

Tarea evaluativa de Optimización

C312

Rachel Mojena González

12 de noviembre de 2025

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

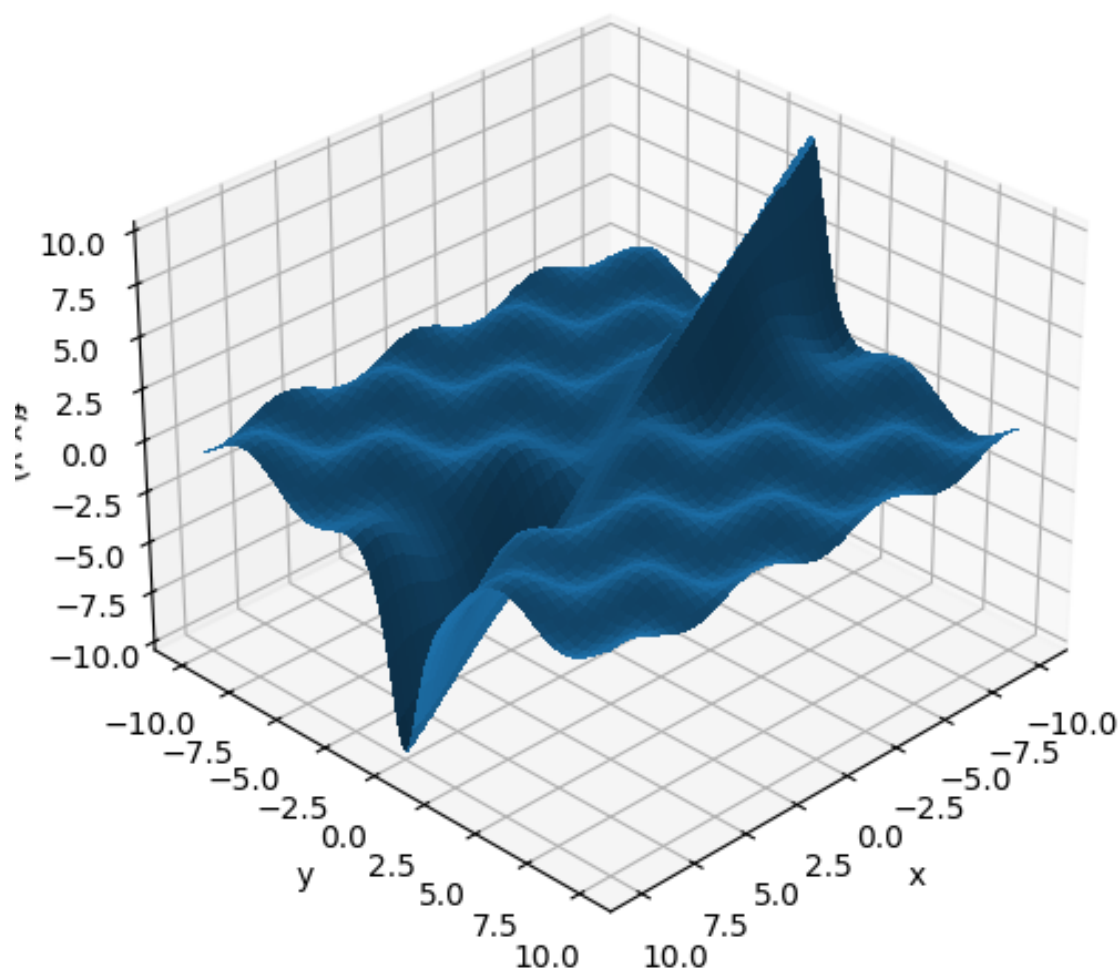


Figura 1: Superficie/curva de la función usada en el experimento.

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^∞ en todo el plano. Dado que $f(x, 0) = -x \rightarrow -\infty$ cuando $x \rightarrow +\infty$, el problema irrestricto 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 2×2 , 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

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

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

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

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.

4. Algoritmos

Métodos implementados: Máximo descenso 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 α_k y actualiza

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

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

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*.

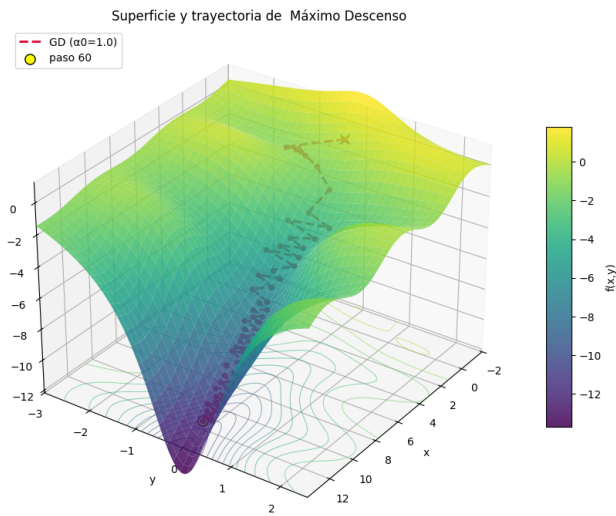


Figura 2: Superficie de Máximo descenso desde 0,0

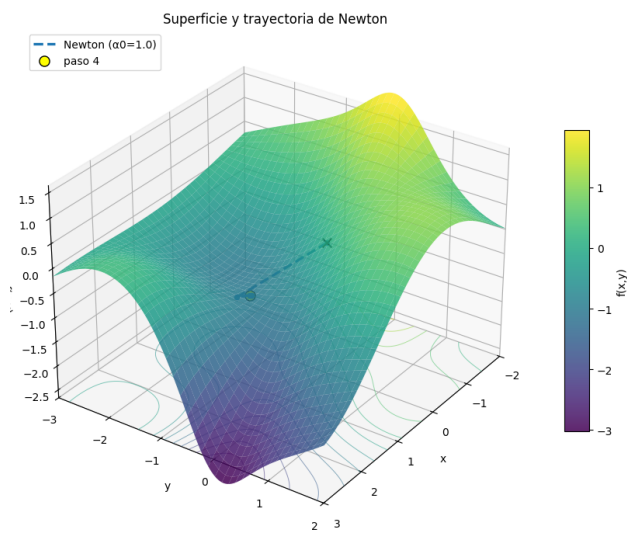


Figura 3: Superficie de Newton desde 0,0

¿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 , ∇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 α 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.

5. 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 α_0 utilizado.

Inicio	Método	α_0	Pasos	f
(-50,-50)	GD	0,5	13	-0,981300
	GD	1,0	4	-0,981300
	Newton	0,5	13	-1,004811
	Newton	1,0	7	-1,004811
(-50,0)	GD	0,5	19	-0,842857
	GD	1,0	5	-0,929623
	Newton	0,5	5	-99,997552
	Newton	1,0	34	-100,001859
(-50,50)	GD	0,5	12	-0,978811
	GD	1,0	3	-0,978811
	Newton	0,5	12	-0,978811
	Newton	1,0	5	-0,978811
(0,-50)	GD	0,5	17	-1,000000
	GD	1,0	4	-1,000000
	Newton	0,5	17	-1,000000
	Newton	1,0	4	-1,000000
(0,0)	GD	0,5	1000	-50,906769
	GD	1,0	1000	-50,994332
	Newton	0,5	16	-0,987785
	Newton	1,0	4	-0,987785
(0,50)	GD	0,5	14	-1,000000
	GD	1,0	3	-1,000000
	Newton	0,5	14	-1,000000
	Newton	1,0	4	-1,000000
(50,-50)	GD	0,5	13	-1,018700
	GD	1,0	4	-1,018700
	Newton	0,5	11	-1,018700
	Newton	1,0	4	-1,018700
(50,0)	GD	0,5	1000	-72,925831
	GD	1,0	1000	-72,925831
	Newton	0,5	33	-100,001859
	Newton	1,0	5	-99,923259
(50,50)	GD	0,5	12	-1,021190
	GD	1,0	3	-1,021190
	Newton	0,5	10	-1,021190
	Newton	1,0	3	-1,021190

Cuadro 1: Resultados por punto (I): primero GD y debajo Newton.

Inicio	Método	α_0	Pasos	f
(90,-10)	GD	0,5	15	-1,756858
	GD	1,0	5	-1,756858
	Newton	0,5	14	-1,756858
	Newton	1,0	3	-1,756858
(90,0)	GD	0,5	844	-100,001859
	GD	1,0	844	-100,001859
	Newton	0,5	42	-100,001859
	Newton	1,0	24	-100,001859
(90,10)	GD	0,5	27	-2,532801
	GD	1,0	12	-2,532801
	Newton	0,5	10	-0,994424
	Newton	1,0	5	-0,994424
(100,-20)	GD	0,5	9	-1,101877
	GD	1,0	6	-1,101877
	Newton	0,5	6	-1,101877
	Newton	1,0	9	-1,101877
(100,-10)	GD	0,5	30	-2,576982
	GD	1,0	11	-2,576982
	Newton	0,5	10	-2,572774
	Newton	1,0	8	-2,576982
(100,0)	GD	0,5	25	-100,001859
	GD	1,0	25	-100,001859
	Newton	0,5	38	-100,001859
	Newton	1,0	38	-100,001859
(100,10)	GD	0,5	12	-1,696013
	GD	1,0	6	-1,696013
	Newton	0,5	3	-1,696012
	Newton	1,0	3	-1,696013
(100,20)	GD	0,5	20	-1,233257
	GD	1,0	8	-1,233257
	Newton	0,5	12	-0,277657
	Newton	1,0	5	-0,277657

Cuadro 2: Resultados por punto (II): primero GD y debajo Newton.

Análisis

- Cercanos a la restricción $x \approx 100$ (p.ej., (90,0)). Newton alcanza el mínimo global de caja ($f \approx -100,001859$) con menos iteraciones que descenso por gradiente: 42 y 24 pasos para $\alpha_0 = 0,5$ y 1,0 respectivamente, mientras descenso por gradiente requiere ~ 844 pasos en ambos casos. Al no estar aún activa la restricción, los pasos proyectados de descenso por gradiente sufren más retrocesos; en cambio, el paso de Newton, guiado por la curvatura, aproxima más rápido la dirección tangente adecuada y activa la restricción en menos iteraciones.
- Exactamente sobre la restricción (100,0). Ambos métodos alcanzan el mínimo global ($f \approx -100,001859$). Aquí, descenso por gradiente lo hace en ~ 25 pasos y Newton en ~ 38 .

- Semicentral $(50, 0)$. Con $\alpha_0 = 0,5$, Newton llega al global en 33 pasos (descenso por gradiente se topa con el tope de 1000 iteraciones). Con $\alpha_0 = 1,0$, Newton cae muy cerca ($f \approx -99,923$) en solo 5 pasos. Un α_0 mayor acelera la llegada a una región que contiene al óptimo, pero no garantiza la mejor evaluación de la función objetivo final.
- Lejanos al borde: $(\pm 50, \pm 50)$ y $(0, \pm 50)$. Ambos métodos convergen a estacionarios locales con $f \approx -1$. La parte trigonométrica induce una topografía no convexa con múltiples pozos; lejos de la restricción que contiene al óptimo, las cuencas interiores dominan y la proyección rara vez activa el plano $x = 100$ donde está el global.
- Centro $(0, 0)$. Descenso por gradiente (con muchas iteraciones) alcanza $f \approx -51$, mientras Newton cae rápido en un estacionario interior con $f \approx -0,988$. En $(0, 0)$ el gradiente apunta hacia $+x$ y $-y$, lo que favorece que descenso por gradiente se dirija hacia valores de x grandes (reduciendo f por el término $-x/(1 + y^2)$). Newton puede estabilizarse en un pozo interior cercano si la proyección no activa la restricción $x = 100$ pronto.

5.1. Gráficas comparativas

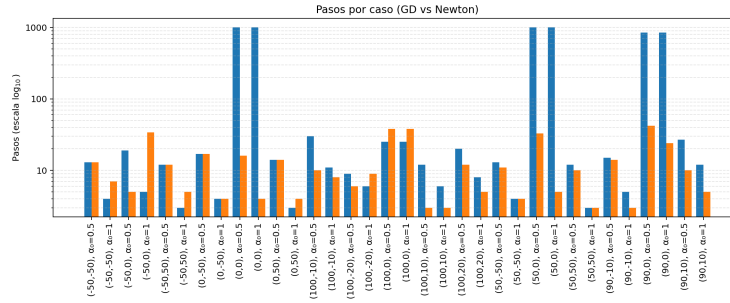


Figura 4: Barras agrupadas del número de pasos por caso, donde cada categoría muestra dos barras: descenso por gradiente y Newton. Barras más bajas significan menos pasos y, por tanto, trayectorias más cortas. Los picos en torno a 10^3 delatan casos donde se agotó el máximo de iteraciones, mientras que diferencias amplias entre barras indican ventaja clara de un método.

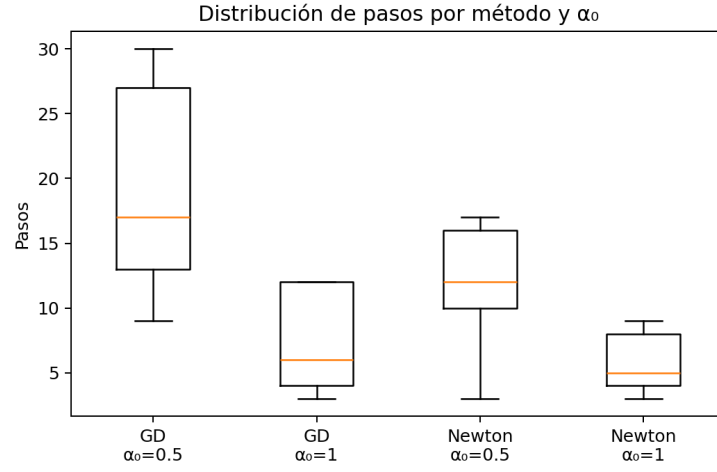


Figura 5: Diagrama de cajas de los pasos por método y α_0 . La línea central es la mediana.

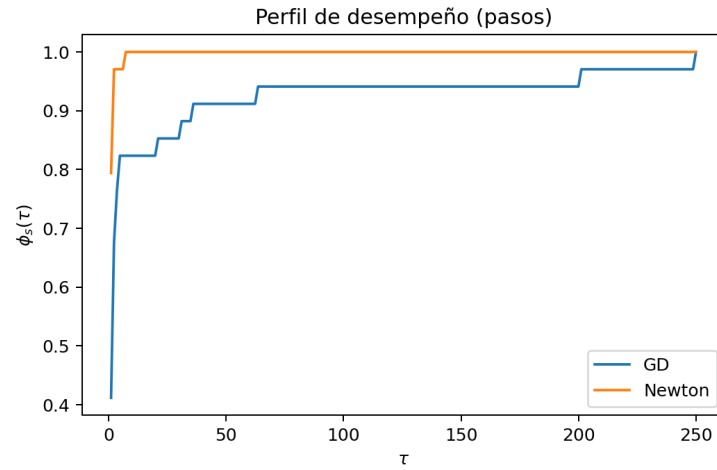


Figura 6: Para cada τ en el eje x , $\phi_s(\tau)$ es la fracción de casos donde el método s usa a lo sumo τ veces los pasos del mejor método en ese caso. Curvas más altas indican mejor desempeño global.

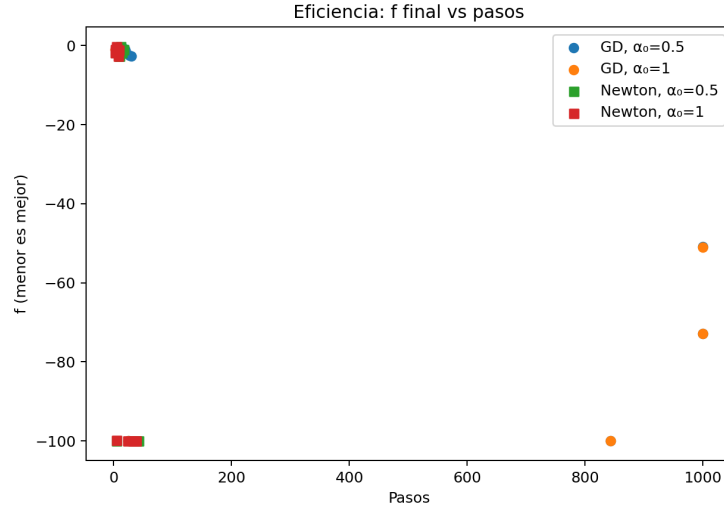


Figura 7: Dispersión del valor final de la evaluación de la función objetivo frente al número de pasos. Cada marcador identifica método y α_0 .

Conclusiones.

- Dominio cerca del óptimo de caja. En vecindad de $x = 100$, Newton suele lograr mejores evaluaciones de la función objetivo con menos pasos que descenso por gradiente $(90, 0)$, gracias a el modelo de curvatura que activa antes la restricción. Una vez sobre la cara activa (caso $(100, 0)$), descenso por gradiente puede igualar o superar en pasos por moverse de forma más directa en la dirección tangente permitida.
- α_0 mayor tiende a acortar trayectorias, pero no garantiza la mejor evaluación de la función objetivo final: con proyección, un paso grande puede dejar al método cerca de la restricción que contiene al óptimo sin llegar lo suficiente (véase $(50, 0)$ con $\alpha_0 = 1,0$ frente a $0,5$).
- Desde arranques interiores o lejanos al borde, ambos métodos pueden caer en estacionarios locales ($f \approx -1$). Descenso por gradiente, al seguir el gradiente hacia $+x$, puede alcanzar valores mucho más bajos (como $f \approx -51$ desde $(0, 0)$).
- Newton domina en proximidad al borde, y el ajuste de α_0 presenta un claro compromiso entre menos pasos y una mejor evaluación de la función objetivo final.

Repositorio