x^2}) \in (0, \tan(1)]$, por lo que $f(x, y) \geq -2\tan(1)$.

Una vez establecida la continuidad de la función y la naturaleza de sus variables, procedemos a analizar su comportamiento diferencial, mediante el cálculo del gradiente y de la matriz hessiana.

El gradiente de $f(x, y)$ está dado por:

$$\nabla f(x, y) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}$$

donde:

$$\frac{\partial f}{\partial x} = 2xe^{-x^2} \left(\tan^2(e^{-x^2}) + 1\right)$$

$$\frac{\partial f}{\partial y} = 2ye^{-y^2} \left(\tan^2(e^{-y^2}) + 1\right)$$

Por otra parte, la matriz Hessiana de $f(x, y)$ está compuesta por las segundas derivadas parciales:

$$H(f) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

donde:

$$\frac{\partial^2 f}{\partial x^2} = 2e^{-x^2}(\tan^2(e^{-x^2}) + 1) \left(1 - 2x^2(1 + 2e^{-x^2}\tan(e^{-x^2}))\right)$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 0$$

$$\frac{\partial^2 f}{\partial y^2} = 2e^{-y^2}(\tan^2(e^{-y^2}) + 1) \left(1 - 2y^2(1 + 2e^{-y^2}\tan(e^{-y^2}))\right)$$

3. Análisis de convexidad

Recordemos primeramente la definición de función convexa:

Definición 1. Sea $f : C \subseteq R^n \to R$ una función definida en un conjunto convexo C. Se dice que f es convexa si para todos $x_1, x_2 \in C$ y para todo $\lambda \in [0, 1]$ se cumple que:

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

Tomemos los puntos $x_1 = (0, 0)$, $x_2 = (10, 0)$, $\alpha = \frac{1}{2}$.

El punto convexo correspondiente es el punto medio
$\bar{x} = \alpha x_1 + (1 - \alpha)x_2 = \frac{1}{2}(0, 0) + \frac{1}{2}(10, 0) = (5, 0)$.

Calculemos los valores de f en estos puntos:

$f(x_1) = f(0,0) = -2\tan(1) \approx -3{,}1148,$
$f(x_2) = f(10,0) \approx -7{,}44 \times 10^{-44},$
$f(\bar{x}) = f(5,0) \approx -2{,}78 \times 10^{-11}.$
La condición de convexidad (para $\alpha = \frac{1}{2}$) exige
$f(\bar{x}) \leq \frac{1}{2}f(x_1) + \frac{1}{2}f(x_2).$
Pero el lado derecho es
$\frac{1}{2}f(x_1) + \frac{1}{2}f(x_2) \approx -1{,}5574,$
mientras que $f(\bar{x}) \approx -2{,}78 \times 10^{-11} > -1{,}5574$? Wait, no: -small ¿-large, yes -$10^{-11} > -3, sof(bar) > average, violation.$
Yes, f() ¿1$\frac{}{2f(x_1)+\frac{1}{2}f(x_2)}$,
y la desigualdad de convexidad falla para esos puntos. Concluimos que f no es convexa sobre el segmento que une x_1 y x_2 (y, por tanto, no es convexa globalmente) ya que existe esta violación de la desigualdad de convexidad.

Habiendo establecido que la función no es convexa mediante el contraejemplo anterior, podemos ahora examinar las implicaciones de este resultado en términos de su matriz Hessiana. Retomando la relación fundamental entre convexidad y la matriz Hessiana:

Lema 1. Sea $f : R^n \to R$ una función dos veces diferenciable en un conjunto convexo $C \subseteq R^n$, entonces, la matriz Hessiana $Hf(x) = \nabla^2 f(x)$ es semidefinida positiva para todo $x \in C$, si y solo si f es convexa en C.

Recíprocamente, si f no es convexa en C, entonces existe al menos un punto $x_0 \in C$ tal que $Hf(x_0)$ no es semidefinida positiva.

Luego, según el lema anterior, se concluye que:
$Hf(x,y)$ no es semidefinida positiva.

4. Determinación teórica del mínimo global de la función

Analicemos el comportamiento de la función en el punto (0, 0), comenzando por evaluar directamente su valor:
$f(0,0) = -2\tan(1) \approx -3{,}1148.$

Dado que la función es continua, acotada inferiormente por $-2\tan(1)$, y a medida que $\|(x,y)\| \to \infty$, $f(x,y) \to 0 > -2\tan(1)$, podemos afirmar que este valor representa un mínimo global de la función.

Ahora bien, para determinar si este mínimo es único, examinemos bajo qué condiciones se alcanza el mínimo. La igualdad $f(x,y) = -2\tan(1)$ ocurre si y sólo si $e^{-x^2} = 1$ y $e^{-y^2} = 1$, lo cual implica $(x,y) = (0,0)$. Esto nos lleva a concluir que el mínimo global es efectivamente único y se alcanza exclusivamente en el origen.

Para complementar este análisis, verifiquemos las condiciones de optimalidad mediante el cálculo de derivadas parciales. Un punto (x^*, y^*) es crítico cuando el gradiente se anula en él. Al calcular las derivadas parciales y evaluarlas en (0, 0), obtenemos:
$\frac{\partial f}{\partial x}(0,0) = 0,\ \frac{\partial f}{\partial y}(0,0) = 0,$

confirmando así que el origen satisface la condición necesaria de primer orden para ser un extremo local. Además, la matriz Hessiana en (0,0) es $(2(\tan^2(1) + 1), 2(\tan^2(1) + 1)) > 0$, positiva definida.

En resumen, mediante una combinación de evaluación directa, análisis de unicidad y verificación de condiciones de optimalidad, hemos establecido que la función alcanza su mínimo global único en (0, 0), con valor mínimo:

$f_{\text{mín}} = -2\tan(1)$.

Este resultado fundamental proporciona el marco teórico necesario para interpretar y evaluar el desempeño de los algoritmos de optimización aplicados a esta función.

5. Aproximación de la solución mediante algoritmos

En el contexto de la optimización numérica, la búsqueda de mínimos de funciones no lineales multivariables requiere el uso de algoritmos iterativos eficientes. Esta sección presenta dos métodos fundamentales para resolver problemas de optimización sin restricciones: el Método de Newton y el Método Quasi-Newton.

La selección de estos dos algoritmos se basa en las características de la función objetivo. La función es suave y dos veces diferenciable, con una dimensión baja (2 variables), lo que hace factible el cálculo de la Hessiana exacta para el Método de Newton. Además, presenta regiones planas lejos del mínimo donde el gradiente es muy pequeño, pero la información de curvatura (Hessiana) permite tomar pasos más grandes y efectivos. El Método Quasi-Newton se selecciona como una aproximación eficiente que evita el cálculo directo de la Hessiana, permitiendo una comparación con el método exacto.

Estos métodos se prefieren sobre el Método de Máximo Descenso (gradiente descendente) porque este último, siendo de primer orden, toma pasos pequeños en regiones planas, lo que resulta en convergencia lenta o estancamiento. Sobre el Método de Región de Confianza, aunque es útil para funciones no convexas, en experimentos numéricos se observó que tiende a detenerse prematuramente en regiones planas debido a gradientes pequeños, similar al comportamiento de métodos de primer orden.

El Método de Newton es un algoritmo iterativo de segundo orden para encontrar mínimos de una función diferenciable $f : R^n \to R$. Utiliza la aproximación cuadrática de la función alrededor del punto actual, resolviendo el sistema lineal dado por la Hessiana y el gradiente.

Dado un punto inicial $x^{(0)} \in R^n$, las iteraciones se definen como:

$$x^{(k+1)} = x^{(k)} - H_f(x^{(k)})^{-1} \nabla f(x^{(k)}), \quad k = 0, 1, 2, \ldots$$

donde $H_f(x^{(k)})$ es la matriz Hessiana en $x^{(k)}$.

El proceso se repite hasta que se cumple un criterio de convergencia, como $\|\nabla f(x^{(k)})\| < \epsilon$ o cambio pequeño en x.

Cerca del mínimo, si la Hessiana es positiva definida, exhibe convergencia cuadrática. Sin embargo, requiere el cálculo e inversión de la Hessiana en cada iteración, lo que es costoso en dimensiones altas, pero viable aquí.

El Método Quasi-Newton es una mejora del método de Newton que aproxima la inversa de la Hessiana sin calcular derivadas segundas explícitamente.

La actualización es:

$$x^{k+1} = x^k - B_k \nabla f(x^k)$$

donde $B_k \approx H_f(x^k)^{-1}$, actualizada usando la condición de secante $B_{k+1} y_k = s_k$, con $s_k = x^{k+1} - x^k$, $y_k = \nabla f(x^{k+1}) - \nabla f(x^k)$.

En la variante BFGS:

$$B_{k+1} = B_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{B_k y_k y_k^T B_k}{y_k^T B_k y_k}$$

Ofrece convergencia superlineal y es más eficiente cuando la Hessiana es cara de computar.

5.1. Observaciones sobre los experimentos realizados

Se implementaron los algoritmos utilizando la biblioteca SciPy en Python, minimizando la función desde diferentes puntos iniciales con tolerancia 10^{-8}.

Desde $x_0 = (1, 1)$, ambos convergen rápidamente al mínimo global $(0, 0)$ con valor $\approx -3{,}1148$: Quasi-Newton (BFGS) en 3 iteraciones (8 evaluaciones de función), Método de Newton (Newton-CG) en 4 iteraciones (8 evaluaciones).

Desde $x_0 = (5, 5)$, Newton-CG converge en 2 iteraciones, mientras que BFGS se detiene inmediatamente considerando el punto como crítico debido al gradiente muy pequeño ($\approx 10^{-10}$).

Desde $x_0 = (10, 10)$, Newton-CG falla por pérdida de precisión numérica (gradiente $\approx 10^{-43}$), y BFGS se detiene sin avanzar.

Estas observaciones destacan la robustez del Método de Newton en regiones planas gracias a la información exacta de curvatura, aunque sensible a Issues numéricos extremos. Quasi-Newton es eficiente cerca del mínimo pero menos efectivo en plateaus lejanos. En general, para esta función, se recomienda iniciar cerca del origen o usar técnicas de inicialización múltiples.