

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

12 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	3
1.1. Objetivos	3
2. Análisis Teórico	3
2.1. Problema de Optimización	3
2.2. Análisis de la función	3
2.3. Gradiente y Hessiana	3
2.4. Convexidad	4
2.5. Existencia de solución	4
2.6. Selección de los Algoritmos	4
2.7. Método de Descenso del Gradiente	5
2.8. Método de Newton	5
3. Diseño Experimental	5
3.1. Configuración de Experimentos	5
3.2. Puntos Iniciales	6
3.3. Parámetros del Algoritmo	6
3.4. Métricas de Evaluación	6
4. Resultados	7
4.1. Resultados Generales	7
4.1.1. Tasa de Éxito	7
4.1.2. Número de Iteraciones	7
4.2. Análisis de Convergencia	8
4.3. Análisis Visual de Trayectorias	9
4.4. Impacto de la Distancia Inicial	10
5. Conclusiones	11
5.1. Notas Finales	11
5.2. Recomendaciones	12

1. Introducción

1.1. Objetivos

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

2.1. Problema de Optimización

Problema de optimización sin restricciones:

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

Nuestra función objetivo es:

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

2.2. Análisis de la función

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

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

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

Acotación: Teniendo en cuenta que:

$$-1 \leq \sin(x) \leq 1 \quad y \quad -1 \leq \cos(y) \leq 1 \quad (3)$$

y que la parte cuadrática satisface:

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

podemos llegar a la cota inferior:

$$f(x, y) \geq -2 \quad (5)$$

2.3. Gradiente y Hessiana

El gradiente de f está dado por:

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

La matriz Hessiana es:

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

2.4. Convexidad

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 $\sin(x) = \cos(y) = 1$) por lo que podemos afirmar que $f(x, y)$ no es estrictamente convexa pero si es convexa dado que su matriz Hessiana es semidefinida positiva.

2.5. Existencia de solución

Conocemos que la función es continua, no está 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:

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

□

Luego existe mínimo global es decir solución para nuestro problema.

2.6. 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 acotados 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.

2.7. Método de Descenso del Gradiente

Tomamos un punto inicial x_0 y buscamos una dirección d_0 tal que f disminuya su valor en el rayo que comienza en x_0 con dirección d_0 . Para el caso de este método d se iguala a $-\nabla f$ en cada iteración. Actualiza la solución mediante:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) \quad (8)$$

donde α_k se determina mediante búsqueda de línea de Armijo para nuestra implementación. Como condición de parada calculamos la distancia de una iteración a otra y comparamos con un epsilon, fundamentado en la aparente convergencia de la sucesión numérica obtenida durante el proceso iterativo. Otra condición que no hemos implementado pero que también es de parada es que la norma del gradiente de la función en la iteración actual sea lo suficientemente pequeña. Además se define un número máximo de iteraciones para evitar ciclos infinitos.

Usa solo como información ∇f . Posee convergencia lineal. Puede hacer zigzag.

2.8. Método de Newton

El método de Newton utiliza la dirección:

$$-\nabla f(x_k) = \nabla^2 f(x_k) d_k \quad (9)$$

y actualiza:

$$x_{k+1} = x_k + \alpha_k d_k \quad (10)$$

Una de las posibles implementaciones es el cálculo de la inversa de la Hessiana pero como esto es computacionalmente costoso decidimos utilizar la variante de resolver el sistema definido por la igualdad anterior.

3. Diseño Experimental

3.1. Configuración de Experimentos

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

3.2. Puntos Iniciales

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., $(30, -5)$, $(-3, 40)$)

3.3. Parámetros del Algoritmo

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

3.4. Métricas de Evaluación

Para cada experimento se registraron:

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

4. Resultados

4.1. Resultados Generales

4.1.1. Tasa de Éxito

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)

4.1.2. Número de Iteraciones

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

4.2. Análisis de Convergencia

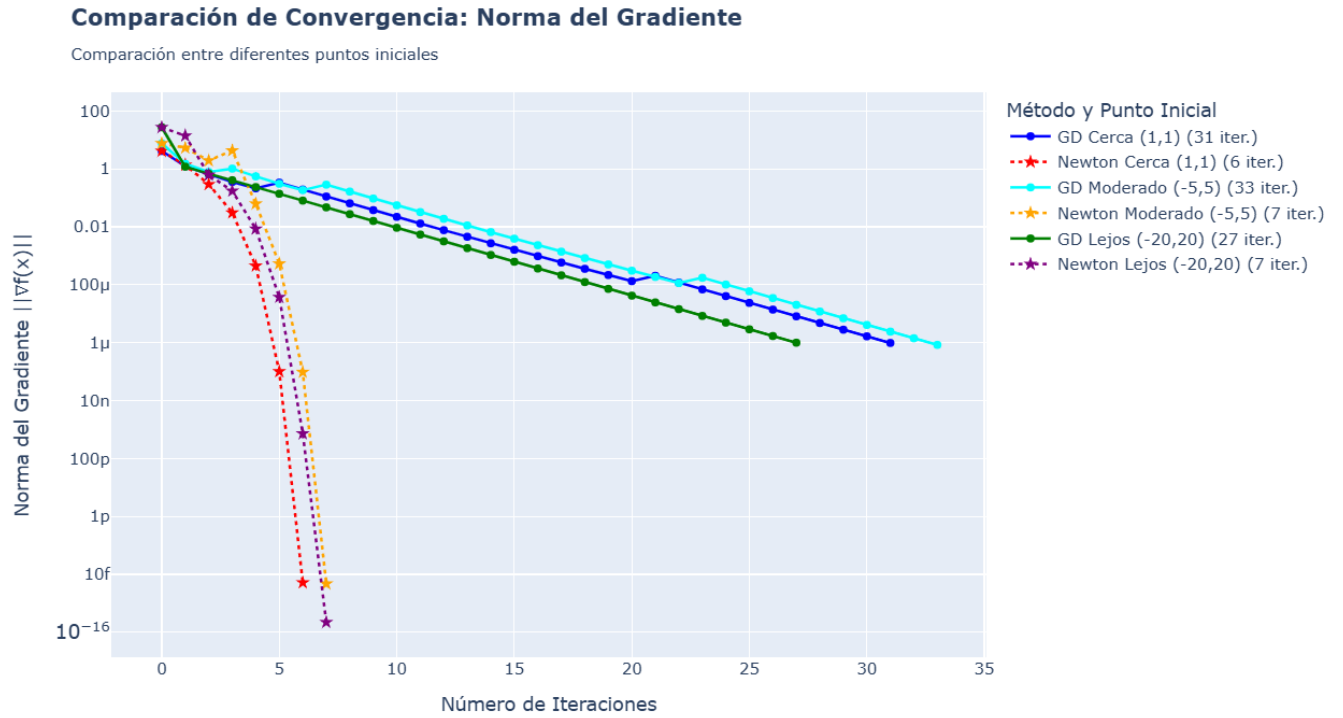


Figura 1: Comparación de la norma del gradiente vs iteraciones para diferentes puntos iniciales.

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 cuadráticamente si el punto inicial está cerca del óptimo, observemos que nuestra apreciación inicial de la vecindad del origen fue correcta.

4.3. Análisis Visual de Trayectorias

Animación: Convergencia de los Métodos de Optimización

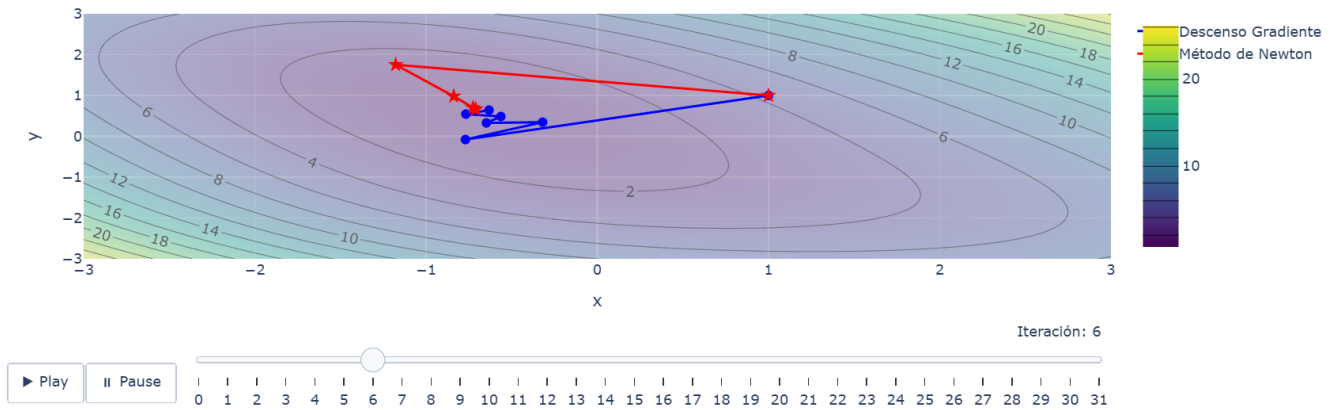


Figura 2: Trayectorias de convergencia en el espacio de búsqueda.

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 iteración ya Newton ha logrado la convergencia mientras que Descenso Máximo necesita 31 iteraciones en total.

4.4. Impacto de la Distancia Inicial

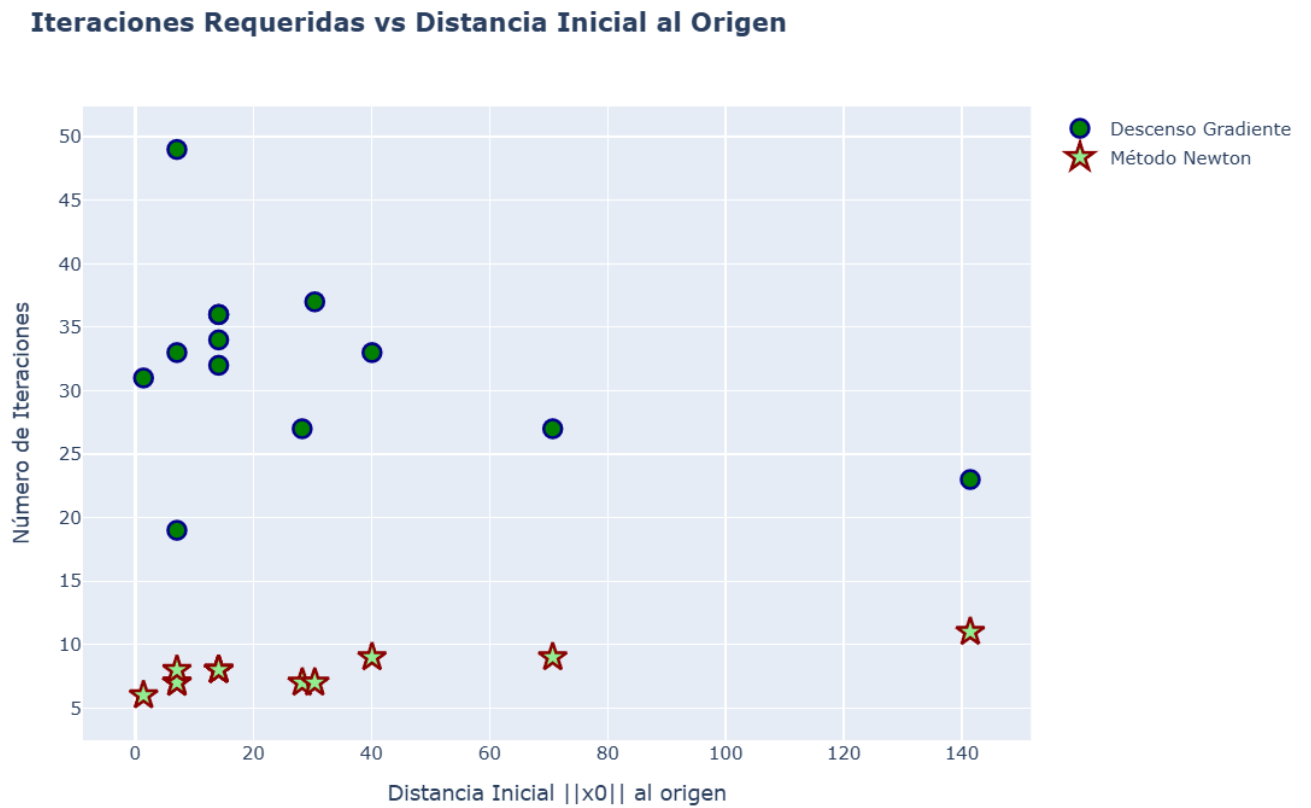


Figura 3: Relación entre la distancia inicial al óptimo y el número de iteraciones requeridas. Newton muestra crecimiento logarítmico mientras que el descenso del gradiente muestra crecimiento irregular.

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

5. Conclusiones

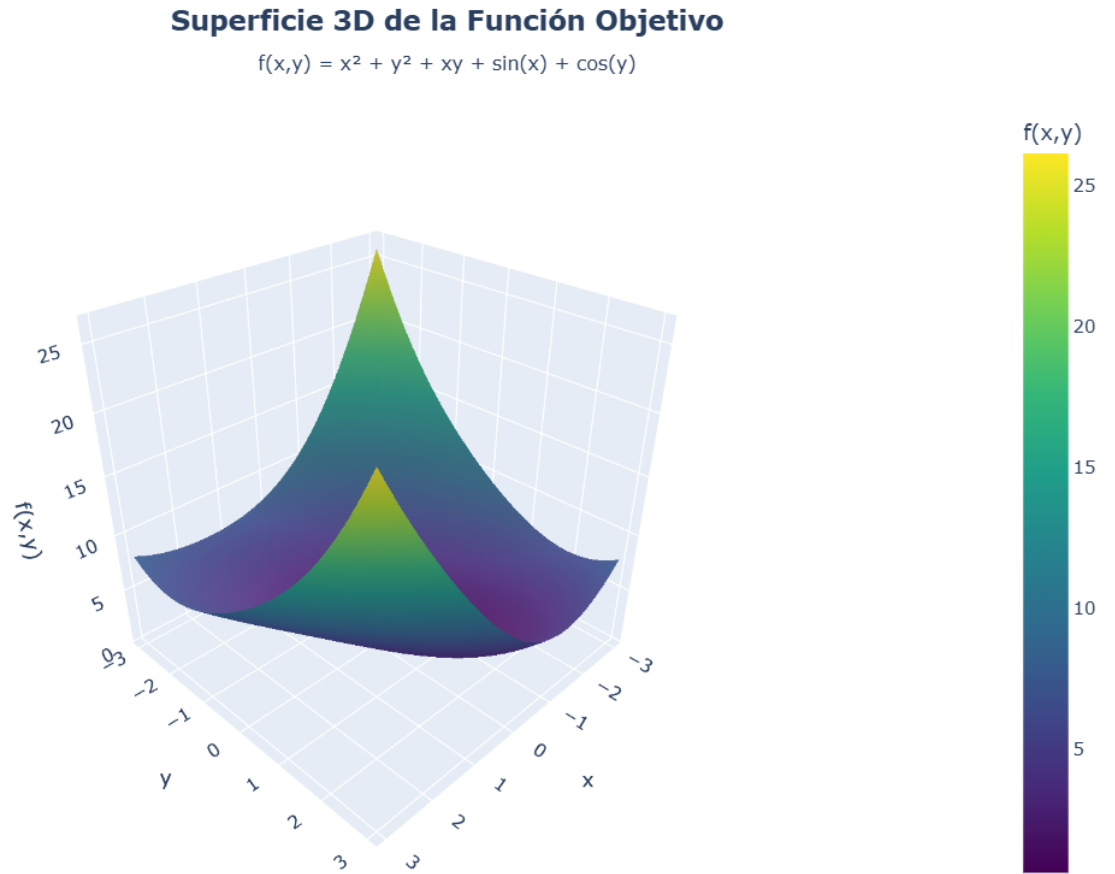


Figura 4: Superficie 3D de la función objetivo.

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 está aproximado con gran exactitud por los resultados 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:

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

5.1. Notas Finales

Se hizo uso consciente de la IA para la generación supervisada del código referente a la graficación en plotly así como para general la plantilla (sin contenido) de este informe.

5.2. Recomendaciones

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

Notar los cambios no monótonos de las iteraciones por categoría 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 y otros factores que pueden influir en la convergencia.

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>