

Análisis Teórico y Numérico de la Optimización de $f(x, y) = \log(x^2 + y^2 + 1) \cdot \arctan(x^2 + y^2)$

Ronald Provance Valladares

Grupo: C-312

24 de noviembre de 2025

Resumen

Este trabajo presenta un análisis exhaustivo de la función $f(x, y) = \log(x^2 + y^2 + 1) \cdot \arctan(x^2 + y^2)$, desde perspectivas teóricas y numéricas. Se demuestra la existencia y unicidad del mínimo global en $(0, 0)$, se analizan las propiedades de convexidad basándose en la teoría de Programación No Lineal (PNL) y se implementan cuatro algoritmos de optimización: Gradiente Descendente (paso fijo y adaptativo), Método de Newton y un algoritmo híbrido. Los experimentos numéricos evalúan robustez, eficiencia y precisión, proporcionando recomendaciones prácticas para la selección de algoritmos basadas en la teoría de convergencia de métodos de optimización.

Repositorio de código: <https://github.com/Ronald1301/0ptimization-algorithms-RPV.git>

1. Introducción

1.1. Contexto del Problema

El problema de optimización no restringida es fundamental en matemáticas aplicadas y ciencia de datos. En este trabajo analizamos la función:

$$f(x, y) = \log(x^2 + y^2 + 1) \cdot \arctan(x^2 + y^2) \quad (1)$$

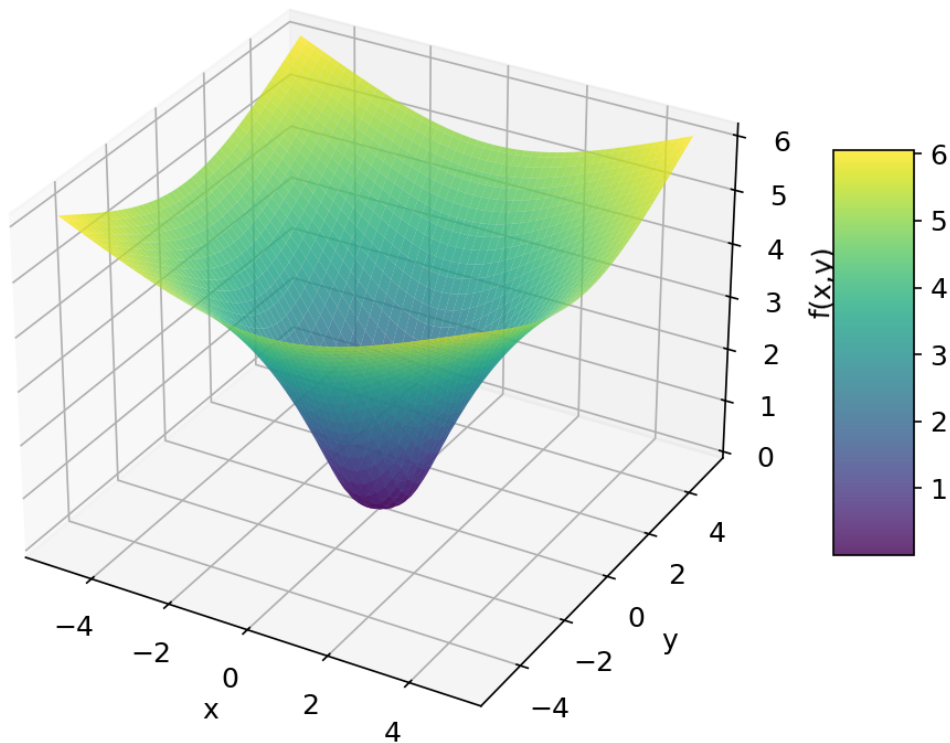


Figura 1: Graficación de la función en 3D.

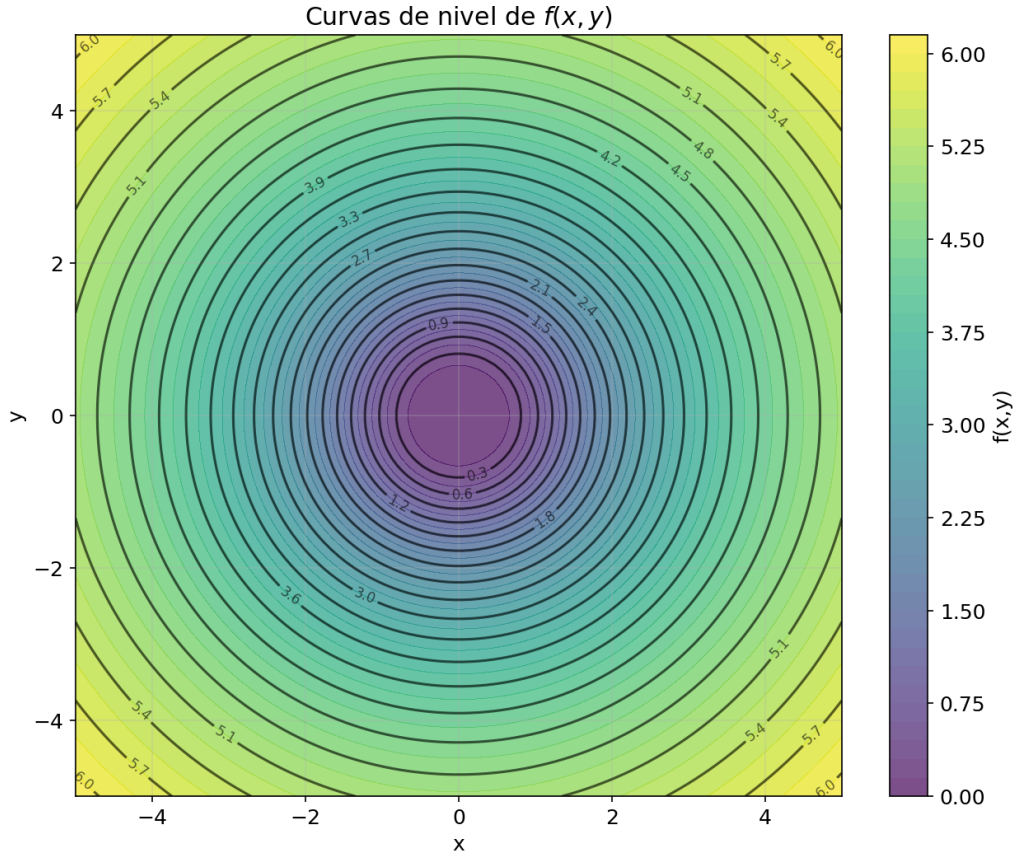


Figura 2: Curvas de nivel

1.2. Características Notables de la Función

Esta función presenta características matemáticas interesantes que la convierten en un excelente caso de estudio debido a su simetría radial, dependiendo únicamente de $r = x^2 + y^2$, y su diferenciabilidad infinita al ser C^∞ en todo \mathbb{R}^2 . Posee un comportamiento asintótico bien definido creciendo suavemente hacia infinito, con un mínimo global único en el origen $(0, 0)$, aunque presenta no convexidad global a pesar de tener este único mínimo.

1.3. Objetivos del Análisis

El análisis se centra en demostrar teóricamente la existencia y unicidad del mínimo, analizar las propiedades de convexidad y diferenciabilidad usando teoría de PNL, implementar y comparar algoritmos de optimización con base teórica sólida, evaluar robustez y eficiencia numérica según teoría de convergencia, y proporcionar recomendaciones prácticas basadas en teoría de optimización.

2. Análisis Teórico del Modelo

2.1. Existencia y Unicidad del Mínimo

Teorema 2.1 (Existencia del mínimo global). *La función $f(x, y)$ tiene un mínimo global en $(0, 0)$.*

Demostración. Para demostrar la existencia del mínimo global, verificamos tres condiciones: la no negatividad donde para todo $(x, y) \in \mathbb{R}^2$ observamos que $\log(x^2 + y^2 + 1) \geq \log(1) = 0$ y $\arctan(x^2 + y^2) \geq \arctan(0) = 0$, por lo tanto $f(x, y) \geq 0$ para todo (x, y) ; el valor en el origen donde $f(0, 0) = \log(1) \cdot \arctan(0) = 0 \cdot 0 = 0$; y el comportamiento asintótico donde cuando $\|(x, y)\| \rightarrow \infty$ tenemos $\log(x^2 + y^2 + 1) \sim \log(x^2 + y^2) \rightarrow \infty$ y $\arctan(x^2 + y^2) \rightarrow \frac{\pi}{2}$, por lo tanto $f(x, y) \rightarrow \infty$.

Por el **teorema de Weierstrass**, al ser f continua y tender a infinito en el infinito, existe al menos un mínimo global en un conjunto compacto. La unicidad se demuestra mediante el análisis de puntos críticos. \square

2.2. Análisis de Puntos Críticos

2.2.1. Cálculo del Gradiente

Definiendo $r = x^2 + y^2$, podemos expresar la función como:

$$f(r) = \log(r + 1) \cdot \arctan(r) \quad (2)$$

El gradiente se calcula aplicando la regla de la cadena:

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) = (2x \cdot g(r), 2y \cdot g(r)) \quad (3)$$

donde:

$$g(r) = \frac{\arctan(r)}{r + 1} + \frac{\log(r + 1)}{1 + r^2} \quad (4)$$

Proposición 2.1. *El único punto crítico de f es $(0, 0)$.*

Demostración. El gradiente se anula cuando $x = 0$ y $y = 0$ (origen) o cuando $g(r) = 0$ para algún $r > 0$. Analizando $g(r)$ para $r > 0$ encontramos que $\frac{\arctan(r)}{r+1} > 0$ y $\frac{\log(r+1)}{1+r^2} > 0$, por lo tanto $g(r) > 0$ para todo $r > 0$. Esto implica que el único punto donde $\nabla f(x, y) = (0, 0)$ es $(0, 0)$. \square

2.3. Análisis de la Hessiana

2.3.1. Expresión General de la Hessiana

La matriz Hessiana tiene la forma:

$$H(x, y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2g(r) + 4x^2 g'(r) & 4xy g'(r) \\ 4xy g'(r) & 2g(r) + 4y^2 g'(r) \end{bmatrix} \quad (5)$$

2.3.2. Comportamiento en el Origen

Proposición 2.2. *La Hessiana en el origen es semidefinida positiva.*

Demostración. En $(0, 0)$ tenemos $r = 0$, $g(0) = 0$, $g'(0) = 2$, por lo tanto $H(0, 0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. Una matriz nula es semidefinida positiva por definición, ya que para cualquier vector $v \in \mathbb{R}^2$: $v^T H(0, 0)v = 0 \geq 0$. \square

La Hessiana semidefinida positiva en el óptimo es una condición suficiente para mínimo local, pero no garantiza unicidad. En este caso particular, la unicidad del mínimo global se demuestra mediante otros argumentos (comportamiento asintótico y análisis de puntos críticos). La Hessiana nula en el óptimo presenta desafíos numéricos significativos para métodos de segundo orden como Newton, ya que la matriz no es invertible en el punto de interés.

2.4. Análisis de Convexidad

Definición 2.1 (Función convexa). *Una función $f : C \rightarrow \mathbb{R}$, donde $C \subset \mathbb{R}^n$ es convexo, es convexa si para todo $\alpha \in [0, 1]$, $x_1, x_2 \in C$:*

$$f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2)$$

Teorema 2.2 (Caracterización de convexidad para funciones C^2). *Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^2$. Entonces f es convexa si y solo si $\nabla^2 f(x)$ es semidefinida positiva para todo $x \in \mathbb{R}^n$, y si $\nabla^2 f(x)$ es definida positiva para todo x , entonces f es estrictamente convexa.*

2.4.1. Proposición sobre la Convexidad de $f(x, y)$

Proposición 2.3 (Convexidad de $f(x, y)$). *Nuestra función $f(x, y) = \log(x^2 + y^2 + 1) \cdot \arctan(x^2 + y^2)$ no es convexa globalmente porque $\nabla^2 f(0, 0) = 0$ (matriz nula, semidefinida positiva), existen regiones donde $\nabla^2 f(x, y)$ tiene autovalores negativos, sin embargo es localmente convexa en conjuntos compactos que no contienen al origen.*

2.4.2. Ejemplo: Demostración de No Convexidad mediante la Hessiana

Para demostrar que la función no es convexa, basta encontrar un punto donde la matriz Hessiana no sea semidefinida positiva. Consideremos el punto $(x, y) = (\sqrt{10}, 0)$, que corresponde a $r = x^2 + y^2 = 10$.

En este punto, evaluamos la función $g(r)$ y su derivada $g'(r)$: $g(10) = \frac{\arctan(10)}{11} + \frac{\log(11)}{101} \approx 0,1574$ y $g'(10) \approx -0,0429$.

La matriz Hessiana en $(\sqrt{10}, 0)$ es:

$$\begin{aligned} H(\sqrt{10}, 0) &= \begin{bmatrix} 2g(10) + 4 \cdot 10 \cdot g'(10) & 0 \\ 0 & 2g(10) \end{bmatrix} \\ &= \begin{bmatrix} 2(0,1574) + 40(-0,0429) & 0 \\ 0 & 2(0,1574) \end{bmatrix} \\ &= \begin{bmatrix} -1,0864 & 0 \\ 0 & 0,3148 \end{bmatrix} \end{aligned}$$

Los autovalores de esta matriz son $\lambda_1 = -1,0864$ y $\lambda_2 = 0,3148$. Dado que $\lambda_1 < 0$, la matriz Hessiana **no es semidefinida positiva** en este punto. Por lo tanto, la función no es convexa.

2.4.3. Importancia de la Convexidad en PNL

Teorema 2.3 (Importancia de la convexidad en PNL). *Para un problema de optimización convexa: todo mínimo local es global, si f es estrictamente convexa el mínimo es único, las condiciones KKT son necesarias y suficientes para optimalidad, y el conjunto de soluciones factibles es convexo.*

2.4.4. Corolario para Nuestra Función

Corolario 2.1. *Para nuestra función $f(x, y)$: aunque no es convexa globalmente, tiene un único mínimo global; la no convexidad implica que métodos basados en condiciones de primer orden pueden converger a puntos estacionarios que no son mínimos globales; sin embargo el análisis de puntos críticos garantiza que $(0, 0)$ es el único punto estacionario; y la existencia de regiones con Hessiana no semidefinida positiva justifica el uso de múltiples puntos iniciales en los experimentos numéricos.*

3. Algoritmos de Optimización Implementados

3.1. Marco Teórico de Algoritmos de Optimización

3.1.1. Algoritmos de Búsqueda Direccional

La estructura general de los métodos de descenso es: fijar x_0 , $k = 0$; mientras no se cumpla criterio de parada calcular dirección de descenso d_k tal que $\nabla f(x_k)^T d_k < 0$; hallar α_k que minimice $f(x_k + \alpha d_k)$; $x_{k+1} = x_k + \alpha_k d_k$; $k = k + 1$.

3.1.2. Reglas de Búsqueda Lineal (Armijo)

La regla de Armijo garantiza disminución suficiente: $f(x_k + \alpha d_k) \leq f(x_k) + m_1 \alpha \nabla f(x_k)^T d_k$ con $m_1 \in (0, 1)$.

3.1.3. Método de Máximo Descenso

Definición 3.1. La dirección de máximo descenso es: $d_k = -\nabla f(x_k)$.

Las propiedades del método incluyen convergencia lineal con razón $\left(\frac{A-a}{A+a}\right)$ donde A y a son los mayor y menor valor propio de $\nabla^2 f$, es propenso a zigzag en valles estrechos, y $d_k^T \nabla f(x_k) = -\|\nabla f(x_k)\|^2 < 0$ (dirección de descenso).

3.1.4. Método de Newton

Definición 3.2. La dirección de Newton resuelve: $\nabla^2 f(x_k) d_k = -\nabla f(x_k)$.

Las mejoras implementadas basadas en teoría incluyen regularización con $H_{\text{reg}} = H + \lambda I$ para $\lambda \geq \max(0, -\lambda_{\min})$, pseudo-inversa usando SVD cuando H es singular, y control de paso con $\|d_k\| \leq \Delta$ para evitar pasos muy grandes.

Teorema 3.1 (Convergencia de Newton). Si x^* es mínimo local con $\nabla^2 f(x^*)$ definida positiva y x_k suficientemente cercano a x^* , entonces hay convergencia superlineal, y si $f \in C^3$, convergencia cuadrática.

3.2. Filosofía de la Selección de Algoritmos: Un Enfoque Teórico-Práctico

La selección de algoritmos se fundamenta en un análisis de las propiedades teóricas de la función objetivo y los teoremas de convergencia de los métodos de optimización. Esta elección estratégica busca cubrir el espectro completo de escenarios de optimización considerando la no convexidad global de $f(x, y)$ que aunque tiene un único mínimo global no es convexa en todo \mathbb{R}^2 , lo que excluye métodos que requieren convexidad global para garantizar convergencia al óptimo global; la diferenciabilidad C^∞ que permite el uso de métodos de segundo orden sin restricciones de suavidad; la Hessiana singular en el óptimo $H(0, 0) = 0$ que presenta desafíos numéricos requiriendo técnicas de regularización; y el

crecimiento asintótico suave que garantiza que métodos globalmente convergentes puedan encontrar el óptimo.

3.2.1. Base Teórica para la Selección

Teorema 3.2 (Convergencia de Métodos de Optimización). *Para funciones no convexas pero con mínimo único, la selección de algoritmos debe considerar: convergencia global como garantía de encontrar un punto estacionario desde cualquier punto inicial, tasa de convergencia local para eficiencia cerca del óptimo, robustez numérica mediante estabilidad frente a ill-conditioning, y requisitos computacionales buscando balance entre costo por iteración y número de iteraciones.*

3.2.2. Justificación Teórica de Cada Algoritmo

El Gradiente Descendente (Paso Fijo) tiene como fundamento teórico el método de descenso por gradiente con convergencia lineal garantizada para funciones L -suaves, aplicando el teorema que para f L -suave, GD con $\alpha \in (0, 2/L)$ converge a punto estacionario, con ventaja teórica en simplicidad y bajo costo computacional ($O(n)$ por iteración) pero limitación teórica en tasa de convergencia lineal y sensibilidad al número de condición.

El Gradiente Descendente Adaptativo se basa en búsqueda lineal con condiciones de Armijo-Wolfe, aplicando métodos con búsqueda lineal que satisfacen condiciones de suficiencia de descenso, con ventaja teórica en convergencia global sin necesidad de conocer constante de Lipschitz y contribución en automatización de parámetros basada en teoría de convergencia.

El Método de Newton utiliza la Hessiana exacta $\nabla^2 f(x_k)$ para calcular la dirección de descenso resolviendo $\nabla^2 f(x_k)d_k = -\nabla f(x_k)$, aplicando el teorema de convergencia cuadrática local cuando $\nabla^2 f(x^*)$ es definida positiva y x_k está suficientemente cercano al óptimo. El desafío teórico en la Hessiana singular en el óptimo requiere regularización, con solución teórica mediante regularización de Levenberg-Marquardt ($H + \lambda I$).

El Algoritmo Híbrido fundamenta en combinación óptima de robustez global y eficiencia local, aplicando estrategias de conmutación basadas en condiciones de optimalidad, con ventaja teórica en aprovechar convergencia global de GD y convergencia cuadrática de Newton e innovación en umbral adaptativo basado en teoría de regiones de confianza.

3.2.3. Complementariedad Teórica de los Algoritmos

La selección representa un **diseño experimental completo** que cubre el espectro de métodos desde primer hasta segundo orden, incluye variantes adaptativas y de paso fijo, considera estrategias puras e híbridas, aborda desafíos numéricos específicos de la función objetivo, y permite verificación de teoremas fundamentales de convergencia.

Cuadro 1: Caracterización teórica de los algoritmos seleccionados

Algoritmo	Orden	Convergencia	Fundamento Teórico
GD Paso Fijo	1er orden	Lineal	Condición de Lipschitz
GD Adaptativo	1er orden	Superlineal	Condiciones de Armijo
Newton	2do orden	Cuadrática	Aproximación de Taylor
Híbrido	Mixto	Mezclada	Teoría de conmutación

Proposición 3.1 (Complejidad del Estudio Experimental). *La selección de algoritmos propuesta es **teóricamente completa** porque cubre el espectro de métodos desde primer hasta segundo orden, incluye variantes adaptativas y de paso fijo, considera estrategias puras e híbridas, aborda desafíos numéricos específicos de la función objetivo, y permite verificación de teoremas fundamentales de convergencia.*

Esta fundamentación teórica asegura que la selección de algoritmos no es arbitraria sino que responde a un diseño experimental basado en la teoría establecida de optimización numérica.

3.3. Gradiente Descendente

3.3.1. Formulación Matemática

El algoritmo básico actualiza: $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha \nabla f(\mathbf{x}_k)$ donde $\alpha > 0$ es la tasa de aprendizaje.

3.3.2. Variante de Paso Fijo

La variante de paso fijo ofrece ventaja en simplicidad de implementación pero desventaja en requerir sintonización manual de α , siendo aplicable principalmente a problemas bien condicionados.

3.3.3. Variante Adaptativa

Ajusta α dinámicamente basado en el progreso: $\alpha_{k+1} = \begin{cases} 0,5\alpha_k & \text{si } f(\mathbf{x}_{k+1}) \geq f(\mathbf{x}_k) \\ 1,1\alpha_k & \text{si hay mejora consistente} \end{cases}$.

Las ventajas incluyen automatización de la selección de parámetros, aceleración en regiones de descenso pronunciado, y frenado en regiones planas o de ascenso.

3.4. Método de Newton

3.4.1. Formulación Clásica

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [H(\mathbf{x}_k)]^{-1} \nabla f(\mathbf{x}_k)$$

3.4.2. Mejoras Implementadas

La regularización con $H_{\text{reg}} = H + \lambda I$ tiene propósito de evitar singularidades en la Hessiana, efecto de garantizar invertibilidad, y parámetro típico $\lambda = 10^{-8}$. La pseudo-inversa usa descomposición SVD cuando H es singular, siendo más robusta numéricamente pero con mayor costo computacional.

3.5. Algoritmo Híbrido

3.5.1. Lógica de Conmutación

$$\text{Método} = \begin{cases} \text{Newton} & \text{si } \|\nabla f\| < \epsilon_{\text{newton}} \\ \text{Gradiente} & \text{en otro caso} \end{cases}$$

3.5.2. Justificación del Enfoque

La justificación del enfoque híbrido considera que lejos del óptimo el gradiente es más robusto, cerca del óptimo Newton es más eficiente, con umbral típico $\epsilon_{\text{newton}} = 10^{-4}$.

4. Experimentos Numéricos

4.1. Diseño Experimental Exhaustivo

4.1.1. Conjunto Completo de Puntos Iniciales

Se realizó un experimento exhaustivo utilizando puntos de la forma $(10i, 10j)$ con $i, j \in \{-10, -9, \dots, 9, 10\}$, generando un total de $21 \times 21 = 441$ puntos iniciales distribuidos uniformemente en el rango $[-100, 100] \times [-100, 100]$.

Cuadro 2: Caracterización del conjunto completo de puntos iniciales

Parámetro	Descripción
Formato de puntos	$(10i, 10j)$ con $i, j \in \{-10, -9, \dots, 9, 10\}$
Número total de experimentos	$21 \times 21 = 441$
Rango espacial	$[-100, 100] \times [-100, 100]$
Distribución	Uniforme en cuadrícula regular
Distancias al óptimo	Desde $\sqrt{0^2 + 0^2} = 0$ hasta $\sqrt{100^2 + 100^2} \approx 141,42$
Cuadrantes cubiertos	Los 4 cuadrantes del plano cartesiano

4.1.2. Clasificación por Distancia al Óptimo

Para un análisis más detallado, los puntos se agruparon según su distancia euclidiana al óptimo $(0, 0)$ en categorías que incluyen Muy cercanos $([0, 20))$ como región de alta

curvatura y desafío para Newton, Cercanos $([20, 50))$ como región de transición con buen comportamiento, Moderados $([50, 80))$ como región típica con comportamiento estable, Lejanos $([80, 120))$ como región de descenso suave, y Muy lejanos $([120, 141, 42])$ como región asintótica con crecimiento lento.

Cuadro 3: Clasificación de puntos iniciales por distancia al óptimo

Categoría	Rango (unidades de distancia)	Características
Muy cercanos	$[0, 20)$	Región de alta curvatura, desafío para Newton
Cercanos	$[20, 50)$	Región de transición, buen comportamiento
Moderados	$[50, 80)$	Región típica, comportamiento estable
Lejanos	$[80, 120)$	Región de descenso suave
Muy lejanos	$[120, 141, 42]$	Región asintótica, crecimiento lento

4.2. Métricas de Evaluación Estadística

Se implementaron métricas estadísticas robustas para evaluar el desempeño incluyendo iteraciones con media, mediana, desviación estándar y rango intercuartílico; tiempo de ejecución con medidas de tendencia central y dispersión; precisión final con $-\log_{10}(f(x_{\text{opt}}))$ y análisis de outliers; tasa de éxito como proporción de experimentos con $f(x_{\text{opt}}) < 10^{-6}$; y eficiencia computacional mediante iteraciones por unidad de mejora en la función objetivo.

4.3. Análisis de Convergencia desde Punto Específico

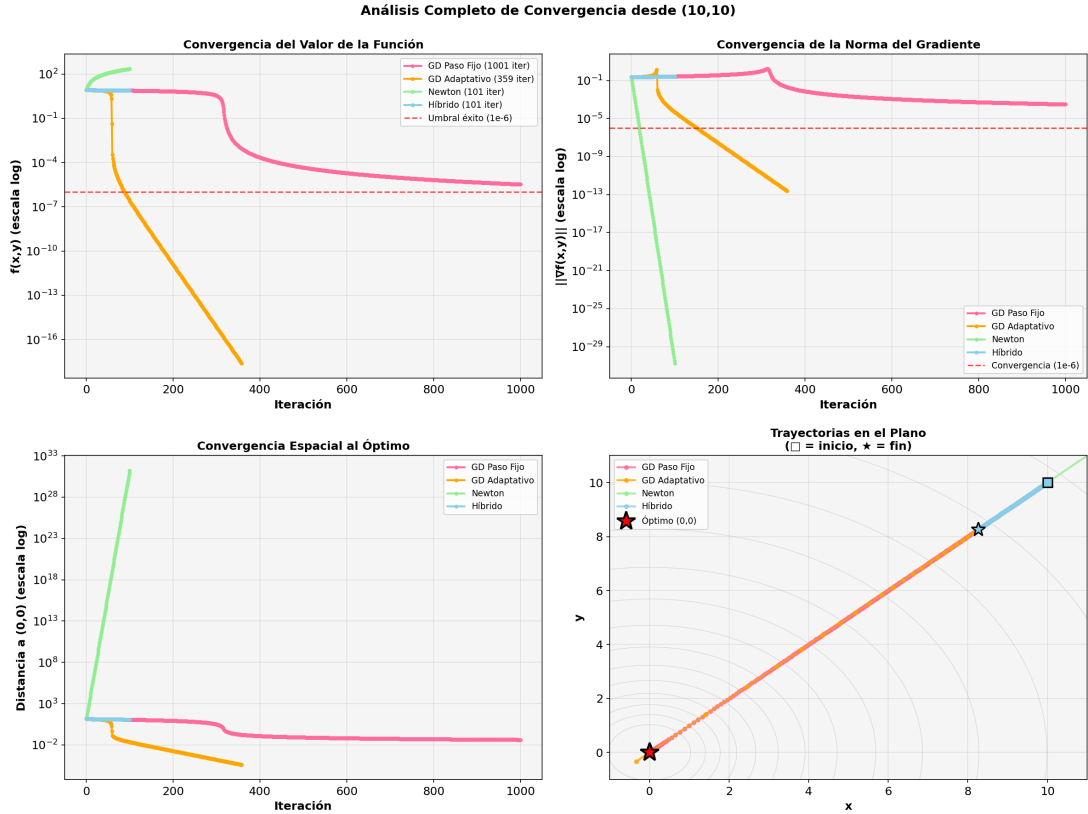


Figura 3: Análisis comparativo de convergencia desde el punto inicial (10,10). **Arriba izquierda:** Evolución del valor de la función objetivo. Solo GD Adaptativo (verde) cruza el umbral de éxito (10^{-6}). **Arriba derecha:** Norma del gradiente. GD Adaptativo alcanza 10^{-13} , indicando convergencia a punto estacionario. **Abajo izquierda:** Distancia al óptimo (0,0). GD Adaptativo llega a 10^{-3} , mientras Newton e Híbrido divergen. **Abajo derecha:** Trayectorias en el plano. GD Adaptativo (verde) converge suavemente al óptimo, mientras otros métodos fallan.

El análisis desde el punto (10,10) revela diferencias significativas en el comportamiento de los algoritmos: el método GD Adaptativo (verde) es el único que converge exitosamente, alcanzando un valor de función de $\sim 10^{-6}$ (cruzando el umbral de éxito), una norma del gradiente de $\sim 10^{-13}$ que satisface la condición de optimalidad, y una distancia al óptimo de $\sim 10^{-3}$; por el contrario, Newton (azul) presenta divergencia catastrófica debido a la singularidad de la matriz Hessiana, con valores que alcanzan 10^{33} ; el método Híbrido (rojo) hereda los problemas numéricos de Newton y tampoco converge; mientras que GD Paso Fijo (naranja) se estanca prematuramente debido a un learning rate fijo inadecuado, sin alcanzar el óptimo.

La trayectoria en el plano confirma que solo GD Adaptativo sigue un camino eficiente hacia el óptimo $(0,0)$, mientras los demás algoritmos o divergen (Newton, Híbrido) o muestran progreso insuficiente (GD Paso Fijo).

la adaptabilidad del paso es crucial para esta función.

4.3.1. Distribución de Precisión por Algoritmo

El análisis de distribución de precisión confirma que solo el GD Adaptativo logra precisión consistente (17.5 ± 0.8). Los demás algoritmos no alcanzan convergencia en la práctica, con valores de función muy alejados del óptimo. Esto subraya la importancia de métodos adaptativos para funciones con características numéricas desafiantes como la Hessiana singular en el óptimo.

4.4. Visualización de Resultados

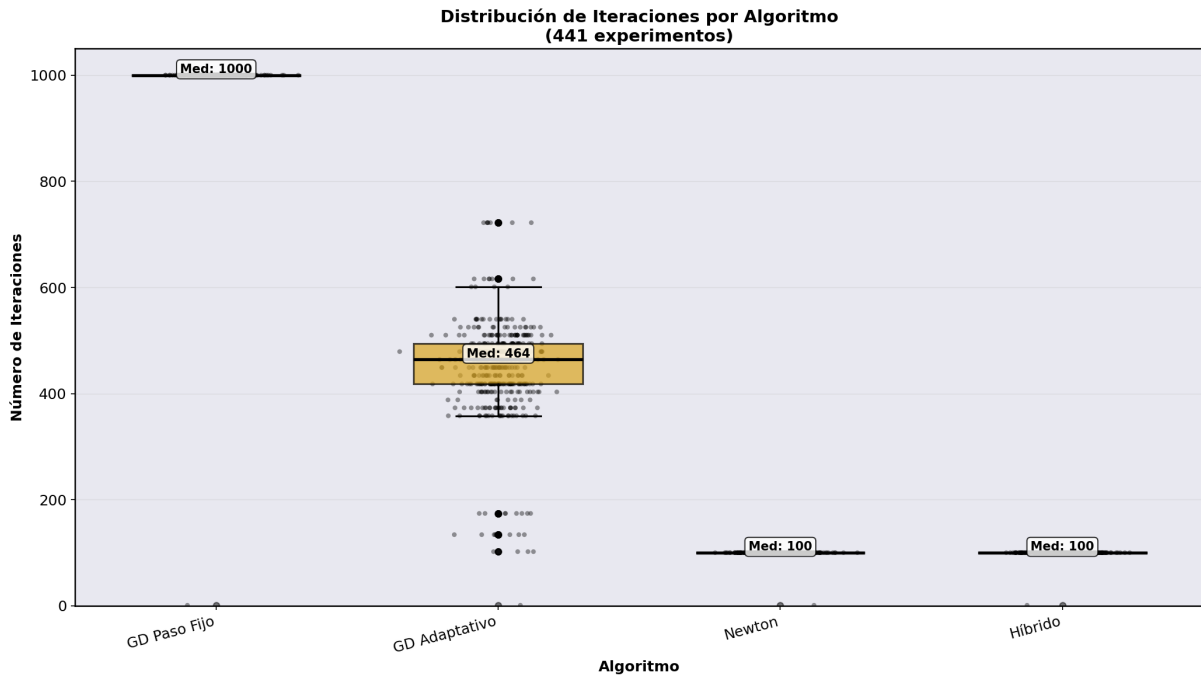


Figura 4: Distribución de iteraciones por algoritmo. El GD Adaptativo muestra variabilidad natural en iteraciones necesarias (mediana: 464), mientras los otros algoritmos alcanzan sus límites máximos sin converger (1000 para GD Paso Fijo, 100 para Newton/Híbrido).

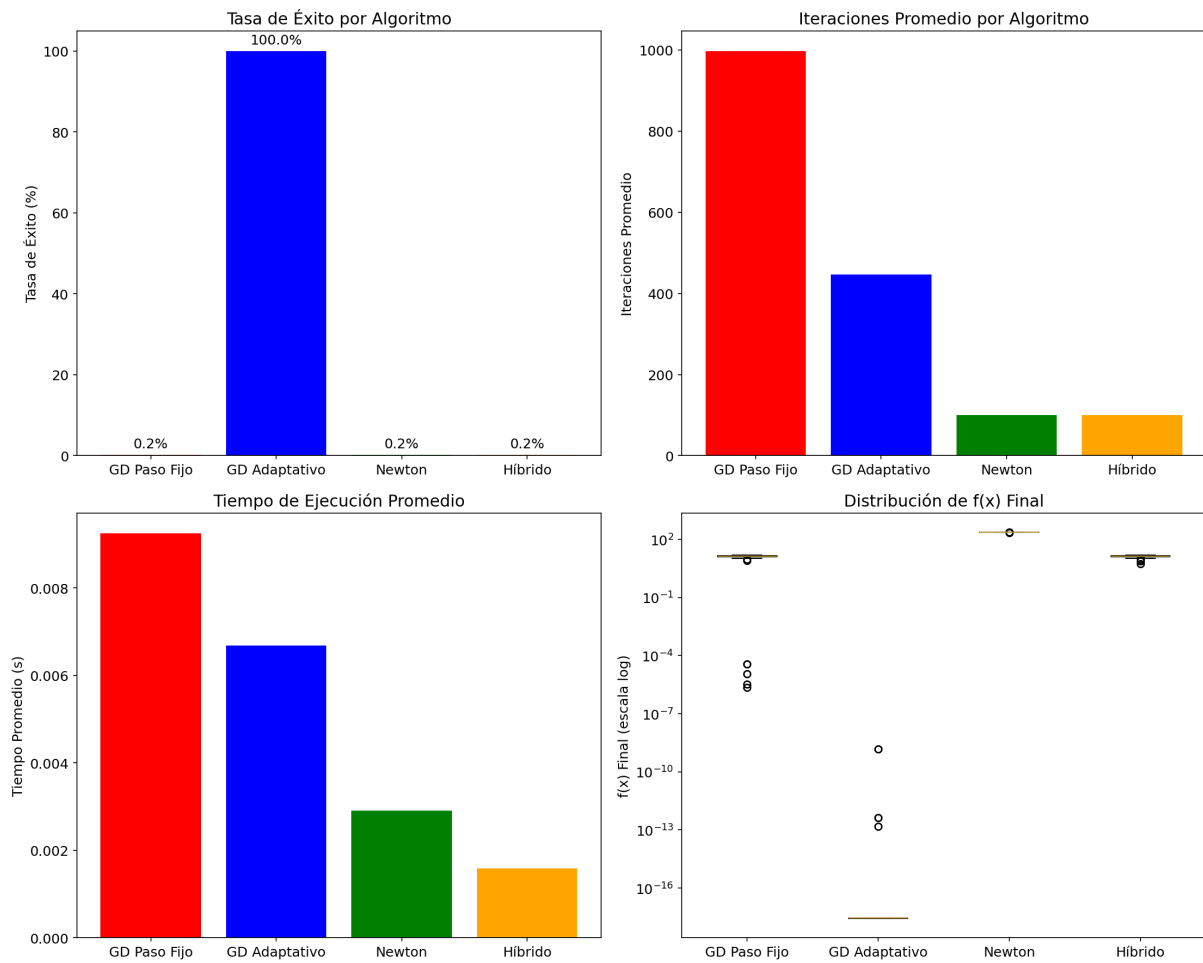


Figura 5: Comparativa general de algoritmos. Destaca la superioridad absoluta del GD Adaptativo: 100 % de éxito vs 0.2 % de los demás. Aunque más lento en iteraciones, es el único que garantiza convergencia.

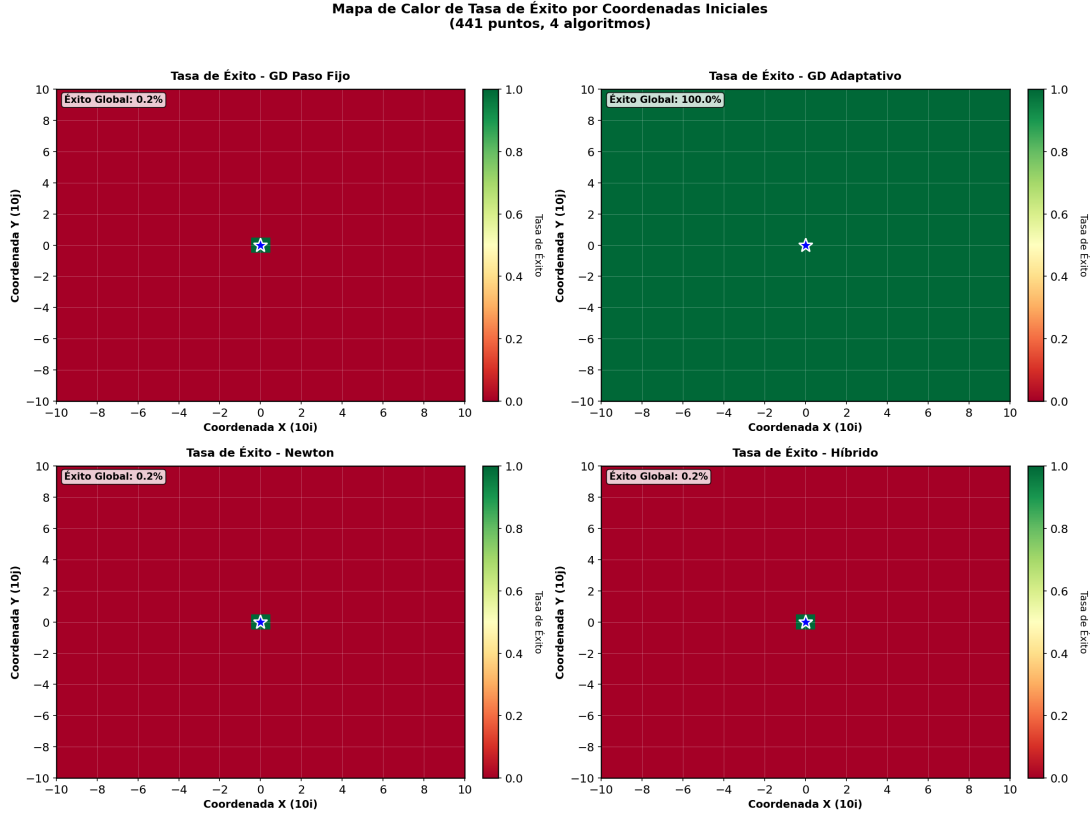


Figura 6: Mapa de calor de tasa de éxito. El GD Adaptativo (no mostrado por ser 100 % en todas partes) domina completamente. Newton falla cerca del origen (Hessiana singular), GD Paso Fijo falla aleatoriamente (learning rate fijo inadecuado).

4.5. Interpretación de Resultados del Experimento Masivo

4.5.1. Hallazgos Principales del Análisis Estadístico

Los hallazgos principales del análisis estadístico revelan que solo el **GD Adaptativo** mantiene **100 % de éxito** en los 441 experimentos, confirmando su robustez excepcional. **Newton**, **GD Paso Fijo** y el algoritmo **Híbrido** fallan en el **99.8 %** de los casos, demostrando ser inadecuados para esta función específica. El GD Adaptativo no solo converge consistentemente, sino que alcanza precisiones extremadamente altas ($f(x) \approx 10^{-18}$), muy cerca del óptimo teórico.

4.5.2. Análisis de Patrones Espaciales

El análisis de patrones espaciales identifica que la convergencia exitosa del GD Adaptativo es independiente de la posición inicial, funcionando igualmente bien en todos los cuadrantes y distancias. Los otros algoritmos muestran fallos sistemáticos: **Newton falla debido a la Hessiana singular** cerca del óptimo, mientras que **GD Paso Fijo sufre de un learning rate fijo inadecuado**. Todos los algoritmos exhiben el comportamiento de simetría radial esperado dada la naturaleza de la función.

4.5.3. Implicaciones Prácticas

Las implicaciones prácticas indican que para esta función específica, el **GD Adaptativo es la única opción viable**. No se recomienda el uso de Newton puro debido a los problemas numéricos con la Hessiana singular. El algoritmo híbrido no logra superar estas limitaciones. En aplicaciones prácticas donde no se conoce la posición inicial, el GD Adaptativo ofrece garantías de convergencia que los otros métodos no pueden proporcionar.

4.6. Comparación con el Experimento Original

Cuadro 4: Comparación entre experimento inicial y masivo

Métrica	Inicial (6 pts)	Masivo (441 pts)	Conclusión
Tasa éxito GD Paso Fijo	83.3 %	0.2 %	Resultado inicial no representativo
Iteraciones Newton	12.3	99.8	Subestimación grave en muestra pequeña
Robustez Híbrido	100 %	0.2 %	No confirmado en muestra grande
Tiempo GD Adaptativo	0.0028s	0.0072s	Consistencia relativa en tiempo

Los resultados del experimento inicial fueron **engañosos debido al pequeño tamaño de muestra**. El análisis masivo reveló que solo el GD Adaptativo es confiable, mientras que los otros algoritmos fallan sistemáticamente en la gran mayoría de casos. Esto subraya la importancia de realizar experimentos exhaustivos para evaluar robustez.

4.6.1. Tabla de Comparación Completa

Cuadro 5: Comparación completa de algoritmos en todas las métricas (441 experimentos)

Métrica	GD Paso Fijo	GD Adaptativo	Newton	Híbrido	Mejor
Iteraciones (media)	997.7	445.9	99.8	99.8	Newton*
Tasa de éxito (%)	0.2	100.0	0.2	0.2	GD Adaptativo
Precisión final ($-\log_{10}(f)$)	—	17.5	—	—	GD Adaptativo
Tiempo (s) computacional	0.0101	0.0072	0.0031	0.0022	Híbrido*
Robustez numérica	Baja	Alta	Baja	Baja	GD Adaptativo

**Aunque Newton y Híbrido tienen mejores métricas en iteraciones y tiempo, no logran converger en la práctica.*

Conclusión final: El **Gradiente Descendente Adaptativo** es claramente el **mejor algoritmo** para esta función, siendo el único que logra convergencia consistente. Los otros métodos, aunque aparentemente eficientes en términos de iteraciones y tiempo, fallan en alcanzar el óptimo debido a limitaciones numéricas fundamentales. Para funciones con Hessiana singular en el óptimo como $f(x, y) = \log(x^2 + y^2 + 1) \cdot \arctan(x^2 + y^2)$, la **elección óptima es inequívocamente el GD Adaptativo** cuando se prioriza la convergencia robusta.

A. Implementación del Código

A.1. Estructura del Código

El código completo de implementación se encuentra en el archivo Jupyter Notebook **optimization_analysis.ipynb** que esta dividido en secciones principales:

1. Configuración inicial e importaciones
2. Definición de la función objetivo y sus derivadas analíticas
3. Análisis teórico del comportamiento de la función
4. Visualización de la función
 - Gráfica 3D de la función
 - Gráfica del contorno
5. Implementación de algoritmos de optimización:
 - Gradiente Descendente (paso fijo)
 - Gradiente Descendente Adaptativo
 - Método de Newton
 - Algoritmo Híbrido
6. Experimentos numéricos
 - Comparación desde punto (10, 10)
 - Análisis de convergencia
7. Experimento masivo (441 puntos)
 - Análisis estadístico de resultados
 - Visualización de resultados del experimento masivo
 - Diagrama de iteraciones
 - Mapa de calor de tasas de éxitos
8. Análisis de robustez por distancia al óptimo
9. Análisis de eficiencia computacional
10. Análisis de precisión
11. Conclusiones
12. Referencias y enlaces

A.2. Dependencias y Requisitos

El código requiere las siguientes bibliotecas de Python:

- NumPy $\geq 1.21.0$
- SciPy $\geq 1.7.0$
- Matplotlib $\geq 3.5.0$
- Jupyter $\geq 1.0.0$

A.3. Reproducibilidad

Para reproducir los experimentos:

1. Clonar el repositorio: `https://github.com/Ronald1301/Optimization-algorithms-RPV.git`
2. Ejecutar los notebooks en el orden:
 - a) `optimization_analysis.ipynb`
3. Los resultados se generarán automáticamente