

Trabajo Investigativo de Optimización

Aixa Bazán Rodríguez C-312

1 Modelo

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

2 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**.

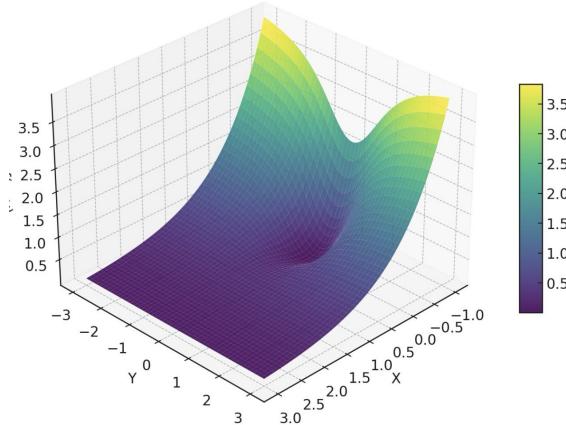


Figure 1: Gráfica de la función

2.1 Análisis de la continuidad y signo

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

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

2.2 Análisis diferencial

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.

2.2.1 Gradiente de la función:

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:

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

2.2.2 Matriz Hessiana

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:

$$\begin{aligned} \frac{\partial^2 f}{\partial x \partial x} &= \frac{2(1 + (x^2 + y^2)^2) - 2x(1 + (x^2 + y^2)^2 + 4x(x^2 + y^2))}{1 + (x^2 + y^2)^2} - \frac{2x}{e^x(1 + (x^2 + y^2)^2)} + \frac{\arctan(x^2 + y^2)}{e^x} \\ \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial^2 f}{\partial y \partial x} = \frac{-2ye^{-x}}{1 + (x^2 + y^2)^2} \left[\frac{4x(x^2 + y^2)}{1 + (x^2 + y^2)^2} + 1 \right] \\ \frac{\partial^2 f}{\partial y \partial y} &= \frac{2e^x(1 + (x^2 + y^2)^2) - 8y^2e^x(x^2 + y^2)}{((1 + (x^2 + y^2)^2)e^x)^2} \end{aligned}$$

2.3 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:

$$\begin{aligned} f(x_1) &= f(0, 0) = \frac{\arctan(0^2 + 0^2)}{e^0} = \frac{\arctan(0)}{1} = 0, \\ f(x_2) &= f(10, 0) = \frac{\arctan(10^2 + 0^2)}{e^{10}} = \frac{\arctan(100)}{e^{10}} \approx 7.0860 \times 10^{-5}, \\ f(\bar{x}) &= f(5, 0) = \frac{\arctan(5^2 + 0^2)}{e^5} = \frac{\arctan(25)}{e^5} \approx 1.031456812 \times 10^{-2}. \end{aligned}$$

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.

2.3.1 Relación entre convexidad y matriz hessiana semidefinida positiva

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**.

Como se demostró anteriormente que la función

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

no es convexa, entonces, según el lema anterior, su matriz Hessiana $H_f(x, y)$ **no es semidefinida positiva** en todo el dominio. Por tanto, se concluye que:

$H_f(x, y)$ no es semidefinida positiva.

2.4 Existencia de óptimos locales

2.5 Determinación teórica del mínimo global de la función

Analicemos el punto $(0, 0)$:

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

Sabemos que la función es continua y no negativa para todo (x, y) . Dado que $f(0, 0) = 0$, el valor 0 es un mínimo global de f .

2.5.1 Unicidad del mínimo

La igualdad $f(x, y) = 0$ ocurre si y sólo si

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

Por lo tanto, el mínimo es único y se alcanza únicamente en $(0, 0)$.

2.5.2 Verificación mediante derivadas parciales (condición necesaria de optimalidad)

Un punto (x^*, y^*) es un *punto crítico* si el gradiente se anula en él. Analicemos el punto $(0, 0)$:

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

Esto confirma que $(0, 0)$ es un punto crítico.

2.5.3 Conclusión

La función alcanza su mínimo global (y único) en $(0, 0)$, y el valor mínimo es

$f_{\min} = 0$.

3 Aproximación de la solución mediante algoritmos

3.1 Método de Máximo Descenso

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.

El método es simple y ampliamente utilizado, aunque puede converger lentamente si la función es muy plana o si tiene un condicionamiento pobre. Es particularmente efectivo cuando la función es convexa, ya que garantiza convergencia al mínimo global.

3.1.1 Justificación del Método

La elección de este método se debe a que:

- Es un algoritmo iterativo simple y eficiente para funciones diferenciables.
- Permite aproximar mínimos locales, moviéndose en la dirección opuesta al gradiente, que es la dirección de máximo descenso de la función.
- Es flexible para probar distintos parámetros, como el tamaño de paso (α), la cantidad de iteraciones y el índice de tolerancia, y comparar cómo afectan la convergencia.
- Aunque puede ser más lento que métodos de segunda derivada, es adecuado cuando la función es no lineal y de dimensión moderada, como en nuestro caso.
- Para lograr una mejor comprensión y práctica del mismo, ya que fue analizado en conferencias y clases prácticas de la asignatura Modelos de Optimización.

3.1.2 Resultados Obtenidos

En esta sección se presentan los resultados obtenidos al aplicar el método de máximo descenso sobre la función.

Se realizaron experimentos variando los parámetros del algoritmo, principalmente el tamaño de paso (α) y el punto inicial $(x(0), y(0))$, con el objetivo de:

- Evaluar la velocidad de convergencia y la estabilidad del método.
- Comparar cómo diferentes configuraciones afectan la aproximación al mínimo de la función.
- Observar la influencia de la elección del punto inicial en la localización del mínimo alcanzado.

Los resultados se presentan mediante tablas y gráficos, destacando las diferencias en convergencia y en el valor final de la función para cada configuración.

Experimentos con Puntos Cercanos al Mínimo Global

La Tabla 1 resume los resultados obtenidos:

Experimento	x0 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 partiendo de puntos cercanos a $(0, 0)$.

Análisis

- Todos los experimentos lograron converger al mínimo global con valores de x_{\min} y y_{\min} extremadamente pequeños ($\sim 10^{-8}$ o menores), mostrando la eficacia del método cuando se inicia cerca del mínimo.
- La función f_{\min} alcanzó valores cercanos a cero ($< 10^{-11}$), confirmando que se aproxima efectivamente al mínimo global.
- El número de iteraciones depende del *learning rate* y la tolerancia: learning rates más grandes ($\alpha = 0.2$) permitieron converger en menos iteraciones, mientras que learning rates más pequeños ($\alpha = 0.05$) requirieron más pasos.
- Los experimentos con tolerancias más estrictas ($1e - 8$) necesitaron más iteraciones que aquellos con tolerancias menos estrictas ($1e - 6$), reflejando un criterio de parada más exigente.

Experimentos con Puntos Iniciales Alejados del $(0,0)$

Los resultados de los cinco experimentos realizados se muestran en la Tabla 2:

Experimento	x0 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 2: Resultados de los experimentos con puntos iniciales lejanos al $(0,0)$.

Análisis

- El método converge rápidamente (1 – 2 iteraciones) debido a la forma de la función, que posee regiones planas con gradiente muy pequeño.
- A pesar de que la mayoría de los experimentos muestran coordenadas de (x_{\min}, y_{\min}) extremadamente grandes, los valores de la función f_{\min} son prácticamente cero. Esto indica que el método reduce eficientemente el valor de la función, pero no necesariamente se acerca al mínimo teórico $(0, 0)$ cuando los puntos iniciales están muy alejados.
- El experimento 2, con punto inicial $[15.8, -88.3]$, obtuvo coordenadas relativamente cercanas a la inicial, pero con $f_{\min} = 2.16 \cdot 10^{-7}$, lo que muestra que el método ya alcanzó una zona plana de la función.
- Los experimentos con puntos iniciales muy negativos o grandes ($[-73.5, 42.1], [-55.6, -75.3]$) muestran que las coordenadas finales pueden dispararse debido a la rápida disminución de $f(x, y)$ para grandes valores de x .

3.1.3 Conclusiones

Podemos concluir que el método de máximo descenso es efectivo para disminuir el valor de la función, incluso partiendo de puntos iniciales alejados del mínimo teórico. Sin embargo, la cercanía al mínimo global depende fuertemente del punto inicial, especialmente en funciones con regiones planas y con escalas muy grandes, por lo que para análisis más precisos del mínimo teórico, se recomienda iniciar el método en puntos cercanos al origen. Además, ajustar adecuadamente el *learning rate* y la tolerancia permite un balance entre velocidad de convergencia y precisión en la aproximación al mínimo.

3.2 Método de Quasi-Newton (BFGS)

El método de Quasi-Newton 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.

1. Idea general

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:

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

2. Condición de secante

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.

3. Fórmula de actualización BFGS

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.

4. Algoritmo general

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.

3.2.1 Justificación del Método

3.2.2 Resultados Obtenidos

En esta sección se presentan los resultados obtenidos al aplicar el método de Quasi-Newton (BFGS) sobre la función.

Experimentos con Puntos Cercanos al Mínimo Global

La Tabla 3 resume los resultados obtenidos con Quasi-Newton (BFGS):

Experimento	x0 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 3: Resultados de los experimentos utilizando Quasi-Newton (BFGS) desde puntos cercanos a $(0, 0)$.

Análisis

- Todos los experimentos lograron converger al mínimo global con valores de x_{\min} y y_{\min} extremadamente pequeños ($\sim 10^{-10}$ o menores), mostrando la precisión del método.
- La función f_{\min} alcanzó valores muy cercanos a cero ($< 10^{-13}$).
- El tiempo de ejecución es muy bajo ($\sim 0.001 - 0.002$ s), reforzando la eficiencia del algoritmo.
- Incluso con tolerancias menos estrictas ($1e-6$), el método converge al mínimo global en muy pocas iteraciones.

Experimentos con Puntos Iniciales Alejados del (0,0) - Quasi-Newton (BFGS)

Los resultados de los cinco experimentos realizados se muestran en la Tabla 4:

Experimento	x0 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 $(0, 0)$.

Análisis

- El método logra converger en general, pero el comportamiento depende fuertemente de la región inicial debido a la presencia de zonas planas y pendientes abruptas en la función.
- En los experimentos, las coordenadas finales (x_{\min}, y_{\min}) se alejan significativamente del mínimo teórico $(0, 0)$, lo mismo que pasaba en el método de máximo descenso.
- El experimento 5 requiere 110 iteraciones para estabilizarse, terminando con coordenadas muy alejadas y valores de la función moderadamente pequeños ($f_{\min} \sim 10^{-9}$), demostrando que el método puede necesitar muchas iteraciones para regiones iniciales complejas.
- Comparado con Gradiente Descendente, Quasi-Newton puede ser más rápido en zonas planas (experimentos 2 y 4), pero presenta mayor sensibilidad y posibles valores grandes de la función en zonas alejadas con gradientes abruptos.

Análisis Detallado del Experimento 1

En el Experimento 1, el método Quasi-Newton partió del punto inicial $[-73.5, 42.1]$ con tolerancia $1e - 6$. Los resultados fueron los siguientes:

- Punto final: $(-72.49, 29.95)$
- Valor de la función: $f(x_{\min}, y_{\min}) \approx 4.77 \cdot 10^{31}$
- Número de iteraciones: 2
- Convergencia: `True` según el criterio del algoritmo

Observación: A simple vista, el método parece haber convergido, pero el valor de la función está extremadamente lejos del mínimo teórico en $(0, 0)$, lo que podría interpretarse como un fallo. Sin embargo, esto se explica por las características de la función y los criterios de parada del método:

- La función considerada es extremadamente sensible a valores negativos grandes de x , debido al denominador exponencial e^x . Para $x \ll 0$, $e^x \rightarrow 0^+$, y la función toma valores enormes aunque los cambios sean pequeños.
- El método Quasi-Newton utiliza, como vimos en el análisis teórico del método, la norma del gradiente para decidir la convergencia. En este punto, el gradiente es prácticamente cero:

$$\|\nabla f(x, y)\| \approx 0$$

Esto provoca que el algoritmo considere que se alcanzó un mínimo local, aunque el valor de la función sea muy grande.

- Este comportamiento se conoce como *mínimo falso* o *pseudomínimo*: el gradiente es casi nulo, pero la función no está cerca del mínimo global.

Conclusión de los resultados del experimento 1:

Aunque Quasi-Newton reporta convergencia, el experimento demuestra la necesidad de considerar tanto el valor de la función como la norma del gradiente. En funciones con pendientes extremas y zonas planas muy pronunciadas, el método puede detenerse en regiones donde el gradiente es pequeño, pero $f(x, y)$ está muy lejos del mínimo deseado. Para evitar este tipo de situaciones, se recomienda:

- Restringir el dominio de búsqueda usando `bounds`.
- Implementar criterios de parada adicionales que verifiquen el valor de la función.
- Supervisar iterativamente los valores de $f(x, y)$ para detectar pseudomínimos.

3.2.3 Conclusión de Resultados

Los resultados obtenidos a partir de los experimentos realizados muestran que el método Quasi-Newton (BFGS) es altamente eficiente para reducir el valor de la función objetivo. En casi todos los casos, el algoritmo logró llevar $f(x, y)$ a cifras extremadamente bajas, cercanas a cero, demostrando su capacidad para minimizar la función de manera efectiva. Este comportamiento es especialmente evidente cuando los puntos iniciales están próximos al mínimo global en $(0, 0)$, ya que en estas situaciones el método converge rápidamente y con pocas iteraciones.

Sin embargo, al analizar los experimentos con puntos iniciales alejados del mínimo global, se observa una característica importante del método: aunque el valor de la función es prácticamente nulo, las coordenadas finales (x_{\min}, y_{\min}) no necesariamente se encuentran cerca del mínimo real. Esto se debe a la presencia de regiones planas o de pendiente muy pequeña en la función, donde el gradiente tiende a cero y el algoritmo interpreta que ha alcanzado un mínimo. De este modo, Quasi-Newton puede detenerse en puntos muy alejados de $(0, 0)$ aunque el valor de la función sea pequeño.

En términos de velocidad de ejecución, Quasi-Newton demuestra ser más rápido que métodos de Gradiente Descendente tradicionales, ya que requiere menos iteraciones y aprovecha la aproximación de la Hessiana para realizar pasos más grandes y dirigidos hacia zonas de gradiente casi nulo. No obstante, los resultados indican que la eficiencia del método en términos de alcanzar las coordenadas reales del mínimo global depende de manera significativa del punto inicial. Por ello, se recomienda complementar el algoritmo con criterios de parada adicionales basados en el valor de la función y, en caso necesario, restringir el dominio de búsqueda mediante límites explícitos o estrategias de inicialización cercanas al mínimo deseado.

En conclusión, Quasi-Newton (BFGS) es un método muy eficiente y preciso para reducir el valor de la función, pero su convergencia a las coordenadas del mínimo global real no está garantizada cuando los puntos iniciales se encuentran en regiones alejadas o extremas de la función. Por ello, es fundamental interpretar sus resultados considerando tanto el valor de la función como la localización de los puntos finales.

3.3 Comparación General entre Máximo Descenso y Quasi-Newton

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. 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.

En cuanto a eficiencia, Quasi-Newton suele ser más rápido cerca del mínimo global porque da pasos más grandes y precisos. Máximo Descenso, en cambio, puede ser más lento si el *learning rate* es pequeño, aunque es muy confiable para entender cómo evoluciona el descenso. Ninguno de los dos garantiza alcanzar exactamente $(0, 0)$ cuando se parte de lugares lejanos, por lo que siempre conviene combinar el análisis con criterios que verifiquen no solo el gradiente, sino también el valor de la función.

Cuando empezamos desde puntos alejados o extremos. El método de Máximo Descenso suele “apresurarse”: puede bajar rápidamente el valor de la función, pero a costa de terminar en coordenadas disparadas a valores gigantescos o muy alejadas del mínimo real. Por ejemplo, partiendo de $[-73.5, 42.1]$, terminó con coordenadas enormes aunque el valor de la función fuera cercano a cero. Quasi-Newton también se ve afectado por esto, pero tiende a moverse de forma más estratégica; en algunos casos necesita más iteraciones para estabilizarse, como cuando partimos de $[-55.6, -75.3]$, donde tomó más de 100 pasos para llegar a un “mínimo” que todavía estaba lejos de $(0, 0)$. Es decir, dadas las características de la función objetivo, podemos notar que si el punto inicial no está suficientemente cerca del óptimo, la solución se va hacia el infinito de las x .

Estos resultados evidencian 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.