Análisis de Función Irrestricta: Algoritmos de Optimización: Método de Newton vs. Descenso del Gradiente

Melissa Maureen Sales Brito
Universidad de La Habana
Facultad de Matemática y Computación

13 de noviembre de 2025

Resumen

Este trabajo presenta un análisis comparativo entre el método de Newton y el descenso del gradiente aplicados a la minimización de la función $f(x, y) = x^2 + y^2 + xy + \sin(x) + \cos(y)$. Se implementaron ambos algoritmos con búsqueda de línea de Armijo y se evaluó su desempeño en 26 escenarios diferentes, analizando convergencia, robustez y eficiencia computacional.

Índice

1. Introducción

Dado un problema de optimización se desea analizar teóricamente que tipo de problema es, aplicar algoritmos conocidos adecuados para el problema y analizar la calidad de la solución obtenida. Se hará comparación de resultados y graficación de los mismos para llegar a conclusiones sobre el uso de los algoritmos.

2. Análisis Teórico

Problema de optimización sin restricciones:

$$\min_{x \in \mathbb{R}^n} f(x) \tag{1}$$

Nuestra función objetivo es:

$$f(x, y) = x^2 + y^2 + xy + \sin(x) + \cos(y) \tag{2}$$

La función es el resultado de la suma de funciones elementales continuas por lo que es una función elemental continua.

Variables Independientes: $x, y \in \mathbb{R}$

Es diferenciable en $\mathbb{R}^2$ ya que todos sus términos son diferenciables.

Teniendo en cuenta que:

$$-1 \leq \sin(x) \leq 1 \quad \text{y} \quad -1 \leq \cos(y) \leq 1 \tag{3}$$

y que la parte cuadrática satisface:

$$x^2 + y^2 + xy = \left(x + \frac{y}{2}\right)^2 + \frac{3}{4}y^2 \geq 0 \quad \forall x, y \in \mathbb{R} \tag{4}$$

podemos llegar a la cota inferior:

$$f(x, y) \geq -2 \tag{5}$$

El gradiente de f está dado por:

$$\nabla f(x, y) = \begin{pmatrix} 2x + y + \cos(x) \\ 2y + x - \sin(y) \end{pmatrix} \tag{6}$$

La matriz Hessiana es:

$$\nabla^2 f(x, y) = \begin{pmatrix} 2 - \sin(x) & 1 \\ 1 & 2 - \cos(y) \end{pmatrix} \tag{7}$$

Todos los menores principales de la matriz Hessiana son mayores o iguales a cero pues el valor máximo que puede tomar tanto el seno como el coseno es 1. El determinante en el peor caso es 0 (cuando $sen(x) = cos(y) = 1$) por lo que podemos afirmar que $f(x, y)$ no es estrictamente convexa pero si es convexa dado que su matriz Hessina es semidefinida positiva.

Existencia de solución:

Conocemos que la función es continua, no esta acotada pues solo posee cota inferior. Como sabemos que el seno y coseno son funciones acotadas:

Proposición 1. *La función* $f(x,y) = x^2 + y^2 + xy + \sin(x) + \cos(y)$ *es coerciva.*

Demostración. Notemos que:

$$f(x,y) = x^2 + y^2 + xy + \sin(x) + \cos(y) \geq x^2 + y^2 + xy - 2$$

Por lo tanto:

$$\lim_{\|(x,y)\|\to\infty} f(x,y) = +\infty$$

□

Luego existe mínimo global es decir solución para nuestro problema.

3. Selección de los Algoritmos

Hay dos filosofías generales en las que se basan los algoritmos para funciones irrestrictas:

- Buscar una dirección donde la función objetivo disminuye su valor y tomar un nuevo punto moviéndose en esa dirección.
- Realizar la búsqueda en una región.

Basandonos en esto, en el previo análisis de $f(x,y)$ y en la complejidad de implementación de los algoritmos se seleccionan dos algoritmos que emplean la primera filosofía: Descenso del Gradiente y Método de Newton.

Si calculamos el gradiente de la función en el origen tenemos como resultado $(1,0) \neq (0,0)$ pero si ignoramos los términos trigonométricos entonces se anula el gradiente, al ser actodados entre $[-1,1]$ podemos llegar a una conclusión apresurada de que el óptimo podría estar en una vecindad estrictamente convexa del origen basandonos en la sustitución de la Hessiana en $(0,0)$.

Con esta pista podemos hacer experimentos con los algoritmos aprovechando sus propiedades.

3.1. Método de Descenso del Gradiente

El método de Descenso del Gradiente es un algoritmo iterativo utilizado para minimizar una función diferenciable $f(x)$, basándose únicamente en la información proporcionada por su gradiente.

A partir de un punto inicial x_0, el algoritmo genera una secuencia de aproximaciones x_k que idealmente convergen hacia un mínimo local de f. En cada iteración, se determina una dirección de descenso d_k y un tamaño de paso α_k que indican cómo avanzar hacia una región de menor valor de la función. En este método, la dirección de búsqueda se define como el negativo del gradiente de la función en el punto actual:

$$d_k = -\nabla f(x_k), \tag{8}$$

ya que el gradiente apunta hacia la dirección de mayor incremento de f, y por tanto su opuesto indica la dirección de descenso más pronunciado.

La nueva aproximación se obtiene actualizando el punto actual en la dirección de descenso:

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

donde el parámetro $\alpha_k > 0$ controla el tamaño del paso. El valor de α_k se selecciona mediante una búsqueda de línea con la condición de Armijo, la cual garantiza que el paso elegido produzca una disminución suficiente en el valor de la función. Esto evita tanto pasos demasiado grandes (que podrían causar divergencia) como pasos demasiado pequeños (que ralentizan la convergencia).

El proceso iterativo continúa hasta que se cumple alguno de los siguientes criterios:

- La **distancia** entre dos iteraciones consecutivas $\|x_{k+1}x_k\|$ es menor que una tolerancia prefijada tol, indicando que las actualizaciones se han vuelto insignificantes.
- (Opcional) La **norma del gradiente** $\|\nabla f(x_k)\|$ sea suficientemente pequeña, lo cual sugiere que estamos cerca de un punto crítico.
- Se alcanza el **número máximo de iteraciones** `max_iter`, para evitar ciclos infinitos en caso de no convergencia.

3.2. Método de Newton

El método de Newton varía el método de Descenso Máximo del Gradiente empleando la siguiente igualdad para hallar la dirección:

$$-\nabla f(x_k) = \nabla^2 f(x_k) d_k \tag{9}$$

quee se obtiene al aproximar f por su expansión de Taylor de segundo orden alrededor del punto actual.

actualiza con:

$$x_{k+1} = x_k + \alpha_k d_k \tag{10}$$

Al incluir la información de la curvatura (Hessina), el método es más certero al hallar la dirección y como en casi todo el dominio H_f es bien condicionada tiene sentido utilizar este método. Para obtener la dirección de descenso tenemos dos opciones clásicas:

- Cálculo de la inversa de la Hessiana:

  Recordemos que el determinante de H_f se anula para $sen(x) = cos(y) = 1$, es decir $x = \frac{\pi}{2} + 2k\pi$ y $y = 2m\pi$, por lo que en ese punto H_f no es invertible. En el resto de puntos podemos usar la fórmula standard para matrices de $2x2$:

  $$H^{-1}(x, y) = \frac{1}{\Delta(x, y)} \begin{pmatrix} 2 - \cos y & -1 \\ -1 & 2 - \sin x \end{pmatrix}$$

  Problemas que podría tener el calcúlo de la inversa: A medida que el determinante se acerca a 0 empeora el condicionamiento de la Hessina por ende se amplifican los errores númericos.

- Resolver el sistema dado por la igualdad (9):

  Resolver este sistema lineal utilizando factorizaciones o métodos lineales, es más estable numéricamente para nuestra función.

4. Diseño Experimental

Se diseñaron 26 experimentos organizados en las siguientes categorías:

Cuadro 1: Categorías de experimentos según distancia inicial

Categoría	Rango	Experimentos	Objetivo
Cerca	$[0, 5)$	4	Comportamiento estándar
Moderado	$[5, 15)$	4	Casos típicos
Lejos	$[15, 30)$	4	Desafío medio
Muy lejos	$[30, 60)$	4	Casos difíciles
Extremo	$[60, 150)$	4	Límites de robustez
Tolerancia	Variable	4	Sensibilidad a ϵ
Asimétrico	Variable	2	Coordenadas dispares

Los puntos iniciales se seleccionaron para cubrir:

- Los cuatro cuadrantes del plano
- Diferentes distancias al origen (desde 1.4 hasta 141.4 unidades)
- Configuraciones asimétricas (e.g., $(30, -5)$, $(-3, 40)$)

Cuadro 2: Parámetros utilizados en los experimentos

Parámetro	Valor	Descripción
α_0	1.0	Tamaño de paso inicial
m_1	0.01	Parámetro de Armijo
τ	0.5	Factor de reducción
ϵ	10^{-6}	Tolerancia estándar
max_iter (GD)	100–2000	Máximo iteraciones
max_iter (Newton)	50–150	Máximo iteraciones

Para cada experimento se registraron:

1. Número de iteraciones hasta convergencia
2. Trayectoria completa: $\{x_0, x_1, \ldots, x_k\}$
3. Estado de convergencia (éxito/fallo)
4. Punto fina
5. Evaluación final de la función

5. Resultados

Cuadro 3: Tasa de éxito por método y categoría

Categoría	**Descenso Gradiente**	**Método Newton**
Cerca (0-5)	100 % (4/4)	100 % (4/4)
Moderado (5-15)	100 % (4/4)	100 % (4/4)
Lejos (15-30)	100 % (4/4)	100 % (4/4)
Muy lejos (30-60)	100 % (4/4)	100 % (4/4)
Extremo (60+)	100 % (4/4)	100 % (4/4)
Asimétrico	100 % (2/2)	100 % (2/2)
Tolerancia	100 % (4/4)	100 % (4/4)
Total	**100 % (26/26)**	**100 % (26/26)**

Cuadro 4: Promedio de iteraciones por categoría

Categoría	**GD**	**Newton**
Cerca (0-5)	31.0	6.0
Moderado (5-15)	34.1	7.7
Lejos (15-30)	27.0	7.0
Muy lejos (30-60)	35.0	8.0
Extremo (60+)	25.0	10.0

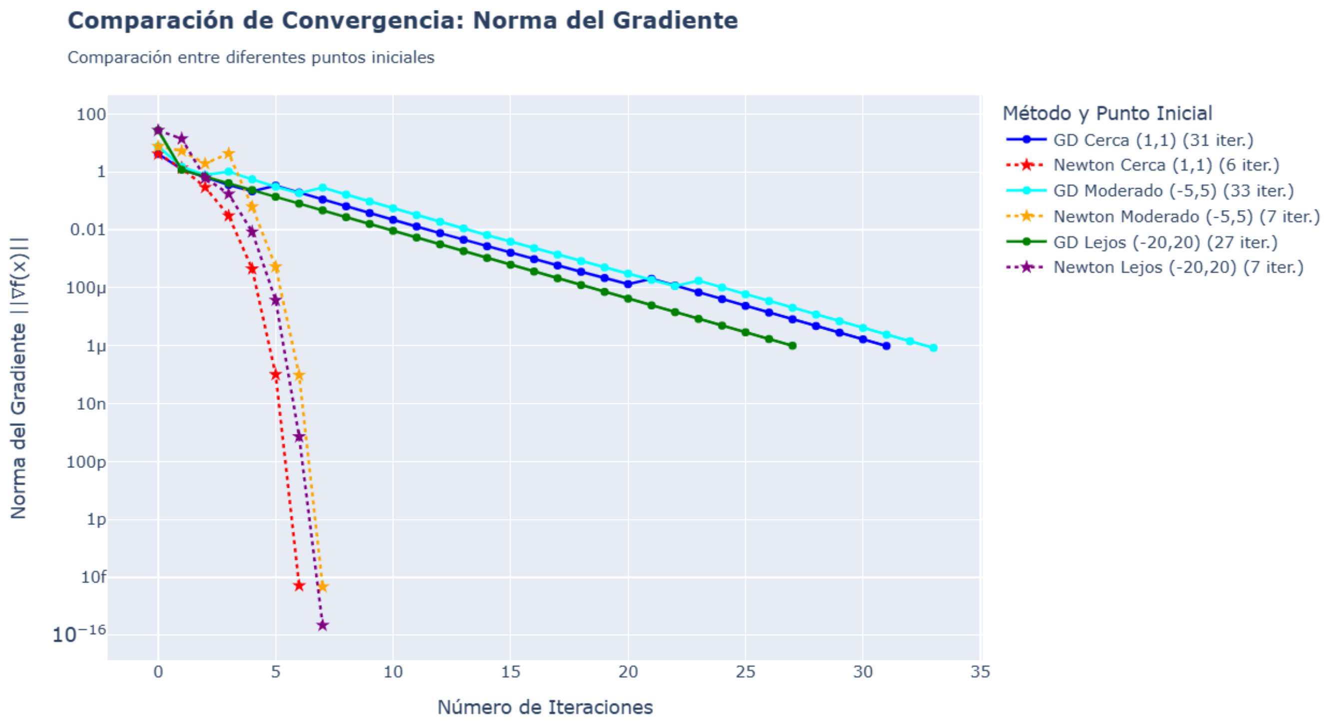

Figura 1: Comparación de la norma del gradiente vs iteraciones para diferentes puntos iniciales.

En la figura se puede apreciar la convergencia lineal ya expuesta del algoritmo de descenso máximo y la propiedad del método de Newton de tener una tendencia a converger cudráticamente si el punto inicial está cerca del óptimo, observemos que nuestra apreciación inicial de la vecindad del origen fue correcta.

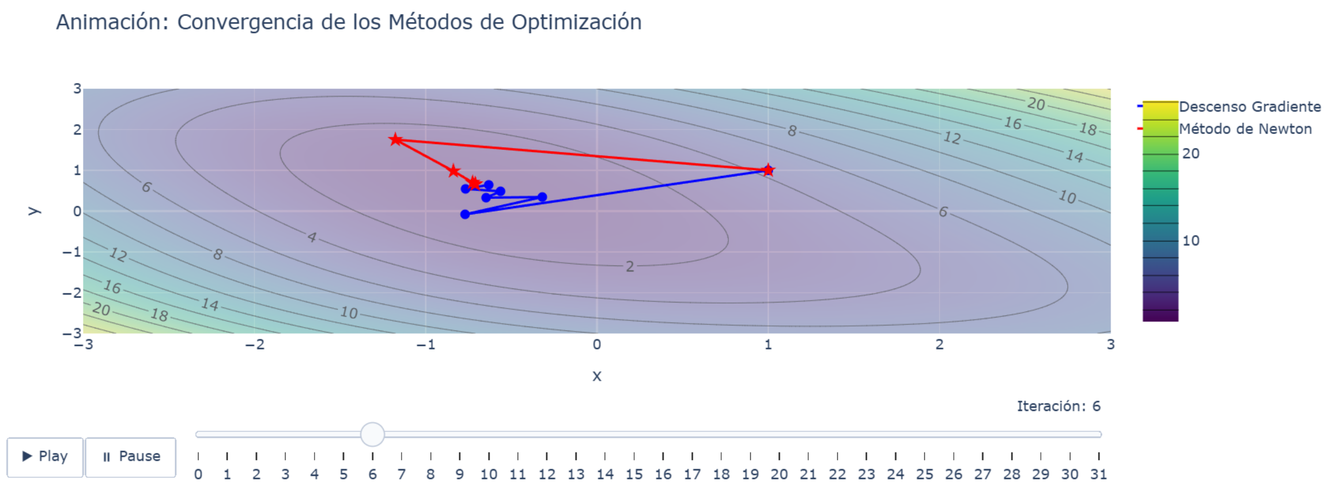

Figura 2: Trayectorias de convergencia en el espacio de búsqueda.

Observemos que Newton sigue una trayectoria más directa mientras que las altas probabilidades de que Descenso Máximo hiciera trayectoria en zigzag se cumplieron. Este es el caso particular de punto inicial cercano al origen, para la tercera iteracion ya Newton ha logrado la convergencia mientras que Descenso Máximo necesita 31 iteraciones en total.

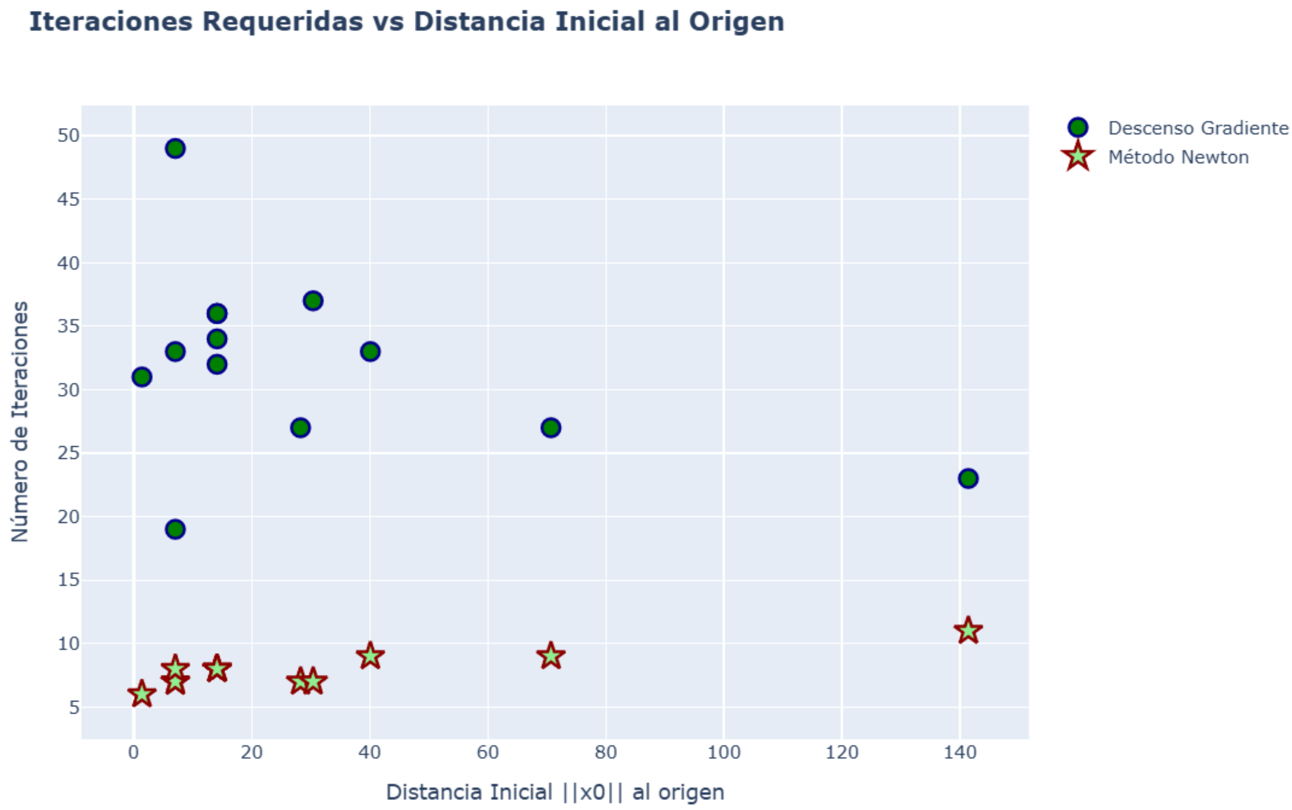

Figura 3: Relación entre la distancia inicial al óptimo y el número de iteraciones requeridas. Newton muestra crecimiento logarítmico mientras que el descenso del gradiente muestra crecimiento irregular.

Observación: La ventaja de Newton se incrementa con tolerancias más estrictas.

6. Conclusiones

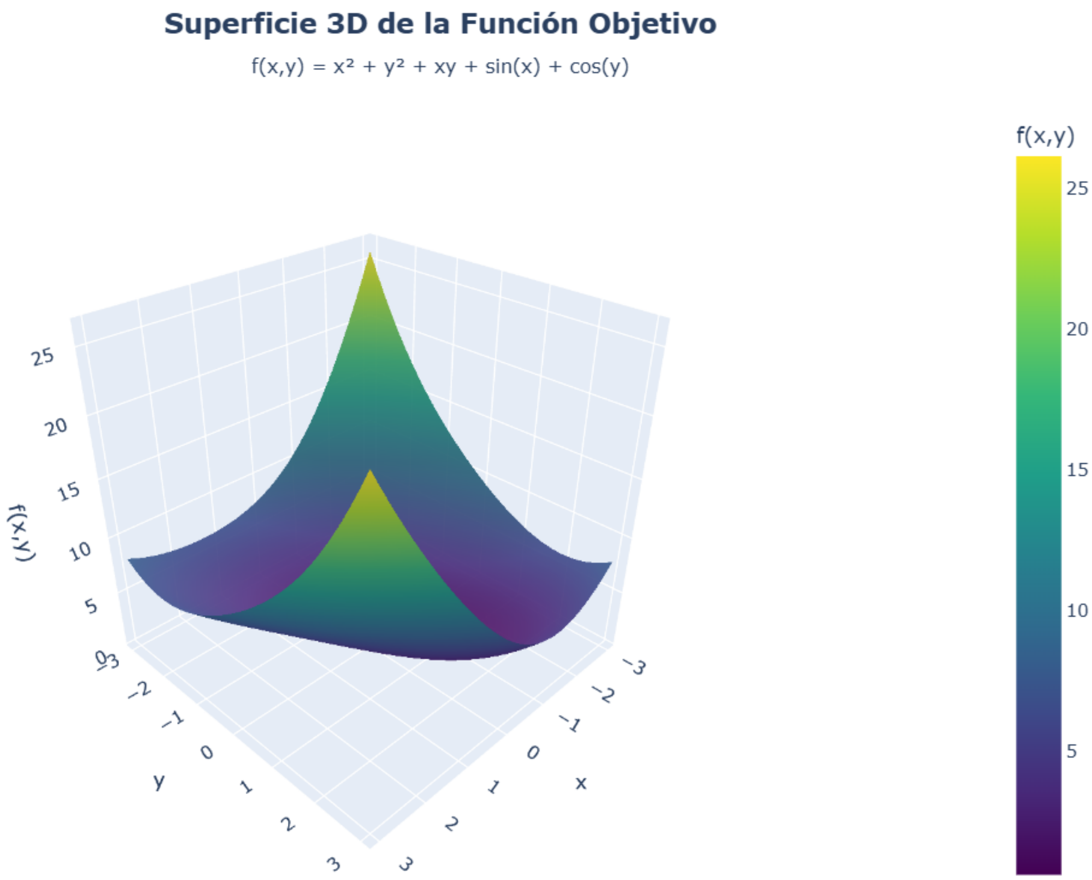

Figura 4: Superficie 3D de la funcón objetivo.

Del gráfico se obtuvo el punto mínimo en la malla utilizada que es $x = -0{,}696970, y = 0{,}636364, f = 0{,}609564$.

Este valor de la función esta aproximado con gran exactitud por los resusltados obtenidos en el 100 % de los experimentos para ambos algoritmos, como ejemplo para el caso cercano tenemos:

Newton:

$final_x : [-0{,}7105124332210255, 0{,}6629971112840694]$, $final_f : 0{,}609254208027612$

Descenso Máximo:

$final_x : [-0{,}7105116125045395, 0{,}6629965782875904]$, $final_f : 0{,}6092542080282398$

Para determinar los puntos de estabilidad se resolvió el sistema no lineal

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

mediante el método de Newton y Descenso Máximo del Gradiente, considerando múltiples puntos iniciales en una malla extensa del plano. El procedimiento converge hacia un único punto crítico dentro de la tolerancia numérica utilizada:

$$(x^*, y^*) \approx (-0{,}71051243,\ 0{,}66299711),$$

en el cual el valor de la función es:

$$f(x^*, y^*) \approx 0{,}6092542.$$

La matriz Hessiana evaluada en dicho punto es:

$$H(x^*, y^*) = \begin{pmatrix} 2 - \sin(-0{,}7105) & 1 \\ 1 & 2 - \cos(0{,}6630) \end{pmatrix} \approx \begin{pmatrix} 2{,}652 & 1 \\ 1 & 1{,}229 \end{pmatrix},$$

cuyos autovalores son aproximadamente:

$$\lambda_1 \approx 0{,}6997, \qquad \lambda_2 \approx 3{,}1644.$$

Dado que ambos autovalores son positivos, la matriz Hessiana es definida positiva en el punto crítico, por lo que éste corresponde a un mínimo local estricto. Además, como la función $f(x, y)$ es coerciva (tiende a $+\infty$ cuando $\|(x, y)\| \to \infty$) y la Hessiana es semidefinida positiva en todo el dominio, se concluye que dicho mínimo local es también el mínimo global de la función.

El número de condición de la Hessiana en el óptimo se estima como:

$$\kappa(H) = \frac{\lambda_{\text{máx}}}{\lambda_{\text{mín}}} \approx \frac{3{,}1644}{0{,}6997} \approx 4{,}52,$$

lo cual indica que la matriz está bien condicionada, garantizando estabilidad numérica en el cálculo de la dirección de Newton.

6.1. Notas Finales

Se hizo uso consciente de la IA para la generación supervisada del código referente a la graficación en plotly.

6.2. Recomendaciones

Analizar dados los resultados de los experimentos variando la toleriancia y la simetría del punto inicial como se comporta cada algoritmo.

Notar los cambios no monótonos de las iteraciones por categoria de distancia(existen casos en los que la distancia del origen es mayor pero la cantidad de iteraciones es menor) para el descenso máximo del gradiente lo que nos da pie a investigar sobre el condicionamiento de la Hessiana según la región, como afecta α en la convergencia del método y otros factores que pueden influir.

Referencias

[1] Bouza Allende, G. (2021). *Optimización Matemática I: Nota de clase*. Universidad de La Habana.

[2] Bouza Allende, G. (2021). *Conferencia 5: Problemas de programación no lineal convexo*. Universidad de La Habana.

[3] Rodríguez, J. (2021). *Clase práctica de algoritmos de optimización*. Universidad de La Habana.

[4] Sales Brito, M. M. (2024). *Repositorio del proyecto*. https://github.com/MaureenSales/OptimizacionProyecto.git