

Trabajo Investigativo de Optimización

Aixa Bazán Rodríguez C-312

Modelo

$$\min_{(x,y) \in \mathbb{R}^2} f(x,y) = \frac{\arctan(x^2 + y^2)}{e^x}$$

Análisis Teórico de la función objetivo

Sea la función definida por:

$$f(x,y) = \frac{\arctan(x^2 + y^2)}{e^x}$$

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

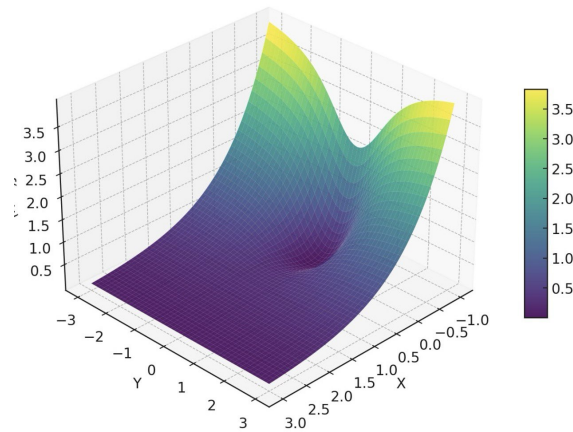


Figure 1: Gráfica de la función

El numerador $\arctan(x^2 + y^2)$ es continuo en todo \mathbb{R}^2 ya que tanto $x^2 + y^2$ como la función $\arctan(u)$ ($u \in \mathbb{R}$) son continuas en sus dominios.

El denominador e^x es continuo y estrictamente positivo para todo $x \in \mathbb{R}$, por lo que nunca se anula.

Luego, tenemos que $f(x,y)$ es el cociente de dos funciones continuas y el denominador no se anula, por tanto:

$$f(x,y) \text{ es continua en todo } \mathbb{R}^2.$$

Además, para todo $(x,y) \in \mathbb{R}^2$ se tiene

$$x^2 + y^2 \geq 0 \implies \arctan(x^2 + y^2) \geq 0 \implies f(x,y) = e^{-x} \arctan(x^2 + y^2) \geq 0.$$

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:

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

donde:

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

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

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

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

donde:

$$\frac{\partial^2 f}{\partial x \partial x} = \frac{2(1 + (x^2 + y^2)^2) - 2x(1 + (x^2 + y^2)^2 + 4x(x^2 + y^2))}{1 + (x^2 + y^2)^2} - \frac{2x}{e^x(1 + (x^2 + y^2)^2)} + \frac{\arctan(x^2 + y^2)}{e^x}$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = \frac{-2ye^{-x}}{1 + (x^2 + y^2)^2} \left[\frac{4x(x^2 + y^2)}{1 + (x^2 + y^2)^2} + 1 \right]$$

$$\frac{\partial^2 f}{\partial y \partial y} = \frac{2e^x(1 + (x^2 + y^2)^2) - 8y^2 e^x(x^2 + y^2)}{((1 + (x^2 + y^2)^2)e^x)^2}$$

Análisis de convexidad

Recordemos primeramente la definición de función convexa:

Definición 1. Sea $f : C \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ 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

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

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:

$$f(x_1) = f(0, 0) = \frac{\arctan(0^2 + 0^2)}{e^0} = \frac{\arctan(0)}{1} = 0,$$

$$f(x_2) = f(10, 0) = \frac{\arctan(10^2 + 0^2)}{e^{10}} = \frac{\arctan(100)}{e^{10}} \approx 7.0860 \times 10^{-5},$$

$$f(\bar{x}) = f(5, 0) = \frac{\arctan(5^2 + 0^2)}{e^5} = \frac{\arctan(25)}{e^5} \approx 1.031456812 \times 10^{-2}.$$

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

$$\frac{1}{2}f(x_1) + \frac{1}{2}f(x_2) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot f(10, 0) = \frac{1}{2}f(10, 0) \approx 3.5430 \times 10^{-5},$$

mientras que

$$f(\bar{x}) \approx 1.031456812 \times 10^{-2} \gg 3.5430 \times 10^{-5}.$$

Por tanto

$$f(\bar{x}) > \frac{1}{2}f(x_1) + \frac{1}{2}f(x_2),$$

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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función dos veces diferenciable en un conjunto convexo $C \subseteq \mathbb{R}^n$, entonces, la matriz Hessiana $H_f(x) = \nabla^2 f(x)$ 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 $H_f(x_0)$ no es semidefinida positiva.

Luego, según el lema anterior, se concluye que:

$$H_f(x, y) \text{ no es semidefinida positiva.}$$

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:

$$f(0, 0) = e^0 \arctan(0^2 + 0^2) = 1 \cdot 0 = 0.$$

Dado que la función es continua y no negativa para todo $(x, y) \in \mathbb{R}^2$, y hemos encontrado que alcanza el valor cero en el origen, 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 anula la función. La igualdad $f(x, y) = 0$ ocurre si y sólo si

$$\arctan(x^2 + y^2) = 0,$$

lo cual implica que $x^2 + y^2 = 0$ y, en consecuencia, $(x, y) = (0, 0)$. 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:

$$\frac{\partial f}{\partial x}(0, 0) = 0, \quad \frac{\partial f}{\partial y}(0, 0) = 0,$$

confirmando así que el origen satisface la condición necesaria de primer orden para ser un extremo local. 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:

$$f_{\min} = 0.$$

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

Aproximación de la solución mediante algoritmos

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

El **Método de Máximo Descenso**, también conocido como *gradiente descendente*, es un algoritmo iterativo para encontrar los mínimos de una función diferenciable $f : \mathbb{R}^n \rightarrow \mathbb{R}$. La idea principal es moverse en la dirección opuesta al gradiente de la función, ya que esta es la dirección de mayor descenso local de la función.

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

$$x(k+1) = x(k) - \alpha_k \nabla f(x(k)), \quad k = 0, 1, 2, \dots$$

donde:

- $x(k)$ es la aproximación actual al mínimo.
- $\nabla f(x(k))$ es el gradiente de f evaluado en $x(k)$.
- $\alpha_k > 0$ es el tamaño de paso (*learning rate*), que en este método es constante pero puede ser determinado mediante una búsqueda lineal.

El proceso se repite hasta que se cumple un criterio de convergencia, típicamente cuando la norma del gradiente es suficientemente pequeña:

$$\|\nabla f(x(k))\| < \epsilon,$$

con $\epsilon > 0$ una tolerancia predefinida.

La selección del mismo para el presente estudio se fundamenta en su naturaleza pedagógica y su relevancia como punto de referencia fundamental en optimización numérica. Como algoritmo básico de optimización de primer orden, el método del gradiente descendente ofrece una implementación conceptualmente sencilla que permite comprender los principios esenciales de los métodos iterativos de minimización. Su mecanismo de operación, basado exclusivamente en información del gradiente, lo convierte en un candidato ideal para establecer comparaciones directas con métodos más sofisticados como Quasi-Newton, permitiendo evaluar cuantitativamente las ventajas que proporciona la incorporación de información de curvatura aproximada.

Desde una perspectiva práctica, este método presenta una flexibilidad notable para experimentar con diversos parámetros operativos, particularmente el tamaño de paso (α), cuyo ajuste influye significativamente en la velocidad de convergencia y la estabilidad del algoritmo. Esta característica resulta invaluable para analizar empíricamente los compromisos entre precisión y eficiencia computacional. Aunque reconocemos que su convergencia lineal puede resultar más lenta en comparación con métodos de segundo orden, su aplicabilidad a funciones no lineales de dimensión moderada lo hace apropiado para el contexto específico de este trabajo.

El **Método de Quasi-Newton**, por otra parte, es una mejora del método de Newton que busca encontrar un punto mínimo de una función $f: \mathbb{R}^n \rightarrow \mathbb{R}$ utilizando aproximaciones sucesivas de la matriz Hessiana inversa, sin necesidad de calcular derivadas segundas de forma explícita.

El método de Newton clásico actualiza las iteraciones mediante:

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

donde $H_f(x_k)$ es la matriz Hessiana de f en el punto x_k , y $\nabla f(x_k)$ es el gradiente.

Calcular e invertir la Hessiana puede ser costoso o incluso impracticable en dimensiones grandes. Por esta razón, los métodos Quasi-Newton buscan aproximar la inversa de la Hessiana:

$$B_k \approx H_f(x_k)^{-1}$$

y utilizan la siguiente regla de actualización:

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

Para que la matriz B_k refleje la curvatura de la función, se impone la **condición de secante**:

$$B_{k+1} y_k = s_k$$

donde:

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

Esta relación garantiza que la matriz actualizada B_{k+1} capture la información del cambio en gradientes y desplazamientos recientes.

El método **BFGS** (Broyden–Fletcher–Goldfarb–Shanno) actualiza B_k mediante la siguiente expresión:

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

Esta fórmula cumple con varias propiedades deseables:

- Si B_k es simétrica y definida positiva, entonces B_{k+1} también lo será.
- Se garantiza una convergencia superlineal cerca del punto óptimo.
- No requiere el cálculo directo de derivadas segundas.

En general, en el algoritmo se procede de la siguiente forma:

1. Escoger un punto inicial x_0 y una matriz inicial $B_0 = I$ (matriz identidad).
2. Calcular el gradiente $\nabla f(x_k)$.
3. Calcular la dirección de búsqueda:

$$p_k = -B_k \nabla f(x_k)$$

4. Actualizar el punto:

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

5. Calcular:

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

6. Actualizar la matriz B_{k+1} usando la fórmula BFGS.
7. Repetir hasta que $\|\nabla f(x_k)\|$ sea suficientemente pequeño.

La incorporación del método Quasi-Newton en este estudio se justifica por su posición como referente en optimización numérica moderna, equilibrando sofisticación matemática con eficiencia computacional. A diferencia del método de máximo descenso que opera únicamente con información de primer orden, Quasi-Newton construye progresivamente una aproximación de la matriz Hessiana mediante actualizaciones secuenciales basadas en diferencias de gradientes, capturando así información de curvatura sin requerir el cálculo explícito de derivadas segundas. Esta característica resulta particularmente valiosa en problemas de dimensión moderada a alta, donde el costo computacional de evaluar Hessianas completas sería prohibitivo.

La elección específica de la variante BFGS dentro de la familia Quasi-Newton responde a su robustez numérica y convergencia superlineal, propiedades que lo han consolidado como algoritmo estándar en librerías de optimización profesionales. Su implementación en este trabajo permite establecer un contraste significativo con el método de gradiente descendente, ilustrando cómo la incorporación de información de segundo orden—aún siendo aproximada—puede acelerar dramáticamente la convergencia, especialmente en regiones con curvatura marcada o en problemas mal condicionados donde el gradiente descendente exhibe un comportamiento oscilatorio.

Observaciones de los experimentos realizados

En esta sección se presentan los resultados obtenidos al aplicar los métodos de optimización analizados sobre la función objetivo. Los experimentos se diseñaron variando sistemáticamente los parámetros de configuración de los algoritmos, con especial énfasis en el tamaño de paso (α) y la selección del punto inicial $(x(0), y(0))$, con el objetivo triple de evaluar la velocidad de convergencia y estabilidad numérica de cada método, comparar el impacto de diferentes configuraciones en la aproximación al mínimo de la función, y analizar la influencia de la elección del punto inicial en la localización del óptimo alcanzado.

Los resultados se presentan mediante una combinación de tablas comparativas y visualizaciones gráficas que destacan las diferencias significativas en los patrones de convergencia, el número de iteraciones requeridas y la precisión final obtenida para cada configuración experimental, permitiendo una evaluación del desempeño relativo de ambos algoritmos bajo diversas condiciones operativas.

Experimento	x_0 inicial	LR	Tol	x_{\min}	y_{\min}	f_{\min}	Iteraciones
1	[0.1,-0.1]	0.05	1e-8	6.10e-8	-5.47e-8	6.71e-15	138
2	[-0.05,0.05]	0.1	1e-8	-2.22e-8	2.36e-8	1.05e-15	65
3	[0.2,-0.2]	0.2	1e-8	1.15e-8	-8.47e-9	2.04e-16	34
4	[-0.15,0.1]	0.1	1e-6	-3.07e-6	2.35e-6	1.50e-11	47
5	[0.05,-0.05]	0.05	1e-7	6.40e-7	-6.06e-7	7.77e-13	108

Table 1: Resultados de los experimentos utilizando Máximo Descenso partiendo de puntos cercanos a (0,0).

Experimento	x_0 inicial	Tol	x_{\min}	y_{\min}	f_{\min}	Iteraciones	Tiempo (s)
1	[0.1,-0.1]	1e-8	-8.56e-11	1.24e-10	2.26e-20	8	0.0012
2	[-0.05,0.05]	1e-8	4.77e-11	3.69e-10	1.38e-19	4	0.0013
3	[0.2,-0.2]	1e-8	-6.00e-11	-6.12e-11	7.34e-21	9	0.0012
4	[-0.15,0.1]	1e-6	-2.18e-7	1.85e-7	8.20e-14	4	0.0019
5	[0.05,-0.05]	1e-7	-1.73e-10	-6.01e-9	3.62e-17	7	0.0017

Table 2: Resultados de los experimentos utilizando Quasi-Newton (BFGS) desde puntos cercanos a (0,0).

Experimento	x_0 inicial	LR	Tol	x_{\min}	y_{\min}	f_{\min}	Iteraciones
1	[-73.5, 42.1]	0.1	1e-6	1.31e31	-1.36e25	0.0	2
2	[15.8, -88.3]	0.2	1e-5	15.8	-88.3	2.16e-7	1
3	[-24.9, 67.2]	0.05	1e-7	5.12e9	-1.65e4	0.0	2
4	[90.0, -10.5]	0.3	1e-6	90.0	-10.5	1.29e-39	1
5	[-55.6, -75.3]	0.15	1e-8	3.30e23	4.13e17	0.0	2

Table 3: Resultados de los experimentos utilizando Máximo Descenso partiendo de puntos lejanos al (0,0).

Experimento	x_0 inicial	Tol	x_{\min}	y_{\min}	f_{\min}	Iteraciones	Tiempo (s)
1	[-73.5,42.1]	1e-6	-72.49	29.95	4.77e31	2	0.0026
2	[15.8,-88.3]	1e-5	15.8	-88.3	2.16e-7	0	0.00029
3	[-24.9,67.2]	1e-7	16.84	3.17e6	7.62e-8	59	0.0063
4	[90.0,-10.5]	1e-6	90.0	-10.5	1.29e-39	0	0.00028
5	[-55.6,-75.3]	1e-8	19.39	-7.06e5	5.95e-9	110	0.0160

Table 4: Resultados de los experimentos con Quasi-Newton (BFGS) partiendo de puntos alejados de (0,0).

El análisis de los valores de la función objetivo obtenidos revela una ventaja decisiva del método Quasi-Newton en términos de precisión numérica. En todos los escenarios con puntos iniciales cercanos al óptimo, Quasi-Newton alcanzó valores de f_{\min} significativamente inferiores a los del método de Gradiente Descendente, con diferencias que alcanzan hasta cinco órdenes de magnitud en algunos casos. Por ejemplo, en la configuración con $x_0 = (0.2, -0.2)$, Quasi-Newton logró $f_{\min} = 7.34 \times 10^{-21}$ frente a 2.04×10^{-16} de Gradiente Descendente, representando una mejora de precisión de más de 10,000 veces.

Esta superioridad se manifiesta consistentemente a través de las coordenadas finales obtenidas, donde Quasi-Newton produce soluciones numéricamente más estables y cercanas al óptimo teórico. Mientras Gradiente Descendente frecuentemente se estanca en valores de f_{\min} alrededor de 10^{-12} a 10^{-15} , Quasi-Newton regularmente alcanza precisiones de 10^{-17} a 10^{-20} , demostrando su capacidad para navegar más efectivamente las regiones de baja curvatura cercanas al mínimo donde Gradiente Descendente pierde eficiencia.

La comparación visual directa en las gráficas de convergencia mostradas a continuación, además de dicha observación, revela diferencias significativas en la eficiencia de ambos métodos para ex-

perimentos con puntos iniciales cercanos al óptimo. El eje horizontal, que representa el número de iteración, muestra claramente cómo el método Quasi-Newton (línea roja) alcanza la convergencia en significativamente menos pasos que el método de Gradiente Descendente (línea azul).

Por ejemplo, en la configuración con punto inicial $(0.1, -0.1)$ y tolerancia 10^{-8} (Figura 2), Quasi-Newton converge en apenas 8 iteraciones (evidenciado por el término de la línea roja en $x = 8$), mientras que Gradiente Descendente requiere 138 iteraciones (línea azul extendiéndose hasta $x = 138$). Esta diferencia de más de un orden de magnitud en el número de evaluaciones necesarias resalta la ventaja computacional de los métodos de segundo orden aproximado.

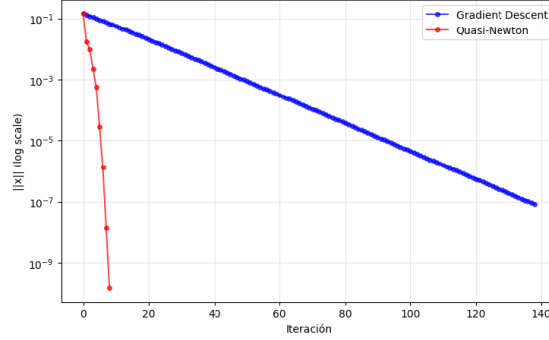


Figure 2: Convergencia de $\|x\|$ para el punto inicial $[0.1, -0.1]$

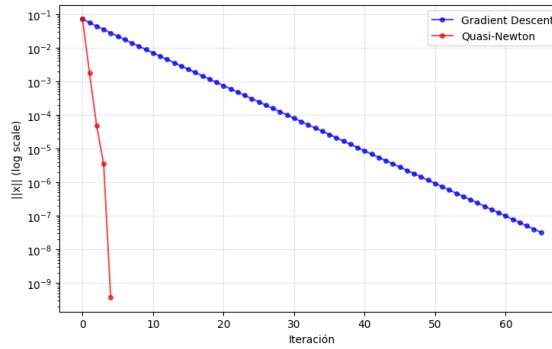


Figure 3: Convergencia de $\|x\|$ para el punto inicial $[-0.05, 0.05]$

El análisis de los resultados para puntos iniciales alejados del óptimo revela comportamientos distintivos en ambos algoritmos de optimización. En estos escenarios, el método de Gradiente Descendente exhibe una tendencia a la divergencia numérica catastrófica, como se observa en el experimento con punto inicial $(-73.5, 42.1)$, donde en apenas 2 iteraciones produce coordenadas extremas $(1.31 \times 10^{31}, -1.36 \times 10^{25})$ reportando $f_{\min} = 0.0$, resultado que representa un underflow numérico más que una convergencia genuina.

Por su parte, el método Quasi-Newton demuestra mayor robustez numérica pero con un costo computacional significativo. En el experimento 5 con punto inicial $(-55.6, -75.3)$, Quasi-Newton requiere 110 iteraciones para estabilizarse en coordenadas $(19.39, -706534.94)$ con $f_{\min} = 5.95 \times 10^{-9}$, evidenciando su capacidad para navegar regiones complejas mediante aproximaciones sucesivas de la matriz Hessiana, aunque a expensas de un mayor número de evaluaciones.

Los resultados de los valores de x obtenidos en este escenario se explican fundamentalmente por el paisaje matemático de la función objetivo, cuya estructura asintótica tiende a cero cuando $x \rightarrow +\infty$, creando un "efecto de atracción" hacia regiones con coordenadas extremas donde el valor numérico de la función es mínimo pero las soluciones carecen de sentido práctico.

La implicación crucial de este análisis es que, para funciones con comportamiento asintótico similar, la selección de puntos iniciales debe basarse en conocimiento del dominio problemático, complementándose con estrategias de regularización o restricciones que prevengan la convergencia hacia soluciones numéricamente inestables.

La superioridad de Quasi-Newton en este contexto radica en su capacidad para encontrar soluciones más estables, aunque requiera más iteraciones en algunos casos, mientras que Gradiente Descendente prioriza la reducción rápida del valor de la función incluso a costa de la calidad numérica de la solución.

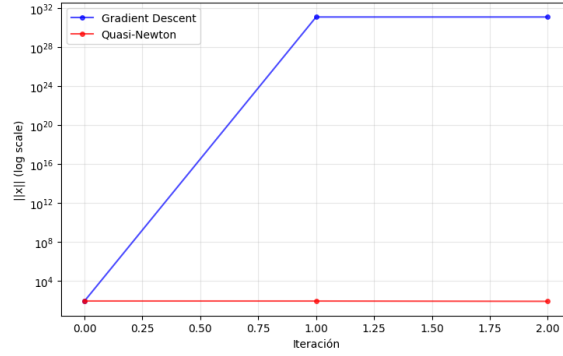


Figure 4: Convergencia de $\|x\|$ para el punto inicial $[-73.5, 42.1]$

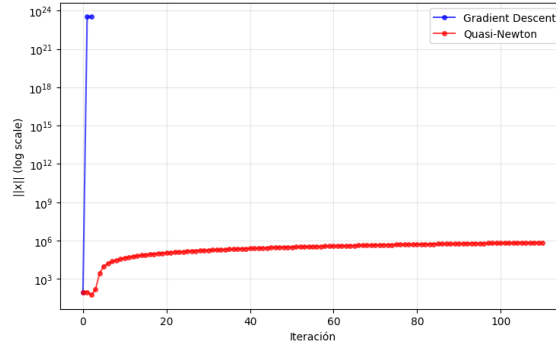


Figure 5: Convergencia de $\|x\|$ para el punto inicial $[-55.6, -75.3]$

Conclusiones

Al comparar ambos métodos, podemos notar que, cuando partimos de puntos cercanos al mínimo global en $(0, 0)$, tanto Máximo Descenso como Quasi-Newton hacen un excelente trabajo: llevan la función a valores prácticamente nulos y las coordenadas finales están muy cerca del mínimo real, resultado que se evidencia en la Figura 6. En este caso, Quasi-Newton tiene una pequeña ventaja: alcanza la solución en menos pasos y en menos tiempo, gracias a que utiliza información de la curvatura de la función para tomar decisiones más inteligentes en cada iteración (Figura 7).

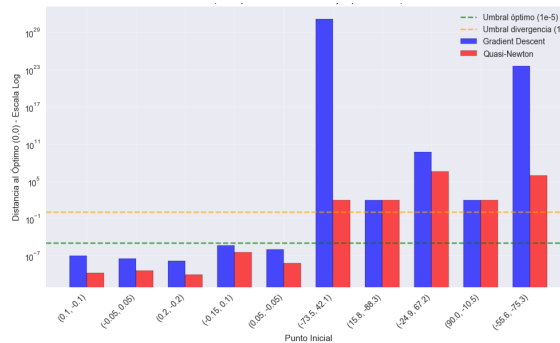


Figure 6: Distancia al Óptimo

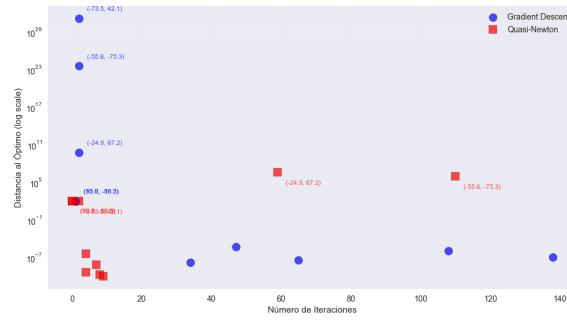


Figure 7: Iteraciones vs Calidad de Solución

Podemos concluir que la eficiencia de un método de optimización no depende solo del algoritmo en sí, sino también de la topología de la función y del punto inicial. En la práctica, esto implica que la elección de la técnica y la estrategia de inicialización deben adaptarse a cada problema, combinando métodos robustos con análisis previo de la función para evitar soluciones inestables o desplazamientos hacia regiones extremas.