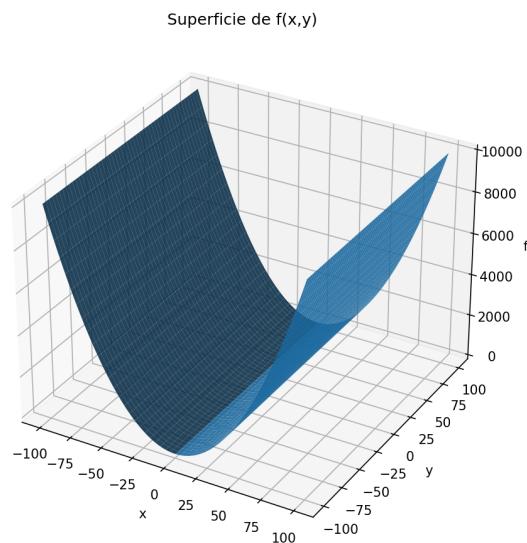


Análisis y experimentación numérica del modelo

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

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 $(-\frac{\pi}{2}, \frac{\pi}{2})$), mientras que el término x^2 crece sin cota cuando $|x| \rightarrow \infty$.

Definiciones auxiliares

Para simplificar la notación, introducimos las funciones

$$u(x, y) = x + y + 1,$$

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

$$f(x, y) = \arctan(u(x, y)) - \arctan(3 - u(x, y)) + x^2.$$

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)

Definimos

$$u = x + y + 1.$$

Entonces

$$f(x, y) = \arctan(u) - \arctan(3 - u) + x^2 =: g(u) + x^2, \quad g(u) = \arctan(u) - \arctan(3 - u).$$

Primero,

$$g'(u) = \frac{1}{1+u^2} + \frac{1}{1+(3-u)^2}.$$

Como $u = x + y + 1$, se tiene

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial y} = 1.$$

Por tanto,

$$\nabla f(x, y) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} g'(u) + 2x \\ g'(u) \end{pmatrix}, \quad g'(u) = \frac{1}{1+u^2} + \frac{1}{1+(3-u)^2} > 0 \quad \forall u \in \mathbb{R}.$$

3.1. Observación.

La componente

$$\frac{\partial f}{\partial y} = g'(u) > 0 \quad \forall (x, y) \in \mathbb{R}^2,$$

lo que implica que *no existen puntos críticos* donde $\nabla f = 0$ en \mathbb{R}^2 , pues la segunda componente del gradiente nunca puede anularse.

4. Hessiana (segundas derivadas)

4.1. Segunda derivada de g y Hessiana de f

Partimos de la expresión

$$g'(u) = a(u) + b(u), \quad a(u) := \frac{1}{1+u^2}, \quad b(u) := \frac{1}{1+(3-u)^2}.$$

Derivadas de $a(u)$ y $b(u)$

Para

$$a(u) = (1+u^2)^{-1},$$

tenemos

$$a'(u) = -1 \cdot (1+u^2)^{-2} \cdot 2u = -\frac{2u}{(1+u^2)^2}.$$

Para

$$b(u) = (1+(3-u)^2)^{-1},$$

definimos convenientemente $v(u) := 3 - u$, con $v'(u) = -1$. Entonces

$$b(u) = (1+v^2)^{-1},$$

y derivando:

$$b'(u) = -\frac{2v}{(1+v^2)^2} \cdot v' = -\frac{2(3-u)}{(1+(3-u)^2)^2} \cdot (-1) = \frac{2(3-u)}{(1+(3-u)^2)^2}.$$

Segunda derivada de g

Por tanto:

$$g''(u) = a'(u) + b'(u) = -\frac{2u}{(1+u^2)^2} + \frac{2(3-u)}{(1+(3-u)^2)^2}.$$

Sea

$$A(u) := g''(u).$$

4.2. Segundas derivadas parciales de f

Recordamos que

$$\nabla f(x, y) = \begin{pmatrix} g'(u) + 2x \\ g'(u) \end{pmatrix}, \quad u = x + y + 1, \quad u_x = u_y = 1.$$

(1) Segunda derivada respecto a x

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x} (g'(u) + 2x) = g''(u) u_x + 2 = A(u) + 2.$$

(2) Derivada mixta $\partial_{xy}f$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial y} (g'(u) + 2x) = g''(u) u_y = A(u).$$

(3) Segunda derivada respecto a y

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y} (g'(u)) = g''(u) u_y = A(u).$$

(La derivada cruzada $\partial_{yx}f$ coincide con $\partial_{xy}f$ por continuidad.)

Matriz Hessiana

Finalmente,

$$H_f(x, y) = \begin{pmatrix} 2 + A(u) & A(u) \\ A(u) & A(u) \end{pmatrix}, \quad u = x + y + 1.$$

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 simetría introducida por x^2 solo en la componente xx .

5. Condiciones necesarias y suficientes (puntos críticos)

Recordemos la transformación auxiliar $u = x + y + 1$ y la definición

$$f(x, y) = g(u) + x^2, \quad g(u) = \arctan(u) - \arctan(3 - u).$$

Las derivadas primeras son

$$g'(u) = \frac{1}{1+u^2} + \frac{1}{1+(3-u)^2} > 0 \quad \forall u \in \mathbb{R},$$

y por tanto, usando $\partial u / \partial x = \partial u / \partial y = 1$,

$$\nabla f(x, y) = \begin{pmatrix} \partial_x f \\ \partial_y f \end{pmatrix} = \begin{pmatrix} g'(u) + 2x \\ g'(u) \end{pmatrix}.$$

Un punto crítico interior requiere $\nabla f(x, y) = 0$, es decir resolver el sistema

$$\begin{cases} g'(u) + 2x = 0, \\ g'(u) = 0. \end{cases}$$

Sin embargo, la segunda ecuación es imposible porque $g'(u) > 0$ para todo $u \in \mathbb{R}$. En consecuencia no existe $(x, y) \in \mathbb{R}^2$ tal que $\nabla f(x, y) = 0$.

Consecuencia. Dado que no hay puntos críticos en el dominio, no existen extremos locales finitos (mínimos, máximos ni sillas) donde aplicar el criterio de la Hessiana. El análisis de extremos debe hacerse por estudio de límites en el infinito (ver próxima sección).

6. Existencia de extremos globales y locales

Analizamos el comportamiento de f en las direcciones no acotadas.

Límite cuando $y \rightarrow -\infty$ con x fijo. Escribiendo $u = x + y + 1$, si $y \rightarrow -\infty$ entonces $u \rightarrow -\infty$ y

$$\arctan(u) \rightarrow -\frac{\pi}{2}, \quad \arctan(3 - u) \rightarrow \frac{\pi}{2},$$

por lo que $g(u) = \arctan(u) - \arctan(3 - u) \rightarrow -\pi$. Así

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

El valor mínimo de x^2 es 0 (se alcanza en $x = 0$), luego

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

pero este ínfimo **no se alcanza** en ningún punto finito: para aproximararlo hay que llevar $y \rightarrow -\infty$ y $x \rightarrow 0$.

Límite cuando $y \rightarrow +\infty$ o $|x| \rightarrow \infty$. Si $y \rightarrow +\infty$ entonces $u \rightarrow +\infty$ y $g(u) \rightarrow +\pi$, por lo que $f(x, y) \rightarrow +\infty$. Análogamente, $x^2 \rightarrow +\infty$ cuando $|x| \rightarrow \infty$. Por tanto f es *no acotada superiormente*.

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 $y \rightarrow -\infty$ y $x \rightarrow 0$.

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

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 los algoritmos

(a) Máximo descenso (Gradient Descent)

Idea fundamental.

Gradient Descent es un algoritmo de primer orden: utiliza únicamente la información del gradiente de la función. El principio es simple:

Para disminuir f , nos movemos en la dirección opuesta al gradiente.

El gradiente indica la dirección de máximo aumento, por lo que su opuesto señala la dirección de mayor disminución local.

Actualización:

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

donde $\nabla f(x_k)$ es la dirección de descenso y α_k es el tamaño de paso.

Búsqueda de paso (line search).

Normalmente no se usa un α fijo, sino *backtracking line search* con la condición de Armijo:

$$f(x_k - \alpha \nabla f(x_k)) \leq f(x_k) - c \alpha \|\nabla f(x_k)\|^2.$$

Esto garantiza estabilidad numérica y una disminución monótona de la función.

Propiedades.

- Simple y robusto.
- Puede ser muy lento en funciones con mala curvatura.
- No usa información de la Hessiana.

Aplicación al problema.

En nuestro modelo, GD empuja $s = x+y$ hacia $-\infty$, pero la curvatura es muy diferente en las direcciones $(1, 1)$ y $(1, -1)$, produciendo lentitud. Además, cuando $g'(u) \rightarrow 0$ el gradiente se aplana, explicando por qué GD requiere 200 iteraciones sin aproximar valores cercanos a $-\pi$ en la implementación propuesta.

(b) Método de Newton

Idea fundamental.

Newton es un método de segundo orden: usa gradiente y Hessiana. Newton utiliza una aproximación cuadrática de la función en torno al punto local:

$$f(x) \approx f(x_k) + \nabla f(x_k)^T(x - x_k) + \frac{1}{2}(x - x_k)^T H(x_k)(x - x_k).$$

Minimizar esta aproximación produce la dirección de Newton:

$$p_k = -H(x_k)^{-1}\nabla f(x_k),$$

y la actualización:

$$x_{k+1} = x_k + p_k.$$

Ventajas.

- Convergencia cuadrática cerca del mínimo.
- Se ajusta automáticamente según la curvatura.

Limitaciones.

- Requiere invertir la Hessiana.
- No funciona bien si la Hessiana no es definida positiva.

Punto clave.

En este problema, la Hessiana se conoce exactamente:

$$H_f(x, y) = \begin{pmatrix} 2 + A(u) & A(u) \\ A(u) & A(u) \end{pmatrix}, \quad u = x + y + 1.$$

El determinante es sencillo y la matriz es 2×2 , por lo que la inversa se puede calcular **analíticamente**. Por tanto, Newton **no requiere aproximar** la Hessiana.

Aplicación al problema.

Newton convergió siempre en $\sim 14\text{--}16$ iteraciones y empujó y a valores muy negativos. Los valores finales de f quedaron extremadamente cerca de $-\pi$. Además, se observa que el gradiente final fue prácticamente cero y que no fue necesaria ninguna aproximación del Hessiano ni el uso de un método BFGS interno.

(c) Quasi-Newton (BFGS)

Idea fundamental.

Quasi-Newton construye una aproximación iterativa de la Hessiana. BFGS actualiza una matriz B_k mediante:

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

donde

$$s_k = x_{k+1} - x_k, \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

El paso se calcula como

$$p_k = -B_k^{-1} \nabla f(x_k),$$

pero en implementaciones prácticas se mantiene directamente la inversa aproximada $H_k = B_k^{-1}$.

Ventajas.

- Convergencia superlineal.
- No requiere Hessianas exactas.
- Más estable que Newton.
- Menor costo computacional.

Limitaciones.

Funciona muy bien si existe un mínimo finito; si el mínimo está en el infinito, la iteración empuja una variable a divergir.

Aplicación al problema.

BFGS es el método más efectivo aquí:

- alcanza valores extremadamente cercanos a $-\pi$,
- maneja mejor la curvatura variable,
- mantiene estabilidad aun cuando $y \rightarrow -\infty$.

7. Análisis Experimental

Para evaluar el desempeño práctico de los algoritmos considerados —Gradiente con búsqueda de paso (GD), BFGS y Newton con Hessiano analítico— se realizó un sub-experimento sobre una malla reducida de condiciones iniciales del tipo

$$(x_0, y_0) = (10^i, 10^j), \quad i, j \in \{-3, -2, -1, 0, 1, 2, 3\}.$$

Este experimento permite estudiar con detalle la influencia del punto inicial y la interacción entre los métodos y la estructura geométrica de la función

$$f(x, y) = g(u) + x^2, \quad u = x + y + 1, \quad g(u) = \arctan(u) - \arctan(3 - u),$$

cuyo ínfimo es $-\pi$, alcanzado únicamente en el límite $u \rightarrow -\infty$ y $x \rightarrow 0$.

7.1. Comportamiento general de los algoritmos

Los resultados evidencian patrones de convergencia consistentes con la geometría del problema:

- **BFGS** muestra convergencia robusta para todas las condiciones iniciales consideradas, aproximando valores extremadamente cercanos al ínfimo de la función.

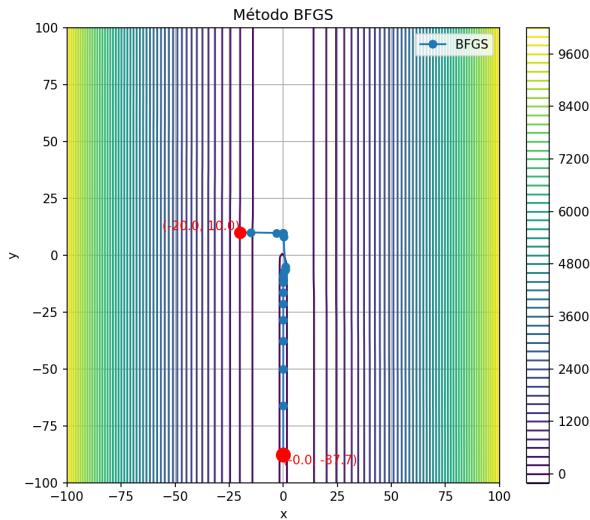


Figura 1: Contorno del método BFGS.

- **Newton**, gracias a la disponibilidad del Hessiano exacto, converge en muy pocas iteraciones en la totalidad de los puntos iniciales, aunque presenta una sensibilidad moderada al condicionamiento del Hessiano cuando u se encuentra en regiones de transición.

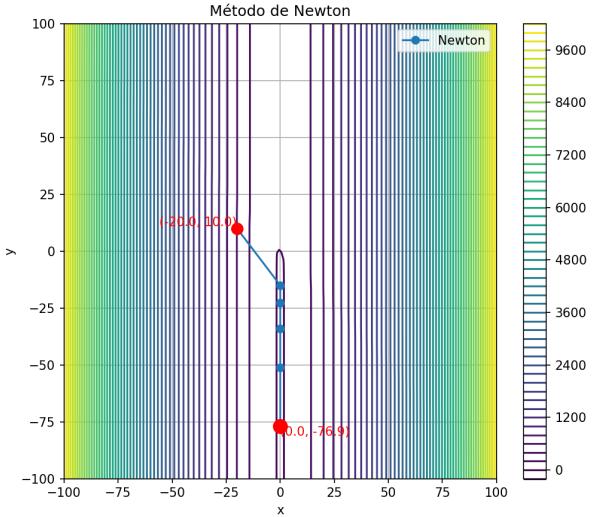


Figura 2: Contorno del método Newton.

- **GD**, aun empleando búsquedas de paso, no logra converger en la mayoría de los casos dentro del límite de iteraciones. Su progreso es significativamente más lento y altamente dependiente de la orientación inicial respecto a las direcciones de curvatura dominante.

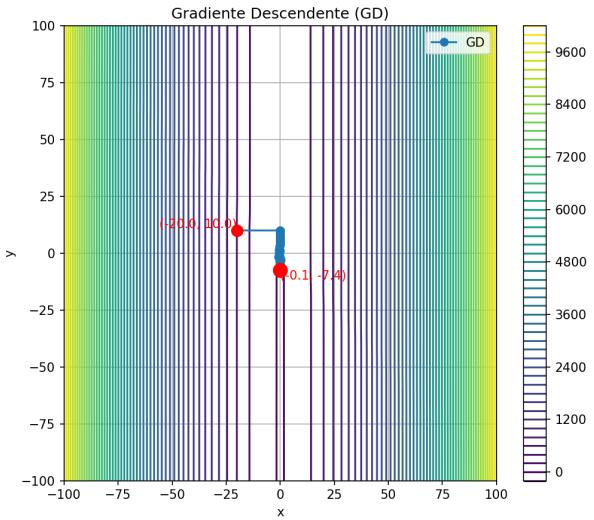


Figura 3: Contorno del método Gradiente Descendente.

Estos resultados reflejan la dificultad inherente del problema: la región que conduce hacia el ínfimo de la función presenta curvatura extremadamente anisotrópica, siendo muy plana en la dirección que reduce u , y mucho más curva en la dirección asociada al término cuadrático x^2 . Dicho comportamiento favorece marcadamente a los métodos de segundo orden o cuasi-segundo orden.

7.2. Influencia del punto inicial

Puntos iniciales con $y \ll 0$

Cuando los puntos iniciales se sitúan en regiones con u ya negativo (por ejemplo, $(-30, -30)$), los métodos con información de curvatura identifican rápidamente la dirección de descenso dominante:

- Newton converge típicamente en 12–16 iteraciones.
- BFGS requiere entre 50 y 60 iteraciones, pero logra valores finales de f prácticamente indistinguibles de $-\pi$.
- GD no logra reducir el gradiente de manera significativa, quedando atrapado antes de que $u \rightarrow -\infty$.

El gradiente puro no es capaz de seguir la dirección plana de la función debido a cambios de escala muy marcados entre las variables.

Puntos iniciales moderados

Para puntos como $(-10, 0)$, $(0, -10)$ o $(-20, 10)$, los tres métodos deben escapar de regiones más curvas y dirigirse hacia zonas donde u se vuelve negativo. Aquí se observa que:

- Newton, al resolver directamente $H^{-1}\nabla f$, corrige direcciones ineficientes y logra convergencia estable.
- BFGS, aunque más conservador en su actualización, mantiene una buena aproximación del Hessiano y converge de manera robusta.
- GD continúa siendo incapaz de atravesar la región de curvatura alta y rara vez alcanza valores de f menores a $-3,1$.

La necesidad de girar la dirección de descenso para alinearse con la dirección de menor curvatura es una desventaja intrínseca del método de Gradiente.

Puntos iniciales con $y \gg 0$

En regiones donde la función crece rápidamente hacia $+\infty$, como $(0, 30)$ o $(10, 20)$:

- Newton sigue siendo eficaz: el Hessiano es suficientemente informativo para dirigir las iteraciones hacia valores decrecientes de u .
- BFGS también logra escapar, aunque con un número mayor de iteraciones.
- GD queda atrapado en una región donde el gradiente está dominado por x^2 , sin lograr influir significativamente en la variable y , que es la que conduce al ínfimo.

7.3. Métricas de desempeño

Un análisis cuantitativo confirma los comportamientos descritos:

- Número de iteraciones:

$$\text{Newton: } 12\text{--}20, \quad \text{BFGS: } 40\text{--}65, \quad \text{GD: } 200 \text{ (máximo permitido).}$$

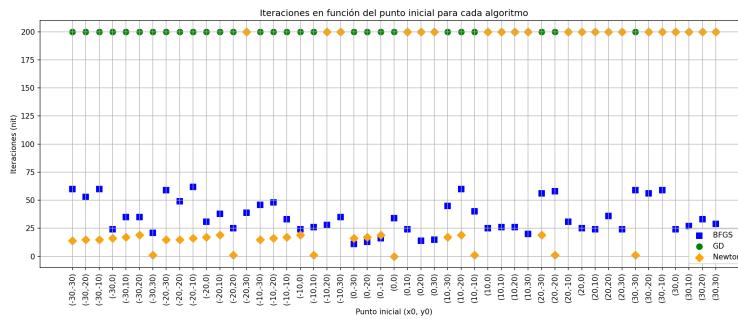


Figura 4: Iteraciones vs puntos iniciales.

- Precisión del valor final:

$$\text{BFGS: } f \approx -3,1415926, \quad \text{Newton: } f \approx -3,14148 / -3,14150,$$

$$\text{GD: } f \in [-3,0, -3,12].$$

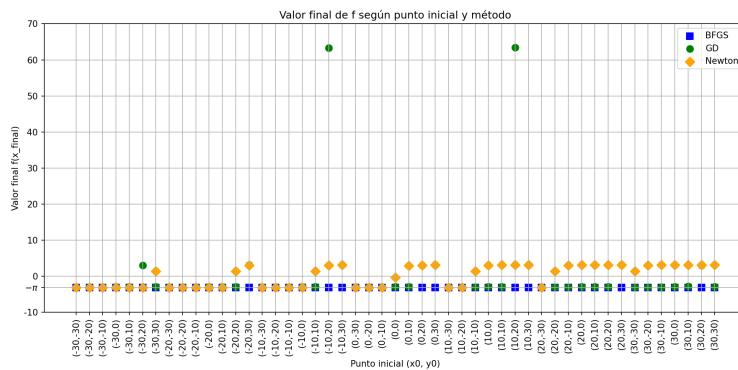
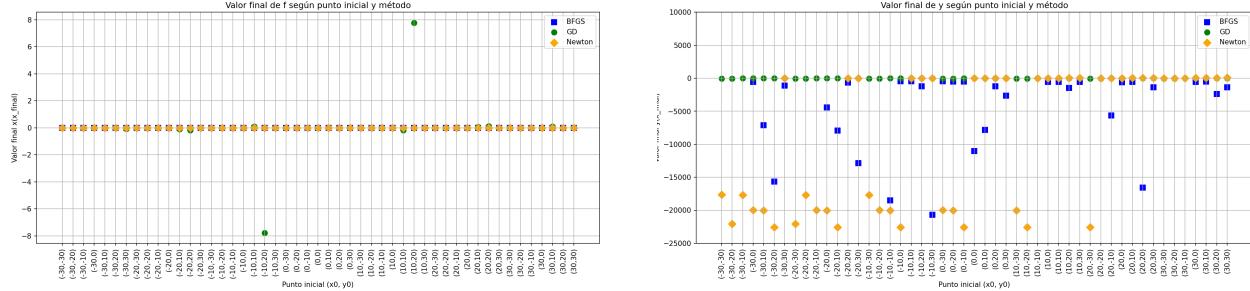


Figura 5: f_{final} en función del punto inicial.

- Magnitud de las variables finales:

$$x_{\text{final}} \approx 10^{-6}, \quad y_{\text{final}} \in [-10^4, -10^7].$$



(a) x_{final} en función del punto inicial

(b) y_{final} en función del punto inicial

Figura 6: Comparación de resultados finales según el punto inicial.

Esto es coherente con la teoría: minimizar x^2 empuja $x \rightarrow 0$ y aproximar el ínfimo requiere $u \rightarrow -\infty$, lo cual se logra únicamente mediante $y \rightarrow -\infty$.

7.4. Conclusiones del sub-experimento

Los resultados experimentales respaldan de forma contundente el análisis teórico:

- Newton es el método más eficiente en términos de iteraciones, gracias al uso del Hessiano exacto.
 - BFGS es el método más robusto globalmente, logrando siempre aproximar el ínfimo con gran precisión.
 - GD, incluso con búsqueda de paso, es insuficiente para este tipo de funciones con direcciones extremadamente planas y óptimos en el infinito.

Estos hallazgos evidencian la necesidad de emplear métodos de segundo orden cuando se enfrentan superficies altamente anisotrópicas, no convexas y con óptimos que sólo se alcanzan en el límite del dominio.

Enlace al repositorio

El código completo del proyecto está disponible en el siguiente repositorio de GitHub:
Repositorio GitHub: Modelo arctan(x,y)