

# Modelos de optimización

Mauricio Medina | Grupo 312

13 de noviembre de 2025

A lo largo de este documento, se estudiará un problema de minimización sin restricciones en  $\mathbb{R}^2$  cuyo objetivo es

$$f(x, y) = 4(x + y + xy)^2 + x^4.$$

Primero se presentará el análisis teórico del modelo (existencia y unicidad del óptimo global, gradiente, Hessiana, puntos estacionarios y clasificación local). Luego, se describen, desde el punto de vista teórico, dos algoritmos que se emplearán en la parte computacional: (i) descenso por gradiente con búsqueda de paso tipo Armijo y (ii) Newton amortiguado con regularización. Seguido de eso se observan gráficas y se analizan experimentos resultados de aplicar estos algoritmos en el modelo.

## Análisis teórico

Sea

$$f(x, y) = 4(x + y + xy)^2 + x^4, \quad (x, y) \in \mathbb{R}^2. \quad (0.1)$$

Introducimos la abreviatura

$$z = x + y + xy.$$

Con esta notación,

$$f(x, y) = (2z)^2 + (x^2)^2,$$

luego  $f$  es suma de cuadrados y por tanto  $f(x, y) \geq 0$  en todo  $\mathbb{R}^2$ , con

$$f(0, 0) = 0.$$

Para que  $f$  alcance la cota inferior 0 es necesario anular ambos sumandos:  $(x^2)^2 = 0$  implica  $x = 0$  y, entonces,  $z|_{x=0} = y$  impone  $y = 0$ . Es decir, el conjunto de ceros de  $f$  es exactamente  $\{(0, 0)\}$ .

**Proposición 1** (Óptimo global y unicidad). *El valor óptimo global de (0.1) es  $f^* = 0$ , alcanzado de forma única en  $(x^*, y^*) = (0, 0)$ .*

*Demostración.* De la descomposición  $f = (2z)^2 + (x^2)^2$  se sigue  $f \geq 0$ . Además,  $f(0, 0) = 0$ , y la discusión anterior muestra que  $f(x, y) = 0$  sólo ocurre en  $(0, 0)$ .  $\square$

Usando  $\partial z/\partial x = 1 + y$  y  $\partial z/\partial y = 1 + x$ , el gradiente de  $f$  es

$$\nabla f(x, y) = \begin{pmatrix} 8z(1 + y) + 4x^3 \\ 8z(1 + x) \end{pmatrix}.$$

La matriz Hessiana viene dada por

$$\nabla^2 f(x, y) = \begin{pmatrix} 12x^2 + 8(1 + y)^2 & 8((1 + x)(1 + y) + z) \\ 8((1 + x)(1 + y) + z) & 8(1 + x)^2 \end{pmatrix} = \begin{pmatrix} 12x^2 + 8(1 + y)^2 & 8 + 16z \\ 8 + 16z & 8(1 + x)^2 \end{pmatrix}.$$

En el origen,

$$\nabla f(0, 0) = \mathbf{0}, \quad \nabla^2 f(0, 0) = \begin{pmatrix} 8 & 8 \\ 8 & 8 \end{pmatrix},$$

cuyos autovalores son 16 y 0 (una dirección plana de orden superior).

Los puntos estacionarios se obtienen de  $\nabla f(x, y) = \mathbf{0}$ . De la segunda componente se deduce

$$8z(1 + x) = 0 \implies z = 0 \quad \text{o} \quad x = -1.$$

*Caso  $z = 0$ :* entonces la primera componente fuerza  $4x^3 = 0$ , de donde  $x = 0$  y, con  $z = 0$ , también  $y = 0$ .

*Caso  $x = -1$ :* aquí  $z = -1$  y la primera componente se reduce a  $8(-1)(1 + y) + 4(-1)^3 = 0$ , equivalente a  $-8(1 + y) - 4 = 0$ , cuyo resultado es  $y = -\frac{3}{2}$ .

**Proposición 2** (Puntos estacionarios). *Los únicos puntos estacionarios de  $f$  son  $(0, 0)$  y  $(-1, -\frac{3}{2})$ .*

Sobre la recta  $x = -t$ ,  $y = t$  se cumple  $z = -t^2$  y

$$f(-t, t) = 4t^4 + t^4 = 5t^4,$$

lo que confirma crecimiento de cuarto orden alrededor del origen. En consecuencia,  $(0, 0)$  es un mínimo local estricto (y además global por la Proposición 1).

En el punto  $(-1, -\frac{3}{2})$ ,

$$\nabla^2 f(-1, -\frac{3}{2}) = \begin{pmatrix} 14 & -8 \\ -8 & 0 \end{pmatrix}, \quad \det = -64 < 0,$$

por lo que la Hessiana es indefinida y  $(-1, -\frac{3}{2})$  es un *punto silla*.

*Observación* (Suavidad, convexidad y (no) coercividad). La función es polinómica y pertenece a  $C^\infty(\mathbb{R}^2)$ . No es convexa globalmente: la Hessiana evaluada en  $(-1, -\frac{3}{2})$  es indefinida. Tampoco es coerciva: sobre  $x = -1$ ,

$$f(-1, y) = 4(-1 + y - y)^2 + (-1)^4 = 5 \quad \text{para todo } y \in \mathbb{R},$$

de modo que  $\|(x, y)\| \rightarrow \infty$  no implica  $f(x, y) \rightarrow \infty$ . Tampoco existe máximo global ya que  $f(x, y) \rightarrow \infty$  cuando  $|x| \rightarrow \infty$  por el término  $x^4$ .

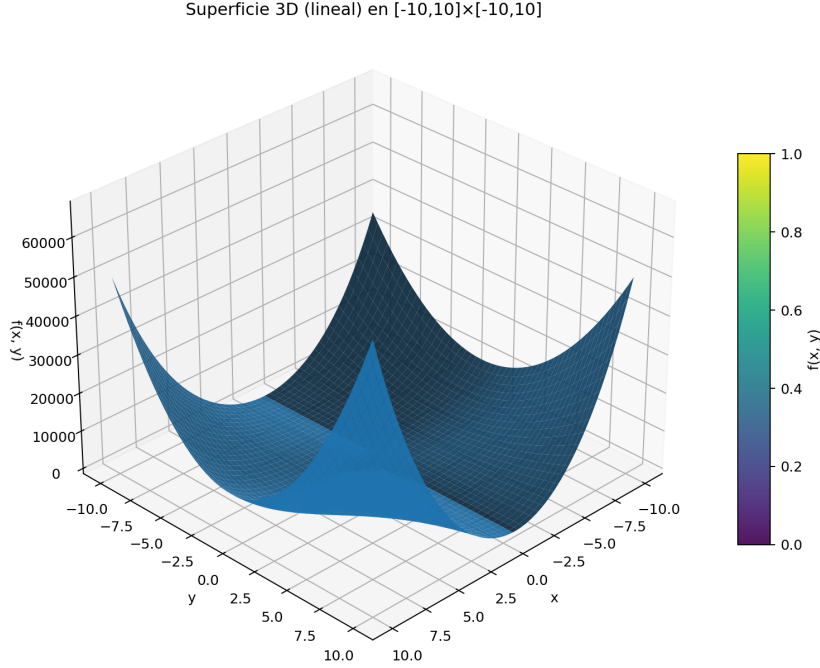


Figura 1: Superficie  $z = f(x, y)$ .

La Figura 1 muestra la superficie de la función representada sobre el dominio cuadrado  $[-10, 10] \times [-10, 10]$  con una escala de colores lineal. El título de la gráfica indica explícitamente este rango y el colorbar de la derecha normaliza los valores de  $f$  en el intervalo aproximado  $[0, 1]$ , de modo que los tonos más claros corresponden a mayores alturas de la superficie.

Geométricamente se observa un paisaje fuertemente dominado por el término  $x^4$ . A medida que  $|x|$  aumenta, la función crece muy rápido y genera las “paredes” empinadas visibles hacia la izquierda y la derecha de la figura. Esta simetría respecto al eje  $y$  es consistente con el hecho de que  $x^4$  no depende de  $y$ . En contraste, la variación en la dirección  $y$  es comparativamente más suave: la curvatura en este eje proviene únicamente del término cuadrático  $4(x + y + xy)^2$ , donde  $y$  aparece acoplado con  $x$ .

En la región central se aprecia un valle que corresponde a la vecindad del mínimo global  $(0, 0)$ . En este entorno la superficie no es un paraboloide perfecto, sino que presenta una zona algo “aplanada” en ciertas direcciones, coherente con la existencia de un autovalor nulo en la Hessiana calculada en el origen. Esto se traduce visualmente en cambios de altura más graduales cerca de  $(0, 0)$  en comparación con el crecimiento explosivo cuando  $|x|$  es grande.

Otro rasgo relevante es la presencia de una especie de “meseta” casi horizontal alineada con  $x = -1$ . Analíticamente se sabe que

$$f(-1, y) = 4(-1 + y - y)^2 + (-1)^4 = 5 \quad \text{para todo } y \in \mathbb{R},$$

de modo que, en la sección  $x = -1$ , la función es constante. En la superficie esta característica se manifiesta como una banda de altura prácticamente uniforme, lo que ilustra de manera gráfica la falta de coercividad: alejándose en la dirección  $y$  con  $x = -1$  la norma de  $(x, y)$  crece, pero el valor de  $f$  permanece fijo.

En conjunto, la gráfica 3D en escala lineal sintetiza bien el comportamiento global del modelo: un mínimo global único en el origen, una meseta a lo largo de  $x = -1$  y un crecimiento muy pronunciado en la variable  $x$ , mientras que la dependencia en  $y$  es más suave y está controlada por el acoplamiento  $x + y + xy$ .

## Algoritmos

En ambos algoritmos, se utiliza la búsqueda de paso tipo Armijo, dada una posición  $x = (x, y)$ , un valor  $f(x)$  y un *vector de descenso*  $d$  (esto es,  $\nabla f(x)^\top d < 0$ ), la regla de Armijo con retroceso busca un tamaño de paso  $\alpha > 0$  tal que

$$f(x + \alpha d) \leq f(x) + c \alpha \nabla f(x)^\top d, \quad (0.2)$$

donde  $c \in (0, 1)$  es un parámetro fijo (típicamente  $c = 10^{-4}$ ). Se parte de  $\alpha_0 > 0$  y se actualiza  $\alpha \leftarrow \rho \alpha$  con un factor  $\rho \in (0, 1)$  (por ejemplo  $\rho = 0,5$ ) hasta que se cumple (0.2). En el notebook asociado a este documento, estos parámetros aparecen como `alpha0`, `rho` y `c`, con un máximo de retrocesos `max_bt` para evitar bucles infinitos.

Si  $d$  es de descenso y  $f$  tiene gradiente localmente, la condición de Armijo garantiza disminución suficiente de  $f$  y evita pasos que no bajen el objetivo. En combinación con una regla razonable para escoger  $d$ , esto produce secuencias  $\{x_k\}$  con  $f(x_{k+1}) < f(x_k)$  y, bajo hipótesis estándar, cualquier punto límite es estacionario.

El método de descenso por gradiente elige  $d_k = -\nabla f(x_k)$  y calcula  $\alpha_k$  mediante Armijo; luego  $x_{k+1} = x_k + \alpha_k d_k$ . Se detiene cuando  $\|\nabla f(x_k)\| < \text{tol}$  o al alcanzar un `maxit`. En notación del código: `gd_with_linesearch` devuelve, además de la solución, el historial de  $f$  y de tamaños de paso  $\alpha_k$ .

En una vecindad del punto óptimo  $(0, 0)$  la superficie presenta un cuenco globalmente suave pero con anisotropía marcada en la curvatura: la Hessiana en el origen posee un autovalor nulo que induce una dirección localmente plana de orden superior, lo que tiende a generar longitudes de paso reducidas en los métodos de primer orden. En regiones donde la combinación  $z = x + y + xy$  permanece aproximadamente constante, la dinámica está esencialmente gobernada por el término  $4x^4$ , de modo que la curvatura efectiva se concentra en la variable  $x$  mientras la variación en  $y$  resulta comparativamente menos pronunciada. Este desbalance entre direcciones fuertemente curvas y direcciones casi degeneradas ralentiza el avance del algoritmo, que progresa de forma cautelosa a medida que se aproxima al mínimo global. La regla de Armijo, no obstante, garantiza una disminución controlada del valor objetivo: cualquier tamaño de paso que produzca un incremento en  $f$  es automáticamente rechazado, lo que proporciona estabilidad numérica y asegura que la secuencia generada sea estrictamente descendente.

El esquema de Newton amortiguado usa, en cada iteración, la dirección

$$d_k = -(\nabla^2 f(x_k) + \lambda_k I)^{-1} \nabla f(x_k),$$

donde  $\lambda_k \geq \lambda_0 > 0$  es un parámetro de amortiguamiento (regularización de Levenberg–Marquardt) que estabiliza el paso cuando la Hessiana es singular o indefinida. El tamaño  $\alpha_k$  se acepta con Armijo; si el retroceso falla (`alpha = 0` en el código), se incrementa  $\lambda_k$

(p.ej.,  $\lambda_k \leftarrow 10\lambda_k$ ). Cuando el paso es aceptado, se suele relajar la regularización (*e.g.*  $\lambda_{k+1} = \max\{\lambda_0, \lambda_k/2\}$ ).

En cuanto a la convergencia, cerca de un minimizador no degenerado (Hessiana definida positiva), Newton con amortiguación recupera la convergencia superlineal/cuadrática típica al hacerse pequeña la regularización. Lejos del óptimo, la búsqueda de paso y el término  $\lambda_k I$  proporcionan robustez global al impedir saltos inestables hacia zonas de Hessiana problemática (como el punto silla).

## Experimentos numéricos y análisis

Cuadro 1: Estadísticos resumen de los 400 experimentos por método, con puntos iniciales muestreados uniformemente en  $[-100, 100]^2$ .

Métrica	Descenso por gradiente	Newton amortiguado
Número de casos	400	400
Tasa de éxito (%)	0.25	66.5
Iteraciones medianas $k$	3000	58
Rango intercuartílico de $k$ (P25, P75)	(3000, 3000)	(49, 1000)
Valor objetivo mediano $f$	0.552544	0.0
Norma del gradiente mediana $\ \nabla f\ $	$1.11 \times 10^{-3}$	0.0
Longitud de camino mediana	155.25	101.73
Máx. valor absoluto de las coordenadas (mediana)	69.61	85.80
Casos cercanos al origen (%)	44.5	58.25
Casos divergentes (%)	12.5	32.75
Fallos por <code>no_step/maxit</code> (%)	99.75	33.5

La Tabla 1 recoge un resumen estadístico global de los 400 experimentos realizados para cada algoritmo, con puntos iniciales tomados aleatoriamente en el cuadrado  $[-100, 100]^2$ . La tasa de éxito indica el porcentaje de casos en los que el método alcanzó la tolerancia prescrita sobre la norma del gradiente antes de agotar el número máximo de iteraciones. Las estadísticas sobre  $k$  describen el esfuerzo iterativo típico: el descenso por gradiente presenta una mediana de 3000 iteraciones y un rango intercuartílico degenerado en (3000, 3000), lo que evidencia que, salvo excepciones, el algoritmo se detiene por la condición de máximo de iteraciones. En contraste, Newton amortiguado alcanza la tolerancia en aproximadamente dos tercios de los casos, con una mediana de sólo 58 iteraciones y un primer cuartil en 49, aunque el tercer cuartil se sitúa en 1000, reflejando la presencia de un subconjunto no despreciable de trayectorias difíciles.

Los valores medianos de la función objetivo y de la norma del gradiente refuerzan este contraste. Mientras que el descenso por gradiente se queda en un valor mediano de  $f \approx 0.55$  y una norma de gradiente del orden de  $10^{-3}$ , el método de Newton amortiguado alcanza en la mediana valores prácticamente nulos tanto para  $f$  como para  $\|\nabla f\|$ , coherentes con la aproximación efectiva al mínimo global. La longitud de camino mediana indica que, pese a usar muchas más iteraciones, el descenso por gradiente recorre una trayectoria algo más larga que Newton, lo que analizaremos después que se debe a pasos más cortos y zigzagueantes en

valles estrechos. El máximo valor absoluto mediano de las coordenadas muestra que ambos métodos exploran regiones de magnitud comparable, si bien Newton puede aventurarse algo más lejos en algunos casos antes de regresar hacia el óptimo.

Las categorías finales permiten interpretar la calidad de las soluciones obtenidas. En descenso por gradiente, aproximadamente un 44.5% de las trayectorias termina en una vecindad del origen que puede considerarse aceptable desde el punto de vista práctico, pero un 12.5% se clasifica como divergente y prácticamente todos los casos fallan por la lentitud del método. Para Newton amortiguado, cerca del 58% de las trayectorias se sitúa en la categoría `origin_near`, coherente con la alta tasa de éxito, mientras que un 32.75% aparece como divergente, asociado mayoritariamente a la meseta no coerciva en torno a  $x \approx -1$  y al punto silla ya analizado.

## Casos representativos

La Tabla 2 resume algunos casos particularmente ilustrativos extraídos del conjunto de experimentos. En todos ellos se muestra el punto inicial  $(x_0, y_0)$  y un resumen de la iteración final de cada algoritmo.<sup>1</sup>

De estos casos se extraen varios patrones coherentes con el análisis teórico del modelo:

- **Descenso por gradiente (GD).** En los casos 0000, 0002, 0010 se puede observar que el punto final está muy cerca del mínimo global  $(0, 0)$  con valores de  $f$  del orden de  $10^{-7}$ , por lo que, desde un punto de vista práctico, el método se comporta como un algoritmo robusto pero lento.
- **Newton amortiguado.** Cuando converge al origen (casos 0000, 0003, 0018), lo hace en un número de iteraciones entre 50 y 70, con valores de  $f$  del orden de  $10^{-12}$  y norma del gradiente menor que  $10^{-8}$ , es decir, con precisión muy superior a la que alcanza GD en el mismo presupuesto de iteraciones. No obstante, el método presenta dos modos de fallo:
  - (a) *Deslizamiento sobre la cuasi-meseta  $x \approx -1$*  (casos 0002, 0010): la iteración se mantiene en torno a  $x = -1$  mientras  $|y|$  crece hasta valores cercanos a  $10^2$ , con  $f$  prácticamente constante. Esta dinámica es consistente con la estructura de la función, que sobre la recta  $x = -1$  presenta variación muy limitada en  $y$ .
  - (b) *Convergencia al punto silla  $(-1, -\frac{3}{2})$*  (caso 0007): pese a no ser un minimizador, la combinación de Hessiana indefinida y término de regularización puede producir un punto fijo de la iteración de Newton en esa región.

Con esto queda claro que los ejemplos donde Newton amortiguado falla están íntimamente ligados a las propiedades globales de la función.

Tras este análisis, las conclusiones cualitativas son que, en este modelo concreto, GD con Armijo actúa como un método conservador: requiere muchas iteraciones pero raramente produce trayectorias patológicas. Newton amortiguado, en cambio, ofrece una convergencia muy rápida al mínimo global cuando la trayectoria entra en su cuenca de atracción, pero puede desviarse hacia la meseta  $x \approx -1$  o hacia el punto silla si la información de curvatura local se interpreta en una región donde la Hessiana es indefinida o casi singular.

---

<sup>1</sup>Los valores numéricos se redondean a tres cifras significativas.

Cuadro 2: Casos representativos de la dinámica de descenso por gradiente (GD) y Newton amortiguado (Nwt).

Caso	$(x_0, y_0)$	GD + Armijo	Newton amortiguado + Armijo
0000	$(-54,53, -36,65)$	$k = 3000$ (se alcanza <b>maxit</b> ); $f_k \approx 7,6 \times 10^{-7}$ , $\ \nabla f_k\  \approx 8,1 \times 10^{-4}$ ; punto final $\approx (-0,029, 0,030)$ , clasificado como <b>origin_near</b> .	Convergencia exitosa en $k = 58$ ; $f_k \approx 1,5 \times 10^{-12}$ , $\ \nabla f_k\  \approx 9,2 \times 10^{-9}$ ; punto final $\approx (-1,1 \times 10^{-3}, 1,1 \times 10^{-3})$ , muy cercano al mínimo global.
0002	$(-21,78, -33,44)$	$k = 3000$ ; $f_k \approx 8,5 \times 10^{-7}$ , $\ \nabla f_k\  \approx 2,7 \times 10^{-4}$ ; punto final cercano al origen, categoría <b>origin_near</b> .	El método de Newton no converge: tras $k = 1000$ iteraciones se obtiene $f_k \approx 1,01$ , $\ \nabla f_k\  \approx 4,6 \times 10^{-4}$ ; el punto final $\approx (-1,00, -4,9 \times 10^2)$ corresponde a un deslizamiento a lo largo de un valle casi plano próximo a $x = -1$ ( <b>diverged</b> ).
0003	$(19,66, -62,65)$	$k = 3000$ ; $f_k \approx 1,11$ , $\ \nabla f_k\  \approx 7,1 \times 10^{-3}$ ; el algoritmo termina en $\approx (-1,03, -39,19)$ , alejado del óptimo, y se clasifica como <b>other</b> .	Convergencia exitosa en $k = 65$ ; $f_k \approx 1,5 \times 10^{-12}$ , $\ \nabla f_k\  \approx 8,8 \times 10^{-9}$ ; punto final $\approx (1,1 \times 10^{-3}, -1,1 \times 10^{-3})$ ( <b>origin_near</b> ).
0007	$(-11,63, 77,30)$	$k = 3000$ ; $f_k \approx 0,948$ , $\ \nabla f_k\  \approx 2,3 \times 10^{-3}$ ; el punto final $\approx (-0,987, 74,46)$ mantiene el valor objetivo casi constante a lo largo de una trayectoria esencialmente vertical ( <b>diverged</b> ).	Convergencia rápida en $k = 13$ con $f_k = 5$ y $\ \nabla f_k\  \approx 8,0 \times 10^{-12}$ ; el punto final es $\approx (-1, -1,5)$ , correspondiente al punto silla teórico.
0010	$(-83,68, -68,02)$	$k = 3000$ ; $f_k \approx 5,1 \times 10^{-7}$ , $\ \nabla f_k\  \approx 7,2 \times 10^{-4}$ ; nuevamente finaliza cerca del origen ( <b>origin_near</b> ).	Newton amortiguado diverge: tras $k = 1000$ iteraciones se obtiene $f_k \approx 1,01$ , $\ \nabla f_k\  \approx 4,6 \times 10^{-4}$ ; el punto final $\approx (-1,00, -4,9 \times 10^2)$ se sitúa sobre la cuasi-meseta con $x \approx -1$ ( <b>diverged</b> ).
0018	$(9,24, 87,53)$	$k = 3000$ ; $f_k \approx 0,952$ , $\ \nabla f_k\  \approx 1,3 \times 10^{-3}$ ; el punto final $\approx (-0,988, 80,86)$ se mantiene lejos del óptimo en una trayectoria predominantemente vertical ( <b>diverged</b> ).	Convergencia en $k = 51$ ; $f_k \approx 1,5 \times 10^{-12}$ , $\ \nabla f_k\  \approx 9,5 \times 10^{-9}$ ; punto final $\approx (-1,1 \times 10^{-3}, 1,1 \times 10^{-3})$ ( <b>origin_near</b> ).

## Visualización de gráficas y observaciones

La Figura 2 muestra las líneas de nivel de  $f(x, y)$  en una ventana restringida alrededor del mínimo global. La escala de colores representa  $\log f(x, y)$ : las zonas más oscuras corresponden a valores pequeños de la función y, en particular, el valle pronunciado que conduce al origen  $(0, 0)$  se visualiza como una banda estrecha y oscura. Las curvas de nivel aparecen fuertemente alargadas, lo que refleja la marcada anisotropía de la curvatura: hay una dirección de fuerte pendiente transversal al valle y una dirección casi plana a lo largo de éste, en concordancia con la presencia de un autovalor pequeño (e incluso nulo en el límite) en la Hessiana.

Sobre este paisaje se superponen las trayectorias de los algoritmos. La ruta azul corresponde

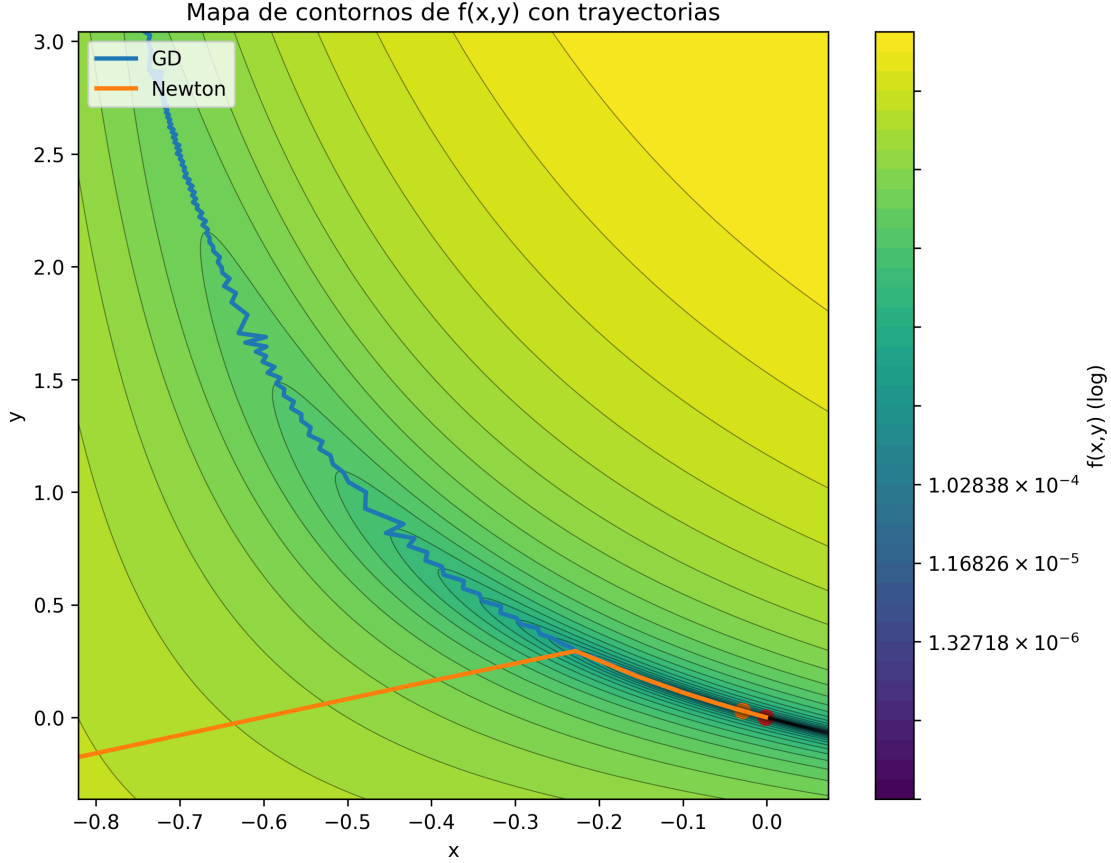


Figura 2: Mapa de contornos de  $f(x,y)$  en escala logarítmica con las trayectorias de descenso por gradiente (GD, azul) y Newton amortiguado (naranja) a partir de un mismo punto inicial.

a descenso por gradiente con búsqueda de Armijo. Se observa un movimiento claramente “serrucho”: el método desciende siguiendo la dirección de gradiente más pronunciada, pero como el valle es estrecho, cada paso tiende a sobrepasar el eje central y la siguiente iteración corrige en la dirección opuesta. Este efecto de zigzag es típico en valles alargados y se traduce en pasos relativamente cortos, especialmente cuando la búsqueda de Armijo reduce el tamaño de paso para respetar la condición de descenso suficiente.

La trayectoria naranja muestra el comportamiento de Newton amortiguado desde el mismo punto inicial. En este caso, los primeros saltos son oblicuos y mucho más largos: la información de la Hessiana permite “precondicionar” el gradiente y aproximar mejor la dirección del valle, de modo que la trayectoria se alinea rápidamente con la dirección de curvatura pequeña y se aproxima al origen en pocas iteraciones. Cerca del mínimo global ambos métodos se concentran en la misma región, pero Newton llega en un número de pasos muy reducido, mientras que GD necesita muchas iteraciones pequeñas para avanzar a lo largo de la dirección casi plana.

En conjunto, la figura ilustra de forma gráfica la diferencia fundamental entre ambos métodos: el descenso por gradiente es robusto pero sensible a la mala condición de las curvas de nivel (zigzag en valles estrechos), mientras que Newton amortiguado aprovecha la estructura de segundo orden para corregir la dirección de búsqueda y lograr una convergencia mucho



más rápida cuando la Hessiana es informativa y el régimen local es favorable.

## Superficie 3D con trayectoria: descenso por gradiente

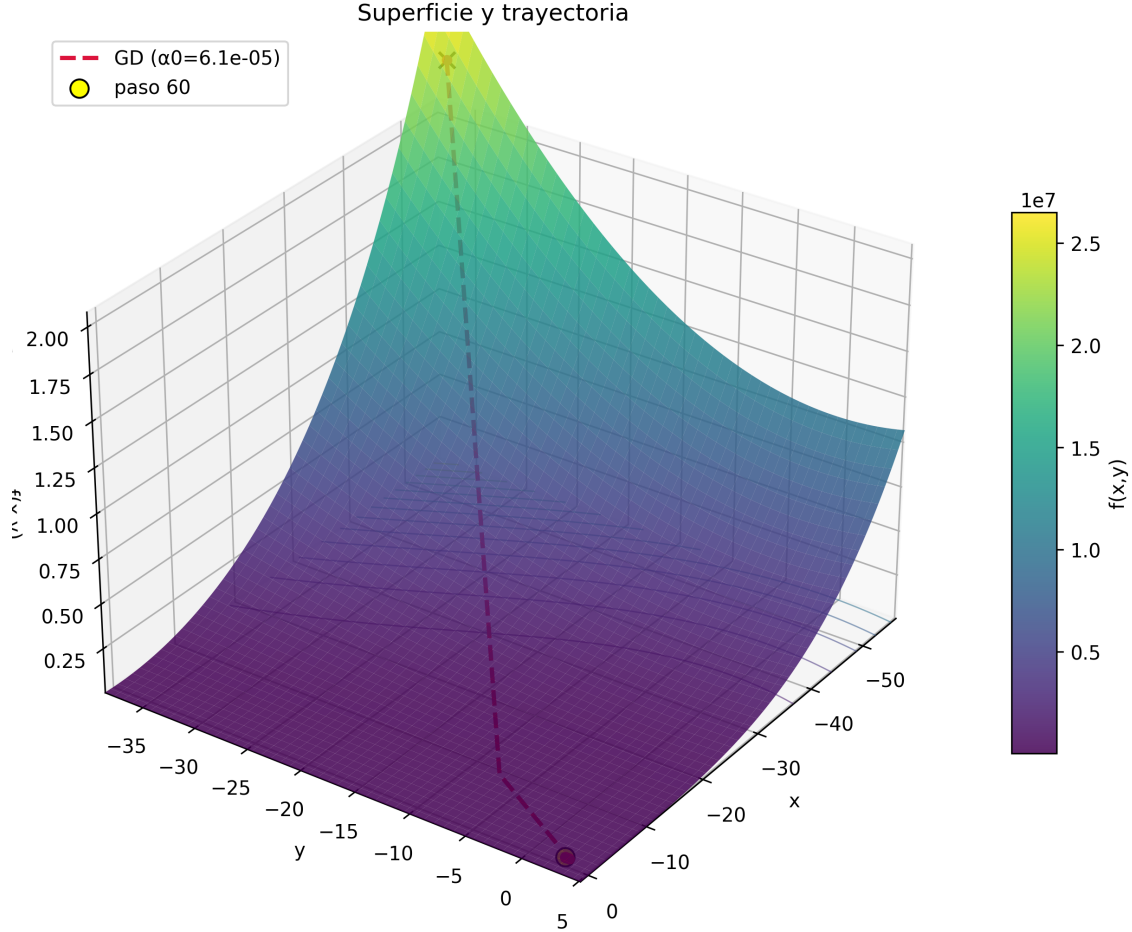


Figura 3: Superficie  $z = f(x, y)$  y trayectoria de descenso por gradiente con búsqueda de Armijo, partiendo de un punto inicial lejano al origen. La cruz marca el punto inicial y el círculo amarillo indica la posición en el paso  $k = 60$ .

Observando la figura 3 se puede notar como en los primeros pasos la pendiente es extremadamente grande y la curvatura está dominada por el término  $x^4$ . En esta zona la búsqueda de paso ajusta automáticamente el tamaño  $\alpha_k$ : los pasos son suficientemente largos como para reducir rápidamente el valor de la función, pero no tan agresivos como para provocar un *overshoot* inestable, lo que aparece en la figura como un tramo prácticamente recto que baja desde la cima hacia la zona de menor altura.

El marcador correspondiente al paso  $k = 60$  (círculo amarillo) señala el momento en que la trayectoria entra en una región de topografía mucho más suave, próxima al valle que conduce al origen. A partir de ahí, el relieve está influido por la combinación  $x + y + xy$  y por la dirección casi plana detectada en el análisis de la Hessiana, donde los tamaños de paso que impone Armijo tienden a reducirse.

## Superficie 3D con trayectoria: Newton amortiguado

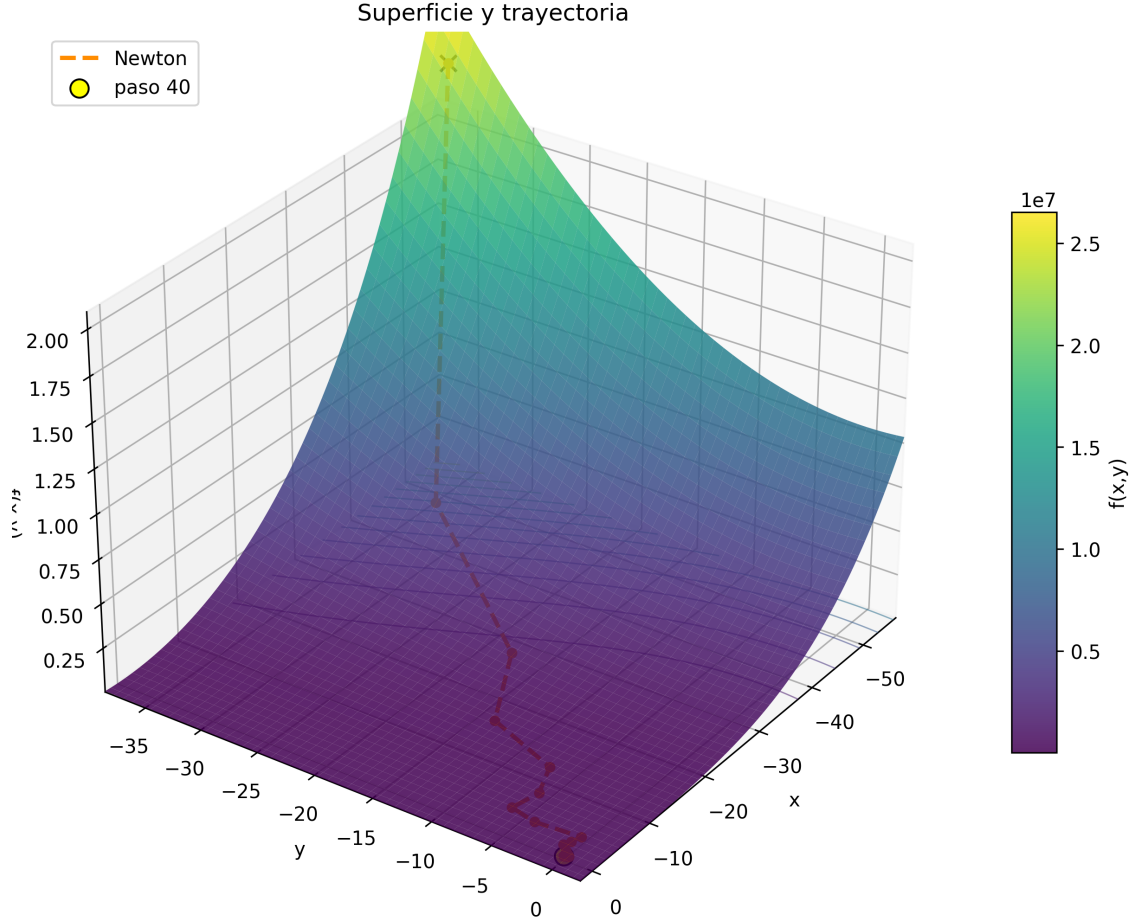


Figura 4: Superficie  $z = f(x, y)$  y trayectoria del método de Newton amortiguado con búsqueda de Armijo. La cruz marca el punto inicial y el círculo amarillo indica la posición en el paso  $k = 40$ .

En cuanto a el análisis de la figura 4 se observa como desde el punto inicial la curva discontinua naranja desciende en muy pocas iteraciones hacia la región donde se localiza el mínimo global, cerca del origen.

En las primeras iteraciones el término de regularización  $\lambda_k$  asegura que la matriz  $\nabla^2 f(x_k, y_k) + \lambda_k I$  sea definida positiva, incluso en una zona donde la Hessiana puede ser mal condicionada. Esto permite construir direcciones de búsqueda bien orientadas, que aprovechan la información de curvatura para realizar pasos oblicuos largos que “cortan” el valle en lugar de seguir el gradiente más pronunciado.

El marcador correspondiente al paso  $k = 40$  indica que, a esa altura, el algoritmo ya se encuentra en una vecindad muy cercana al mínimo, con  $f$  reducido varios órdenes de magnitud respecto al valor inicial. En esta fase, la combinación de la regla de Armijo y la disminución progresiva de  $\lambda_k$  hace que los pasos se acorten y se asemejen a los del método de Newton puro.

Comparada con la trayectoria de descenso por gradiente, la figura pone de manifiesto la principal ventaja de Newton amortiguado en este problema: desde un mismo punto inicial, consigue recorrer la ladera y entrar en la cuenca de atracción del mínimo global en un número de iteraciones significativamente menor, gracias al uso explícito de la Hessiana y al papel estabilizador de la regularización.

## Conclusiones

A partir del análisis teórico realizado se concluye que la función objetivo

$$f(x, y) = 4(x + y + xy)^2 + x^4$$

puede expresarse como suma de cuadrados reales, lo que garantiza su no negatividad en todo  $\mathbb{R}^2$  y permite identificar de forma inmediata a  $(0, 0)$  como único punto donde  $f$  se anula. El estudio detallado del gradiente y de la Hessiana muestra que sólo existen dos puntos críticos: el origen, que resulta ser un mínimo local estricto y, por la descomposición anterior, también mínimo global único, y el punto  $(-1, -\frac{3}{2})$ , clasificado como punto silla. La presencia de una recta sobre la cual  $f(-1, y) = 5$  para todo  $y$  explica tanto la falta de coercividad como la existencia de una meseta casi horizontal en la región  $x \approx -1$ , donde el valor de la función apenas varía aunque la norma de  $(x, y)$  crezca.

Los experimentos numéricos confirman la influencia de esta geometría en el comportamiento de los algoritmos. El descenso por gradiente con búsqueda de línea tipo Armijo produce trayectorias de descenso monótono y se muestra robusto frente a puntos iniciales lejanos: las órbitas permanecen acotadas y, en la mayoría de los casos, el método se aproxima a una vecindad del mínimo global. Sin embargo, la mala condición del valle alrededor del origen, con direcciones casi planas asociadas a autovalores muy pequeños de la Hessiana, implica que el número de iteraciones requerido para alcanzar la tolerancia impuesta sobre la norma del gradiente es elevado; en muchos casos el algoritmo se detiene por alcanzar el máximo de iteraciones aunque el valor de  $f$  ya sea muy pequeño desde el punto de vista práctico.

El método de Newton amortiguado con Armijo, que incorpora información de segundo orden y un término de regularización  $\lambda_k I$ , exhibe un comportamiento complementario. Cuando la trayectoria entra en la cuenca de atracción del mínimo y la Hessiana es suficientemente informativa, la convergencia es muy rápida: se alcanzan valores de  $f$  del orden de  $10^{-12}$  y normas de gradiente por debajo de  $10^{-8}$  en pocas decenas de iteraciones. No obstante, los experimentos también muestran modos de fallo: en varios casos la iteración deriva hacia la meseta con  $x \approx -1$ , recorriendo grandes distancias en la variable  $y$  sin reducir de manera significativa el objetivo, o incluso converge al punto silla  $(-1, -\frac{3}{2})$ . Estas situaciones reflejan la sensibilidad del método frente a Hessianas indefinidas o mal condicionadas, incluso en presencia de regularización.

En conjunto, el estudio pone de manifiesto el compromiso clásico entre robustez y rapidez en métodos de optimización sin restricciones. El descenso por gradiente con Armijo actúa como un método conservador y estable, adecuado como referencia básica cuando se prioriza la seguridad frente a trayectorias patológicas, mientras que Newton amortiguado resulta muy atractivo cuando se dispone de buenos puntos iniciales o se trabaja en regiones donde la curvatura es bien comportada. Como líneas de trabajo futuro se plantea el diseño de reglas

adaptativas para el parámetro de regularización, así como la incorporación de estrategias de reinicio que permitan escapar de la meseta en torno a  $x \approx -1$ , además de la comparación con métodos cuasi-Newton que podrían ofrecer un equilibrio adicional entre el costo computacional, la robustez y la velocidad de convergencia.