

# Tarea evaluativa de Optimización

C312

Rachel Mojena González

10 de noviembre de 2025

$$\text{Superficie } f(x,y) = \cos(x) \sin(y) - x/(y^2+1)$$

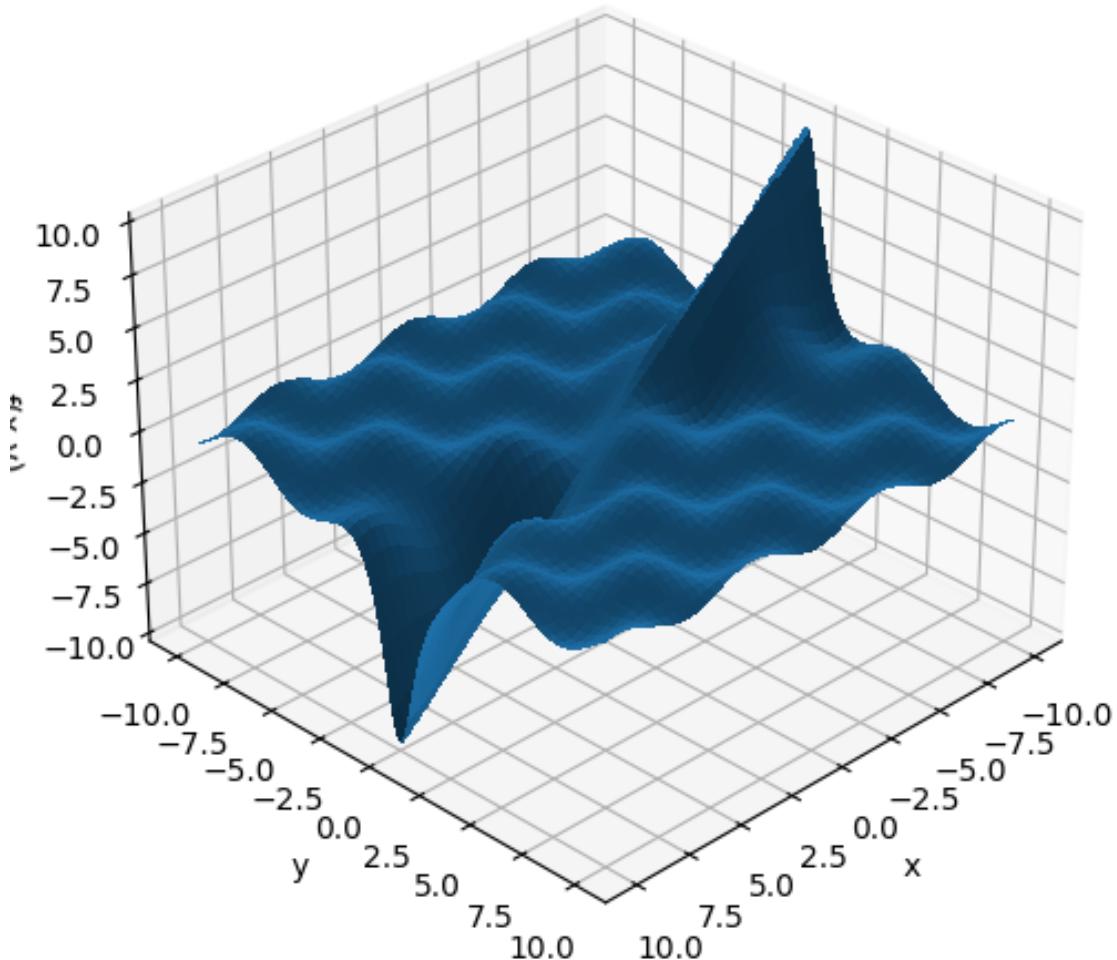


Figura 1: Superficie/curva de la función usada en el experimento.

## 1. Formulación del modelo

Se considera la función objetivo

$$f(x, y) = \cos x \sin y - \frac{x}{y^2 + 1}, \quad (x, y) \in \mathbb{R}^2, \quad (1)$$

la cual es  $C^\infty$  en todo el plano. Dado que  $f(x, 0) = -x \rightarrow -\infty$  cuando  $x \rightarrow +\infty$ , el problema irresticto no está acotado inferiormente y carece de mínimo global. Para efectos de análisis y verificación con KKT, se trabaja la versión acotada

$$\mathcal{B} = [-100, 100]^2 = \{(x, y) : -100 \leq x \leq 100, -100 \leq y \leq 100\}.$$

En forma estándar, las restricciones se escriben como

$$\begin{aligned} g_1(x, y) &= x - 100 \leq 0, & g_2(x, y) &= -x - 100 \leq 0, \\ g_3(x, y) &= y - 100 \leq 0, & g_4(x, y) &= -y - 100 \leq 0. \end{aligned} \quad (2)$$

## 2. Diagnóstico teórico

### 2.1. Gradiente y hessiana

Las derivadas parciales de primer orden son

$$\begin{aligned} f_x(x, y) &= -\sin x \sin y - \frac{1}{y^2 + 1}, \\ f_y(x, y) &= \cos x \cos y + \frac{2xy}{(y^2 + 1)^2}. \end{aligned} \quad (3)$$

La matriz hessiana resulta

$$\nabla^2 f(x, y) = \begin{bmatrix} -\cos x \sin y & -\sin x \cos y + \frac{2y}{(y^2 + 1)^2} \\ -\sin x \cos y + \frac{2y}{(y^2 + 1)^2} & -\cos x \sin y + \frac{2x(1 - 3y^2)}{(y^2 + 1)^3} \end{bmatrix}. \quad (4)$$

### 2.2. No convexidad

Para que  $f$  fuera convexa en  $\mathbb{R}^2$  sería necesario que su hessiana fuese semidefinida positiva (SDP) para todo  $(x, y)$ .

Evalúese en el punto  $(x_0, y_0) = (0, \frac{\pi}{2})$ , donde  $\sin y_0 = 1$  y  $\cos y_0 = 0$ :

$$\nabla^2 f(0, \frac{\pi}{2}) = \begin{bmatrix} -1 & \frac{2(\frac{\pi}{2})}{((\frac{\pi}{2})^2 + 1)^2} \\ \frac{2(\frac{\pi}{2})}{((\frac{\pi}{2})^2 + 1)^2} & -1 \end{bmatrix} = \begin{bmatrix} -1 & \alpha \\ \alpha & -1 \end{bmatrix}, \quad \alpha = \frac{\pi}{((\frac{\pi}{2})^2 + 1)^2} \approx 0,261.$$

Para matrices simétricas  $2 \times 2$ , la condición SDP es equivalente a  $a \geq 0$  y  $\det \geq 0$  para  $H = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ .

Aquí  $a = -1 < 0$ , luego  $H$  no es SDP. De hecho,  $\det(H) = (-1)^2 - \alpha^2 > 0$  y ambos autovalores son negativos  $(-1 \pm \alpha)$ , con lo cual la hessiana es definida negativa en ese punto. Por consiguiente,  $f$  no es convexa.

### 2.3. Puntos estacionarios del problema irrestricto

Los puntos estacionarios interiores satisfacen  $\nabla f = 0$ , esto es

$$\boxed{\sin x \sin y = -\frac{1}{y^2 + 1}}, \quad \boxed{\cos x \cos y = -\frac{2xy}{(y^2 + 1)^2}}. \quad (5)$$

Debido a la periodicidad de las funciones trigonométricas, existen infinitos candidatos en  $\mathbb{R}^2$  y, en particular, dentro de  $\mathcal{B}$ .

## 3. Condiciones KKT en $\mathcal{B}$

Sea

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \sum_{i=1}^4 \lambda_i g_i(x, y), \quad \lambda_i \geq 0.$$

Las KKT necesarias bajo LICQ, que se satisface pues los gradientes activos son afines e independientes.

El análisis se realiza por *conjuntos activos*  $I(x, y) = \{i : g_i(x, y) = 0\}$ .

**Interior**  $I = \emptyset$

Del problema irrestricto se obtiene  $\lambda_i = 0$  y  $f_x = 0$ ,  $f_y = 0$ . Se listan numéricamente algunos candidatos interiores dentro de  $\mathcal{B}$ , por ejemplo  $(x, y) \approx (0,7946, -0,8935)$  con  $f \approx -0,9878$  y  $(x, y) \approx (4,4881, -3,2314)$  con  $f \approx -0,4122$ .

**Borde derecho**  $I = \{1\}$  ( $x = 100$ )

Las ecuaciones KKT quedan

$$\cos(100) \cos y + \frac{200y}{(y^2 + 1)^2} = 0, \quad \lambda_1 = -f_x(100, y) = \sin(100) \sin y + \frac{1}{y^2 + 1} \ (\geq 0).$$

Entre las soluciones en  $[-100, 100]$ , el valor

$$y \approx -0,0043117$$

satisface  $\lambda_1 \approx 1,0022 \geq 0$  y produce

$$f(100, y) \approx -100,001859.$$

**Borde izquierdo**  $I = \{2\}$  ( $x = -100$ )

Se obtiene

$$\cos(100) \cos y - \frac{200y}{(y^2 + 1)^2} = 0, \quad \lambda_2 = f_x(-100, y) = \sin(100) \sin y - \frac{1}{y^2 + 1} \ (\geq 0).$$

Los candidatos resultantes presentan valores significativamente superiores (en valor absoluto menores) al del caso  $x = 100$ ; por ejemplo  $f(-100, -39,2737) \approx -0,7975$ .

**Borde superior**  $I = \{3\}$  ( $y = 100$ )

Las KKT imponen

$$-\sin x \sin(100) - \frac{1}{100^2 + 1} = 0, \quad \lambda_3 = -f_y(x, 100) \ (\geq 0).$$

Entre las soluciones factibles en  $[-100, 100]$ , el mejor valor observado es  $f(x, 100) \approx 0,4966$  en  $x \approx 97,3892$ .

**Borde inferior**  $I = \{4\}$  ( $y = -100$ )

Análogamente,

$$-\sin x \sin(-100) - \frac{1}{100^2 + 1} = 0, \quad \lambda_4 = f_y(x, -100) \ (\geq 0),$$

con mejores valores alrededor de 0,497, lejos del caso  $x = 100$ .

**Esquinas**  $I = \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}$

En las cuatro combinaciones, algún multiplicador resulta negativo, por lo que no satisfacen KKT; además, sus valores de  $f$  no son competitivos.

**Conclusión en  $\mathcal{B}$ .** Por comparación de los candidatos KKT dentro de  $\mathcal{B}$  se obtiene

$$(x, y) = (100, y), \quad y \approx -0,0043117, \quad \lambda_1 \approx 1,0022, \quad \lambda_{2,3,4} = 0,$$

con valor  $f(x, y) \approx -100,001859$ . Dado que el dominio es compacto y  $f$  es continua, este punto corresponde al mínimo global del problema acotado.

## Algoritmos

Métodos implementados: Máximo descenso y Newton.

## Descenso por Gradiente y Newton

**Máximo descenso (gradient descent).** Itera desde  $x_k$  en la dirección de mayor disminución instantánea,  $d_k = -\nabla f(x_k)$ , elige un tamaño de paso  $\alpha_k$  y actualiza

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

Se considera robusto porque funciona en muchas funciones no convexas, rara vez “explota”, y su atracción local es amplia. Su convergencia típica es lineal y puede zigzaguear en valles alargados.

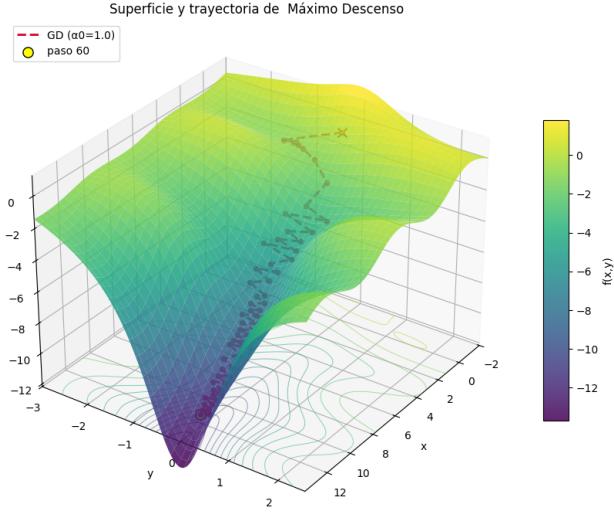


Figura 2: Superficie de Máximo descenso desde 0,0

**Newton.** Construye una aproximación cuadrática local usando  $\nabla f(x_k)$  y la Hessiana  $\nabla^2 f(x_k)$ , y toma  $d_k$  resolviendo

$$\nabla^2 f(x_k) d_k = -\nabla f(x_k), \quad x_{k+1} = x_k + \alpha_k d_k.$$

Si  $\nabla^2 f(x^*) \succ 0$  y estamos suficientemente cerca de  $x^*$ , el método amortiguación adecuada) tiene convergencia *superlineal/cuadrática*.

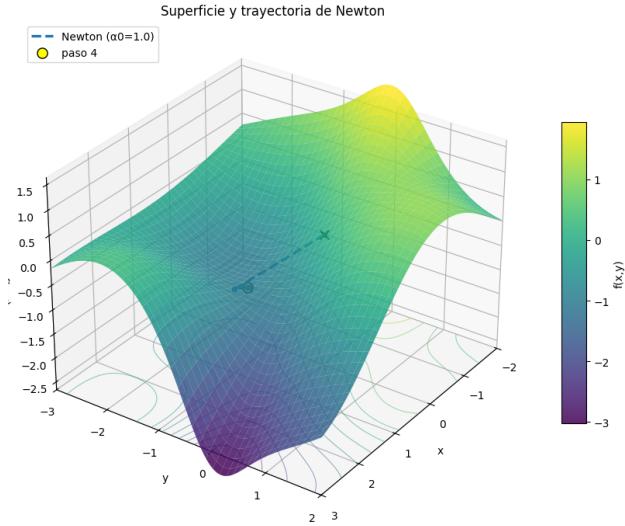


Figura 3: Superficie de Newton desde 0,0

### ¿Qué son las reglas de Armijo/Wolfe?

- **Armijo (descenso suficiente):** dados  $x$  y una dirección de descenso  $d$  con  $\nabla f(x)^\top d < 0$ , se busca  $\alpha > 0$  tal que

$$f(x + \alpha d) \leq f(x) + c_1 \alpha \nabla f(x)^\top d, \quad c_1 \in (0, 1).$$

En backtracking se empieza con  $\alpha = \alpha_0$  y se reduce  $\alpha \leftarrow \rho\alpha$  ( $\rho \in (0, 1)$ ) hasta cumplir la desigualdad.

- **Wolfe:** además de Armijo, impone una condición de curvatura

$$|\nabla f(x + \alpha d)^\top d| \leq c_2 |\nabla f(x)^\top d|, \quad 0 < c_1 < c_2 < 1,$$

Garantiza que no se eligen pasos “demasiado cortos” y controla la curvatura.

### Problema irrestricto vs. restringido a caja

En el **irrestricto**, nuestra  $f$  *no* tiene mínimo global: en  $y = 0$ ,  $f(x, 0) = -x \rightarrow -\infty$ . Con Armijo sigue encontrando pasos que bajan  $f$  y no aterriza.

En la **caja**  $[-100, 100]^2$  (problema restringido) el mínimo *existe* (Weierstrass). Sin embargo, como  $f$  es no convexa, descenso/Newton sólo garantizan llegar a puntos *KKT* (locales) y no al global. De hecho, en esta función el mínimo global está en el borde ( $x = 100$ ,  $y \approx 0$ ), pero hay muchos estacionarios interiores con  $f \approx -1$ .

Trabajamos con una función suave y derivable,  $f(x, y) = \cos x \sin y - \frac{x}{y^2+1}$ , acotada en la caja  $[-100, 100]^2$ . Este modelo “calza” perfecto con *Descenso por Gradiente* y *Newton* porque podemos evaluar  $f$ ,  $\nabla f$  y  $\nabla^2 f$  de forma directa, y la restricción de caja se maneja con una idea simple: después de proponer el paso  $x + \alpha d$ , se recorta (proyecta) a los límites. La búsqueda en línea tipo Armijo ajusta automáticamente  $\alpha$  para asegurar descenso.

Además, ambos métodos se complementan: *Descenso por Gradiente* es estable y barato por iteración, ideal para avanzar desde puntos lejanos; *Newton* aprovecha la información de curvatura y acelera mucho al acercarnos al mínimo.

### Análisis de los experimentos con casos de prueba

Se evaluaron puntos de la forma  $(10n, 10m)$ , incluyendo

$$\left\{ \begin{array}{l} (-50, -50), (-50, 0), (-50, 50), \\ (0, -50), (0, 0), (0, 50), \\ (50, -50), (50, 0), (50, 50), \\ (90, -10), (90, 0), (90, 10), \\ (100, -20), (100, -10), (100, 0), (100, 10), (100, 20) \end{array} \right.$$

y dos tamaños de paso inicial  $\alpha_0 \in \{0,5, 1,0\}$ . Para cada combinación se ejecutaron *Descenso por Gradiente* y *Newton*, registrando el valor final  $f$ , el número de pasos y el  $\alpha_0$  utilizado.

**Resultados y conclusiones.** En términos globales, **Newton** obtuvo el mejor  $f$  en la *mayoría* de las combinaciones, especialmente cuando el arranque está cerca de la pared  $x \approx 100$ ; por ejemplo, desde  $(90, 0)$  se alcanzó el mínimo global de caja  $f \approx -100,001859$  con 42 y 24 pasos para  $\alpha_0 = 0,5$  y  $1,0$  respectivamente, mientras que gradiente necesitó 844 pasos en ambos casos. Desde  $(100, 0)$ , ambos métodos alcanzaron el mínimo global ( $f \approx -100,001859$ ); aquí, gradiente lo hizo en  $\sim 25$  pasos y Newton en  $\sim 38$ , mostrando que cerca del borde “bueno” el gradiente también puede ser

muy eficiente. Desde  $(50, 0)$ , Newton con  $\alpha_0 = 0,5$  llegó al global en 33 pasos (gradiente se quedó en 1000 pasos), pero con  $\alpha_0 = 1,0$  Newton llegó muy cerca ( $f \approx -99,923$ ) en sólo 5 pasos, ilustrando que un  $\alpha_0$  mayor reduce pasos, aunque no siempre mejora  $f$  final. En arranques lejanos al borde (p. ej.,  $(\pm 50, \pm 50)$  o  $(0, \pm 50)$ ) ambos métodos convergen a estacionarios *locales* con  $f \approx -1$ , y en  $(0, 0)$  se observó que gradiente (con muchas iteraciones) alcanzó  $f \approx -51$ , mientras Newton cayó rápidamente en un estacionario interior ( $f \approx -0,988$ ). En resumen: cerca de  $x = 100$ , Newton suele llegar antes al óptimo global y con menos pasos que gradiente; en interiores, Newton puede estabilizarse en estacionarios locales con muy pocos pasos mientras gradiente continúa descendiendo más tiempo; el efecto de  $\alpha_0$  es consistente:  $\alpha_0 = 1,0$  tiende a acortar trayectorias, pero  $\alpha_0 = 0,5$  puede dar el mejor  $f$  en algunos arranques (p. ej.,  $(-50, 0)$ ).

Repositorio