Análisis y experimentación numérica del modelo $f(x,y) = \arctan(x+y+1) - \arctan(-x-y+2) + x^2$

Nombre: Ramón Cherta González
Grupo: C - 311

1. Modelo a analizar

Definimos

$$f(x, y) = \arctan(x + y + 1) \; - \; \arctan(-x - y + 2) \; + \; x^2.$$

Observaciones iniciales

Las funciones arctan son definidas y suaves en todo $\mathbb{R}$. Por tanto,

$$f \in C^\infty(\mathbb{R}^2).$$

Los dos primeros sumandos están acotados (cada arctan toma valores en el intervalo $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$), mientras que el término x^2 crece sin cota cuando $|x| \to \infty$.

Definiciones auxiliares

Para simplificar la notación, introducimos las funciones

$$u(x, y) = x + y + 1, \qquad v(x, y) = -x - y + 2 = 2 - (x + y).$$

Con esta notación alternativa, la función del modelo puede escribirse como

$$f(x, y) = \arctan\big(u(x, y)\big) \; - \; \arctan\big(v(x, y)\big) \; + \; x^2.$$

2. Tipo de variables

Las variables del modelo, x e y, son continuas y toman valores en $\mathbb{R}$. Dado que no se imponen restricciones adicionales, el problema es *no restringido* y su dominio es todo el espacio $\mathbb{R}^2$.

Además, el modelo no incluye variables discretas ni una mezcla de tipos; todas las variables involucradas son estrictamente continuas.

3. Gradiente (derivadas parciales)

Recordemos que la derivada de la función $\arctan(t)$ es

$$\frac{1}{1 + t^2}.$$

Con las funciones auxiliares previamente definidas,

$$u(x, y) = x + y + 1, \qquad v(x, y) = 2 - (x + y),$$

introducimos las expresiones

$$a(x, y) = \frac{1}{1 + u(x, y)^2} \qquad \text{y} \qquad b(x, y) = \frac{1}{1 + v(x, y)^2}.$$

Entonces, el gradiente de f está dado por

$$\nabla f(x,y) = \begin{pmatrix} \dfrac{\partial f}{\partial x} \\[2ex] \dfrac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} a(x,y) + b(x,y) + 2x \\[2ex] a(x,y) + b(x,y) \end{pmatrix}.$$

Observación

Notamos que $a(x,y) > 0$ y $b(x,y) > 0$ para todo $(x,y) \in \mathbb{R}^2$, pues ambos denominadores son estrictamente positivos. En consecuencia, la componente

$$\frac{\partial f}{\partial y} = a(x,y) + b(x,y)$$

es estrictamente positiva en todo $\mathbb{R}^2$.

4. Hessiana (segundas derivadas)

Para calcular las segundas derivadas, recordemos primero que

$$\frac{d}{dt}\left(\frac{1}{1+t^2}\right) = -\frac{2t}{(1+t^2)^2}.$$

Aplicado a las funciones auxiliares

$$u(x,y) = x + y + 1, \qquad v(x,y) = 2 - (x+y),$$

definimos previamente

$$a(x,y) = \frac{1}{1+u(x,y)^2}, \qquad b(x,y) = \frac{1}{1+v(x,y)^2}.$$

Entonces, para cualquier variable $t \in \{x,y\}$,

$$\frac{\partial a}{\partial t} = -\frac{2\,u(x,y)}{(1+u(x,y)^2)^2} \cdot \frac{\partial u}{\partial t}, \qquad \frac{\partial b}{\partial t} = -\frac{2\,v(x,y)}{(1+v(x,y)^2)^2} \cdot \frac{\partial v}{\partial t}.$$

Como

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial y} = 1, \qquad \frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} = -1,$$

podemos derivar ahora cada componente de la Hessiana.

Componentes de la Hessiana

$$f_{xx}(x,y) = -\frac{2u}{(1+u^2)^2} + \frac{2v}{(1+v^2)^2} + 2,$$

$$f_{xy}(x,y) = -\frac{2u}{(1+u^2)^2} + \frac{2v}{(1+v^2)^2},$$

$$f_{yy}(x,y) = -\frac{2u}{(1+u^2)^2} + \frac{2v}{(1+v^2)^2}.$$

Matriz Hessiana

La matriz Hessiana de f viene dada por:

$$Hf(x,y) = \begin{pmatrix} 2 - \dfrac{2u}{(1+u^2)^2} + \dfrac{2v}{(1+v^2)^2} & -\dfrac{2u}{(1+u^2)^2} + \dfrac{2v}{(1+v^2)^2} \\ -\dfrac{2u}{(1+u^2)^2} + \dfrac{2v}{(1+v^2)^2} & -\dfrac{2u}{(1+u^2)^2} + \dfrac{2v}{(1+v^2)^2} \end{pmatrix}.$$

Nota

La estructura muestra que la fila 2 y la columna 2 son iguales (la última fila y columna comparten el mismo término), reflejando la asimetría introducida por x^2 solo en la componente xx.

5. Condiciones necesarias y suficientes (puntos críticos)

Un punto crítico interior debe satisfacer

$$\nabla f(x,y) = 0,$$

lo cual equivale a resolver el sistema

$$\begin{cases} a(x,y) + b(x,y) + 2x = 0, \\ a(x,y) + b(x,y) = 0. \end{cases}$$

Sin embargo, la segunda ecuación es imposible de satisfacer. En efecto, recordemos que

$$a(x,y) = \frac{1}{1+u(x,y)^2} > 0, \qquad b(x,y) = \frac{1}{1+v(x,y)^2} > 0$$

para todo $(x,y) \in \mathbb{R}^2$. Por lo tanto,

$$a(x,y) + b(x,y) > 0,$$

lo cual contradice la condición $a(x,y) + b(x,y) = 0$.

Consecuencia

No existe ningún punto $(x,y) \in \mathbb{R}^2$ tal que $\nabla f(x,y) = 0$. En consecuencia, la función no posee puntos críticos interiores y no es posible aplicar el criterio del Hessiano para clasificar puntos estacionarios, ya que dichos puntos simplemente no existen.

6) Existencia de extremos globales y locales

Analizamos primero el comportamiento de la función cuando las variables tienden a infinito.

Caso $y \to -\infty$ con x fijo. Sea

$$u = x + y + 1 \quad \Rightarrow \quad u \to -\infty \;\Rightarrow\; \arctan(u) \to -\frac{\pi}{2},$$

$$v = 2 - (x + y) \quad \Rightarrow \quad v \to +\infty \;\Rightarrow\; \arctan(v) \to +\frac{\pi}{2}.$$

Entonces,

$$\arctan(u) + \arctan(v) \to -\pi.$$

Por lo tanto,

$$\lim_{y\to-\infty} f(x,y) = -\pi + x^2.$$

El mínimo de x^2 es 0 (para $x = 0$), así que el límite inferior es

$$\inf_{(x,y)\in\mathbb{R}^2} f(x,y) = -\pi.$$

Caso $y \to +\infty$. En este caso,

$$\arctan(u) + \arctan(v) \to +\pi,$$

y debido al término x^2, se obtiene

$$f(x,y) \to +\infty.$$

Conclusión global. La función es no acotada superiormente y su ínfimo es

$$\inf f = -\pi,$$

pero dicho valor no se alcanza para ningún $(x,y) \in \mathbb{R}^2$. Para aproximarse a $-\pi$ es necesario tomar $y \to -\infty$ y $x \to 0$.

No existe mínimo global (el ínfimo no se alcanza) ni máximo global (la función diverge a $+\infty$ cuando $y \to +\infty$ o $|x| \to \infty$).

Extremos locales. Como no existen puntos críticos (no hay soluciones reales de $\nabla f = 0$), la función no presenta mínimos locales, máximos locales ni puntos de silla finitos.

Descripción de algoritmos (resumen, ventajas y limitaciones)

(a) Máximo descenso (Gradient Descent) con búsqueda de paso (backtracking)

Idea. En cada iteración,

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k),$$

donde α_k se elige mediante búsqueda de paso (backtracking Armijo) para asegurar disminución.

Ventajas:

- Simple de implementar.
- No requiere Hessiano.
- Control total sobre el tamaño de paso y los criterios de parada.

Desventajas:

- Convergencia lenta cerca de óptimos (si el problema tuviera uno).
- Sensible a la escala y al condicionamiento; puede requerir muchas iteraciones.
- En este problema (sin mínimo finito), empujará $y \to -\infty$ sin encontrar un punto estacionario — por lo que habría que añadir límites.

Cuándo usarlo en la práctica: como baseline para estudiar la dirección de descenso y comparar trayectorias.

(b) Newton (y variantes con Hessiano aproximado)

Idea. Se usa una aproximación cuadrática local:

$$x_{k+1} = x_k - \alpha_k H_f(x_k)^{-1} \nabla f(x_k),$$

con α_k determinado por damping o line search. Si se dispone del Hessiano, Newton ofrece convergencia cuadrática cerca de un óptimo.

Ventajas:

- Convergencia rápida si el Hessiano es calculable y es definido positivo localmente.
- Requiere menos iteraciones que Gradient Descent en problemas bien condicionados.

Desventajas:

- Computar y factorizar el Hessiano es costoso ($O(n^3)$), aunque en 2D no es un problema.
- Si el Hessiano no es definido positivo, el paso puede ser ascendente; requiere modificaciones (regularización o métodos de región de confianza).
- En este problema, al no existir puntos estacionarios, Newton tenderá a dar pasos grandes que deben ser regulados.

Variantes: Newton-CG (usa productos Hessiano–vector), métodos de región de confianza (`trust-ncg`, `trust-exact`), que protegen contra Hessianos indefinidos.

(c) Quasi-Newton (BFGS)

Idea. Aproximar la inversa del Hessiano mediante actualizaciones secuenciales (BFGS o L-BFGS en dimensiones grandes). El paso es

$$x_{k+1} = x_k - \alpha_k B_k \nabla f(x_k),$$

donde B_k aproxima H^{-1}.

Ventajas:

- Convergencia superlineal práctica sin necesidad de Hessiano explícito.
- Muy buen equilibrio entre costo computacional y rendimiento.
- BFGS es de los mejores métodos genéricos sin restricciones.

Desventajas:

- Si la función no tiene un óptimo finito, BFGS también puede empujar las variables sin límite.
- Sensible a ruido o no diferenciabilidad (no es el caso aquí).

Cuándo usarlo: recomendado como primer intento en `scipy.optimize.minimize` (`method = 'BFGS'`).

Comparación de algoritmos (a partir de los CSV)

BFGS — el mejor método

Convergencia: casi todas las corridas terminan con `success = True`.
Valor final:

$$f \approx -3{,}1415926,$$

extremadamente cerca del valor límite $-\pi$.
Solución encontrada:

$$x \approx 0, \qquad y \ll 0 \quad (\text{típicamente } -10^6 \text{ a } -10^8).$$

Iteraciones: entre 15 y 70, con gradientes muy pequeños.
Interpretación: BFGS captura adecuadamente la geometría de la función, aprovecha la curvatura y dirige el descenso hacia la región $s \to -\infty$. Se mantiene robusto incluso cuando una de las variables diverge numéricamente.

Gradient Descent (GD)

Convergencia: `success = True` pero siempre llega al límite de iteraciones (`nit = 200`).
Valor final: típicamente

$$f \approx -3{,}12,$$

lo cual está lejos del verdadero límite $-\pi$.
Gradiente: permanece lejos de cero al detenerse.
Interpretación: GD requiere muchos más pasos o un line search más sofisticado. Es muy sensible a la escala de la función (curvatura elevada inducida por el término x^2). No logra empujar $s \to -\infty$ dentro de las 200 iteraciones disponibles.

Newton

Convergencia: casi ningún éxito; `success = False` en prácticamente todas las corridas.
Iteraciones: siempre `nit = 200`.
Valor final: aproximadamente

$$f \approx -3{,}135,$$

mejor que GD pero inferior a BFGS.
Problemas: el Hessiano se vuelve mal condicionado o indefinido, generando pasos inestables.
Interpretación: Newton puro no es adecuado cuando:

- la curvatura cambia fuertemente,
- el "mínimo" está al infinito,
- el Hessiano presenta regiones indefinidas.

Sin damping ni métodos trust-region, Newton tiende a fallar numéricamente.

Conclusión final

Método	**Convergencia**	**Calidad de solución**	**Razón**
BFGS	Excelente	Excelente ($f \approx -\pi$)	Maneja bien la curvatura; avanza hacia $s \to -\infty$
GD	Pobre	No alcanza $-\pi$	Requiere más pasos; LR no adaptativo; curvatura difícil
Newton	Mala	Intermedia	Hessiano problemático; pasos inestables

Conclusión (en una frase). BFGS es claramente el método más robusto y preciso para este problema, porque maneja mejor la curvatura y aproxima eficazmente el ínfimo $-\pi$, mientras que GD converge demasiado lento y Newton falla numéricamente.