Tarea Evaluativa de Optimización

Carlos Mazorra Matos c-311

11 de noviembre de 2025

1. Función, dominio y primeras observaciones

- **Función:**
$$f:\mathbb{R}^2\to\mathbb{R},\qquad f(x,y)=x^2+y^2-10\,\cos^2(x)-10\,\cos\left(\frac{x^2}{30}\right).$$
- **Dominio:** $\operatorname{dom} f=\mathbb{R}^2$.
- **Continuidad:** La función $f(x,y)$ es una suma algebraica de los siguientes términos:
  1. **x^2 y y^2:** Son **funciones polinómicas**, y por lo tanto, son **continuas** en todo el dominio $\mathbb{R}^2$.
  2. **$-10(\cos(x))^2$:**
     - $\cos(x)$ es **continua** en $\mathbb{R}$.
     - La función cuadrática $t\mapsto t^2$ es **continua** en $\mathbb{R}$.
     - Por **composición** de funciones continuas, $(\cos(x))^2$ es continua.
     - La multiplicación por la constante -10 mantiene la continuidad.
  3. **$-10\cos\left(\frac{x^2}{30}\right)$:**
     - El argumento del coseno, $p(x)=\frac{x^2}{30}$, es un **polinomio**, por lo tanto, es **continuo** en $\mathbb{R}$.
     - La función $\cos(t)$ es **continua** en $\mathbb{R}$.
     - Por **composición** de funciones continuas, $\cos\left(\frac{x^2}{30}\right)$ es continua.
     - La multiplicación por la constante -10 mantiene la continuidad.

  Por tanto la función es continua en todo el dominio.
- **Cota inferior:** usando $\cos^2(x)\le 1$ y $\cos(t)\le 1$ para todo t, se tiene
$$f(x,y)\ge x^2+y^2-10-10=x^2+y^2-20\ge -20.$$
  En particular, f es *coerciva* (tiende a $+\infty$ cuando $\|(x,y)\|\to\infty$), por lo que admite al menos un mínimo global.

2. Gradiente

Las derivadas parciales son
$$\begin{aligned}\frac{\partial f}{\partial x}(x,y)&=2x-10\,\frac{\mathrm{d}}{\mathrm{d}x}\big(\cos^2 x\big)-10\,\frac{\mathrm{d}}{\mathrm{d}x}\Big(\cos\big(\tfrac{x^2}{30}\big)\Big)\\&=2x+10\,\sin(2x)+\frac{2x}{3}\,\sin\left(\frac{x^2}{30}\right),\\\frac{\partial f}{\partial y}(x,y)&=2y.\end{aligned}$$

En forma vectorial,

$$\nabla f(x,y) = \begin{bmatrix} 2x + 10\,\sin(2x) + \frac{2x}{3}\,\sin(\frac{x^2}{30}) \\ 2y \end{bmatrix}.$$

3. Hessiano y convexidad

La segunda derivada respecto de x es

$$\frac{\partial^2 f}{\partial x^2}(x,y) = 2 + 20\,\cos(2x) + \frac{2}{3}\,\sin\left(\frac{x^2}{30}\right) + \frac{2x^2}{45}\,\cos\left(\frac{x^2}{30}\right),$$

$$\frac{\partial^2 f}{\partial x\,\partial y}(x,y) = 0, \qquad \frac{\partial^2 f}{\partial y^2}(x,y) = 2.$$

Así, el Hessiano es diagonal:

$$\nabla^2 f(x,y) = \begin{bmatrix} 2 + 20\,\cos(2x) + \frac{2}{3}\,\sin(\frac{x^2}{30}) + \frac{2x^2}{45}\,\cos(\frac{x^2}{30}) & 0 \\ 0 & 2 \end{bmatrix}.$$

No convexidad global. El término $20\cos(2x)$ oscila en $[-20, 20]$; existen valores de x (p.ej. $x \approx \frac{\pi}{2}$) donde la entrada $(1,1)$ del Hessiano se hace claramente menor que $2 - 20 < 0$ (los términos restantes pueden ser pequeños), por lo que $\nabla^2 f$ no es semidefinida positiva en todo $\mathbb{R}^2$. Luego, f *no* es convexa globalmente.

4. Puntos estacionarios

Los puntos estacionarios satisfacen $\nabla f(x,y) = 0$.

- De $\frac{\partial f}{\partial y} = 2y = 0$ se obtiene necesariamente $y^* = 0$.
- Para x, debemos resolver la ecuación

$$g(x) = 2x + 10\,\sin(2x) + \frac{2x}{3}\,\sin\left(\frac{x^2}{30}\right) = 0.$$

Es inmediato que $x = 0$ es solución (pues $\sin(0) = 0$).

Por lo tanto, $(0,0)$ es un punto estacionario.

Evaluación de la Hessiana en el punto

Evaluando la Hessiana en el punto $(0,0)$ obtenemos:

$$\nabla^2 f(0,0) = \begin{bmatrix} 2 + 20\,\cos(0) + \frac{2}{3}\,\sin(0) + \frac{2(0)^2}{45}\,\cos(0) & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 + 20(1) + 0 + 0 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 22 & 0 \\ 0 & 2 \end{bmatrix}.$$

Los valores propios de la matriz Hessiana $\nabla^2 f(0,0)$ son los elementos de la diagonal, $\lambda_1 = 22$ y $\lambda_2 = 2$. Como ambos valores propios son **positivos** ($\lambda_1 > 0$ y $\lambda_2 > 0$), el Hessiano es **definido positivo** en $(0,0)$, y por tanto, el punto $(0,0)$ es un **mínimo local estricto**.

Optimalidad global y unicidad

Usando las cotas del inicio,

$$f(x,y) \geq x^2 + y^2 - 20 \geq -20.$$

En $(0,0)$ se obtiene

$$f(0,0) = 0 + 0 - 10\cos^2(0) - 10\cos(0) = -10 - 10 = -20,$$

con lo cual $(0,0)$ *alcanza* la cota inferior. Como los máximos valores que pueden tomar $\cos^2(x)$ y $\cos(\frac{x^2}{30})$ es 1 y $x^2 + y^2 \geq 0$ entonces $(0,0)$ es el **único mínimo global** y

$$\min_{(x,y)\in\mathbb{R}^2} f(x,y) = f(0,0) = -20.$$

5. Descripción General del método de Máximo descenso

El **Método de Máximo Descenso**, también conocido como **Gradiente Descendente** (*Gradient Descent*), es un algoritmo iterativo de optimización utilizado para encontrar mínimos locales de funciones diferenciables de varias variables $f : \mathbb{R}^n \to \mathbb{R}$.

La premisa fundamental del método es que el **gradiente** de una función en un punto dado ($\nabla f(\mathbf{x}_k)$) apunta en la dirección de *máximo ascenso*. Por lo tanto, moverse en la dirección del **gradiente negativo** ($-\nabla f(\mathbf{x}_k)$) garantiza el *descenso* más rápido posible en el valor de la función, al menos localmente.

Partiendo de un punto inicial $\mathbf{x}_0$, se genera una secuencia de puntos $\mathbf{x}_1, \mathbf{x}_2, \ldots$ que se espera converja a un mínimo local de f.

5.1. Fórmula del Método

En cada iteración k, el siguiente punto $\mathbf{x}_{k+1}$ se calcula mediante la siguiente fórmula de actualización:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k)$$

Donde:

- $\mathbf{x}_k \in \mathbb{R}^n$: Es el punto actual en la iteración k.
- $\nabla f(\mathbf{x}_k)$: Es el **gradiente** de f evaluado en $\mathbf{x}_k$. Indica la dirección del máximo ascenso.
- $-\nabla f(\mathbf{x}_k)$: Es la **dirección de descenso**.
- $\alpha_k > 0$: Es el **tamaño del paso**..

5.2. Procedimiento Básico

El algoritmo se desarrolla en los siguientes pasos iterativos:

1. **Inicialización:** Escoger un punto inicial $\mathbf{x}_0$ y establecer un criterio de parada (tolerancia ϵ).
2. **Cálculo del Gradiente:** En la iteración k, calcular el gradiente $\nabla f(\mathbf{x}_k)$.
3. **Criterio de Parada:** Si $\nabla f(\mathbf{x}_k)$ es menor que ϵ, terminar el proceso.
4. **Determinación del Paso:** Elegir el tamaño del paso $\alpha_k > 0$ (fijo, o mediante un método como la búsqueda de paso óptimo).
5. **Actualización:** Calcular el nuevo punto $\mathbf{x}_{k+1}$ usando la fórmula de actualización.
6. **Repetir:** Incrementar $k \leftarrow k + 1$ y volver al paso 2.

5.3. Ventajas y Desventajas

Ventajas

- **Simplicidad:** Es conceptualmente sencillo y fácil de implementar.
- **Eficiencia Computacional:** Solo requiere el cálculo de las derivadas parciales (el gradiente), no necesita la costosa matriz Hessiana ($\nabla^2 f$).
- **Escalabilidad:** Funciona eficientemente en problemas con un número muy grande de variables (alta dimensionalidad).

Desventajas

- **Lentitud:** La convergencia puede ser muy lenta cuando la función tiene contornos alargados o presenta un *condicionamiento pobre* (es decir, cuando los valores propios de la Hessiana son muy diferentes). Esto resulta en un zigzagueo.
- **Dependencia de α:** Es altamente sensible a la elección del tamaño del paso (α). Un α muy grande puede causar **divergencia** (sobrepasar el mínimo) y un α muy pequeño conduce a una convergencia excesivamente **lenta**.
- **Mínimos Locales:** El método está garantizado para converger solo a un **mínimo local** y no necesariamente al mínimo global.

6. Descripción General del método de Newton

El **Método de Newton** es un potente algoritmo iterativo de optimización utilizado para encontrar puntos estacionarios (mínimos, máximos o puntos de silla) de una función dos veces diferenciable $f : \mathbb{R}^n \to \mathbb{R}$. A diferencia del Gradiente Descendente, el método utiliza información de la **curvatura** de la función a través de la matriz Hessiana, lo que le confiere una velocidad de convergencia superior cerca del óptimo.

La función $f(\mathbf{x})$ se aproxima localmente en la iteración k mediante una expansión de Taylor de segundo orden alrededor de $\mathbf{x}_k$:

$$f(\mathbf{x}) \approx f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T(\mathbf{x} - \mathbf{x}_k) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_k)^T \mathbf{H}_f(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k)$$

El siguiente punto $\mathbf{x}_{k+1}$ se determina encontrando el punto crítico (gradiente cero) de esta aproximación cuadrática.

6.1. Fórmula del Método

En cada iteración k, el siguiente punto $\mathbf{x}_{k+1}$ se calcula como:

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

Donde:

- $\nabla f(\mathbf{x}_k)$: Es el **vector gradiente** de f en $\mathbf{x}_k$.
- $\mathbf{H}_f(\mathbf{x}_k)$: Es la **matriz Hessiana** de f en $\mathbf{x}_k$, definida por la matriz de segundas derivadas parciales:

$$\mathbf{H}_f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

El término $\mathbf{p}_k = \mathbf{H}_f(\mathbf{x}_k)^{-1} \nabla f(\mathbf{x}_k)$ es la **dirección de Newton**, que ajusta la dirección y el tamaño del paso basándose en la curvatura local.

6.2. Procedimiento Básico

La forma estándar de implementar la iteración es resolviendo un sistema de ecuaciones lineales en lugar de calcular explícitamente la inversa de la Hessiana.

1. **Inicialización:** Escoger un punto inicial $\mathbf{x}_0$ y un criterio de parada ϵ.
2. **Cálculo:** En la iteración k, calcular el gradiente $\nabla f(\mathbf{x}_k)$ y la Hessiana $\mathbf{H}_f(\mathbf{x}_k)$.
3. **Criterio de Parada:** Si $\nabla f(\mathbf{x}_k) < \epsilon$, detener el algoritmo.
4. **Resolución del Sistema:** Resolver el sistema de ecuaciones lineales para obtener la dirección $\mathbf{p}_k$:
$$\mathbf{H}_f(\mathbf{x}_k)\mathbf{p}_k = \nabla f(\mathbf{x}_k)$$
5. **Actualización:** Calcular el nuevo punto:
$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{p}_k$$
6. **Repetir:** Incrementar $k \leftarrow k + 1$ y volver al paso 2.

6.3. Ventajas y Desventajas

Ventajas

- **Convergencia Rápida:** Posee una **convergencia cuadrática** cerca de un mínimo local estricto, lo que significa que el número de cifras significativas correctas en el resultado se duplica en cada paso. Típicamente converge en muy pocas iteraciones.
- **Paso Automático:** La dirección de Newton incluye automáticamente el tamaño del paso que se requiere para alcanzar el mínimo de la aproximación cuadrática.

Desventajas

- **Alto Costo Computacional:** Requiere calcular, almacenar y, a menudo, invertir (o resolver un sistema lineal con) la matriz Hessiana. Para funciones con n variables, esto implica n^2 cálculos de segundas derivadas por iteración.
- **Problemas con la Hessiana:** Si la matriz Hessiana $\mathbf{H}_f(\mathbf{x}_k)$ no es definida positiva (es singular o indefinida), el método puede fallar, divergir, o conducir a un máximo o punto de silla.
- **Sensibilidad:** Es sensible a la elección del punto inicial, especialmente en regiones no convexas de la función, donde la convergencia puede ser errática.

7. Análisis de Métodos para la Optimización de $f(x, y)$

7.1. Método de Máximo Descenso (Gradiente)

Razones para su Uso

Es conceptualmente sencillo y requiere cálculos relativamente básicos, solo requiere el cálculo del **gradiente** $\nabla f(x, y)$, y funciona bien para explorar regiones amplias del dominio y es robusto para encontrar mínimos locales. Sin embargo la convergencia puede ser muy lenta si los términos oscilatorios del gradiente hacen que la superficie cambie bruscamente y necesita un ajuste cuidadoso del tamaño de paso.

7.2. Método de Newton

Razones para su Uso

Utiliza la **matriz Hessiana** $\mathbf{H}_f(x, y)$, que incorpora la curvatura de la función, puede converger muy rápido cerca de un mínimo local y la dirección y magnitud del paso se ajustan automáticamente, aprovechando la información de la curvatura. Pero también hay que tener en cuenta que debido a los términos oscilatorios $\cos(2x)$, la entrada $(1, 1)$ de la Hessiana puede tomar valores negativos. Esto puede hacer que $\mathbf{H}_f$ sea indefinida, lo que podría desviar el método o hacerlo converger a un máximo o un punto silla.

8. Análisis de Resultados de Algoritmos de Optimización

8.1. Algoritmo: Newton (Método de Newton)

Este algoritmo no utiliza un *learning rate* (lr) y su comportamiento depende fuertemente del punto de inicio (x_0).

Experimentos Convergentes (al óptimo de -20)

- Todos los experimentos que comenzaron con una coordenada $x = 0$ (es decir, x_0 en el eje Y, como [0, 0], [0, 10], [0, 30], etc.) convergieron exitosamente.
- **Velocidad:** La convergencia fue extremadamente rápida. En todos estos casos, el algoritmo encontró el mínimo en 1 solo paso (la trayectoria tiene 2 puntos: el inicial y el final).

Experimentos No Convergentes (a -20)

- Todos los experimentos que comenzaron con una coordenada $x \neq 0$ (p.ej., [10, 0], [30, 10], [-10, 90], [-70, -50]) fallaron en encontrar el mínimo de -20.0.
- **Comportamiento:** En lugar de converger al mínimo global, se dirigieron a otros puntos estacionarios (posiblemente mínimos locales) y se quedaron atascados allí.
- Inicios en [10, 0], [30, 0], [50, 0], [70, 0], [90, 0] y [30, 10], [-30, 50] convergieron a un valor de 1689.45...
- Inicios en [50, 50] y [50, -50] convergieron a un valor de 885.98...
- Inicios en [-10, 0], [-10, 90] convergieron a un valor de 66.48...
- Inicios en [-30, 0], [-50, 0], [-30, -30] y [-50, -10] convergieron a un valor de -10.699...
- Inicios en [-70, 0], [-90, 0], [-70, -50] y [-90, -70] convergieron a un valor de 89.06...

8.2. Algoritmo: Máximo Descenso (Gradiente)

El comportamiento de este algoritmo dependió críticamente de los parámetros: *learning rate* (lr) y número de iteraciones.

8.2.1. Agrupación por Parámetros

Con lr = 0.1

- **Convergencia:** Éxito. Este *learning rate* fue el más efectivo. Logró converger a -20.0 (o un valor muy cercano como -19.999...) desde todos los puntos de inicio probados.
- **Velocidad:** Fue el más rápido para este algoritmo, encontrando el mínimo en menos de 100 pasos (p.ej., $\sim$75-85 pasos para los inicios en el eje Y).

Con lr = 0.01

- **Convergencia:** Éxito, pero lento. Este lr también logró converger al valor de -20.0 desde los diferentes puntos de inicio, pero necesitó el máximo de 1000 iteraciones para hacerlo.
- **Fallo (por iteraciones):** Cuando se limitó a 50 o 200 iteraciones, este lr no fue suficiente y el algoritmo se detuvo prematuramente (p.ej., con x_0=[0, 90], 200 iteraciones solo llegaron a un valor de -17.4).

Con lr = 0.001

- **Convergencia:** Fallo. Este *learning rate* es demasiado pequeño. El algoritmo avanzó muy lentamente y no logró converger en ninguno de los experimentos, ni siquiera con 1000 iteraciones.
- **Ejemplo:** Con x_0=[0, 90], tras 1000 iteraciones, el valor apenas llegó a 127.7, muy lejos del objetivo de -20.0.

9. Comparativa de Algoritmos y conclusiones

- El **Método de Newton** es rápido pero inestable. Cuando funciona (en este caso, cuando x_0 está en el eje Y), es increíblemente eficiente (1 paso). Sin embargo, es muy sensible al punto de inicio y converge a mínimos locales incorrectos si x_0 no está en la cuenca de atracción adecuada.
- El **Máximo Descenso** es lento pero seguro. Demostró ser mucho más robusto, encontrando el mínimo correcto (-20) desde todos los puntos de inicio probados. Su desventaja es que es mucho más lento que Newton y su éxito depende totalmente de la correcta elección de sus parámetros.

9.1. Importancia de los Parámetros (Máximo Descenso)

- La elección del *tamaño del paso* (lr) es fundamental. Un lr de 0.001 es inútilmente pequeño, mientras que 0.1 es muy efectivo.
- El número de iteraciones también es importante. Un lr más pequeño (como 0.01) requiere más iteraciones (más de 200) para alcanzar la convergencia que un lr más grande (como 0.1).

9.2. Valles

La función que se está optimizando parece tener un mínimo global claro en [0, 0] (valor -20), pero también posee múltiples mínimos locales o puntos estacionarios en el eje X que atrapan al algoritmo de Newton si se inicia cerca de ellos.

El entorno figure se utiliza para que LaTeX gestione la colocación de la imagen.

10. Experimentos más interesantes

10.1. Newton: Convergencia Ideal desde $\mathbf{x_0} = [\mathbf{0}, \mathbf{90}]$

Este experimento muestra el mejor escenario y el máximo rendimiento del método de Newton. Encontró el mínimo global en un solo paso, ilustrando su capacidad de convergencia rápida cuando se inicia en la región correcta.

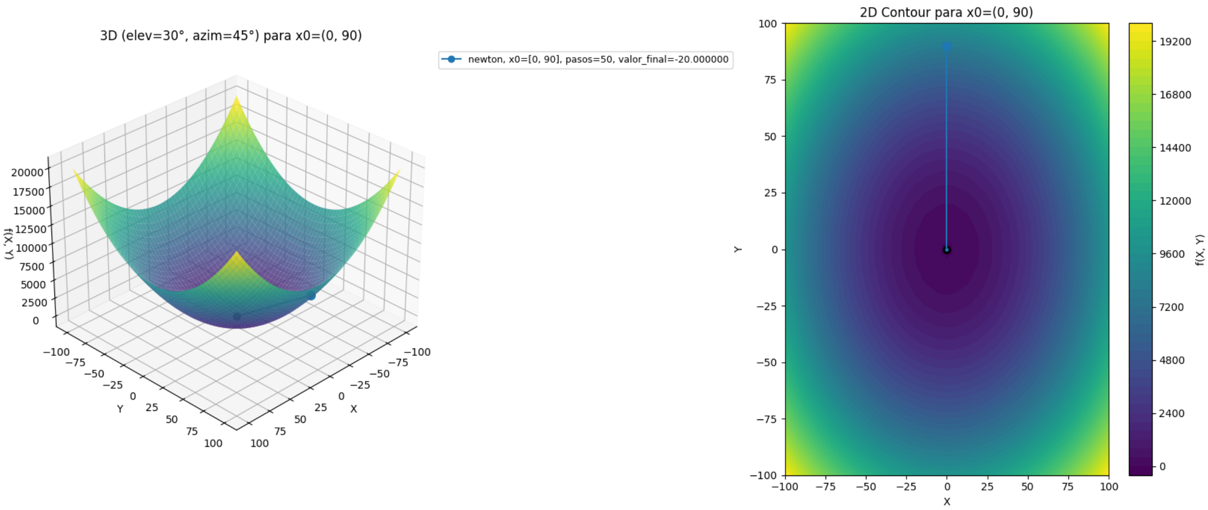

Figura 1: Gráfico de resultados del algoritmo de Newton en $\mathbf{x_0} = [\mathbf{0}, \mathbf{90}]$

10.2. Newton: Sensibilidad al punto inicial $\mathbf{x_0} = [\mathbf{50}, \mathbf{50}]$

Este experimento es el contraste directo del anterior. Muestra la extrema sensibilidad del método de Newton al punto de inicio. A pesar de su rapidez, solo se necesita un ligero cambio en x_0 para que converja a un punto estacionario incorrecto con un valor muy alto, ilustrando la inestabilidad del algoritmo en funciones no convexas.

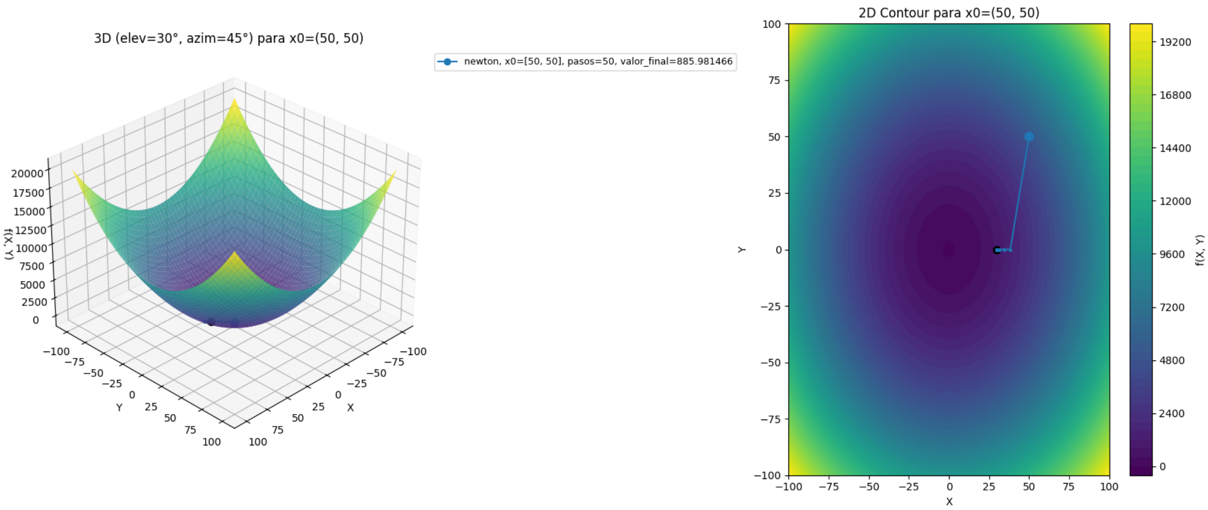

Figura 2: Gráfico de resultados del algoritmo de Newton en $\mathbf{x_0} = [\mathbf{50}, \mathbf{50}]$

10.3. Máximo Descenso (lr=0.1): Rendimiento robusto y eficiente

Este es el caso más eficiente del algoritmo de Máximo Descenso. Se puede comparar directamente con el caso ideal de Newton. Muestra que, con un learning rate (lr) bien elegido, el algoritmo converge al óptimo de manera segura, aunque es significativamente más lento que Newton, demostrando su robustez.

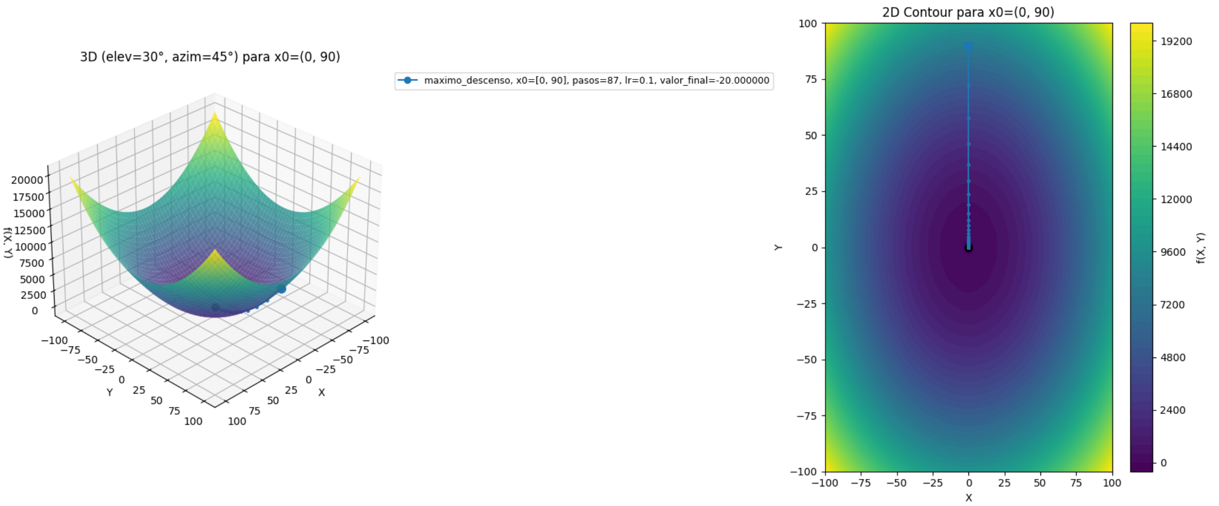

Figura 3: Gráfico de resultados del algoritmo de Máximo Descenso con lr=0.1

10.4. Máximo Descenso (lr=0.01): Convergencia lenta (1000 pasos)

Este experimento ilustra la sensibilidad a los parámetros. Demuestra que un lr diez veces menor sigue llevando a la convergencia, pero a cuesta muchas más iteraciones.

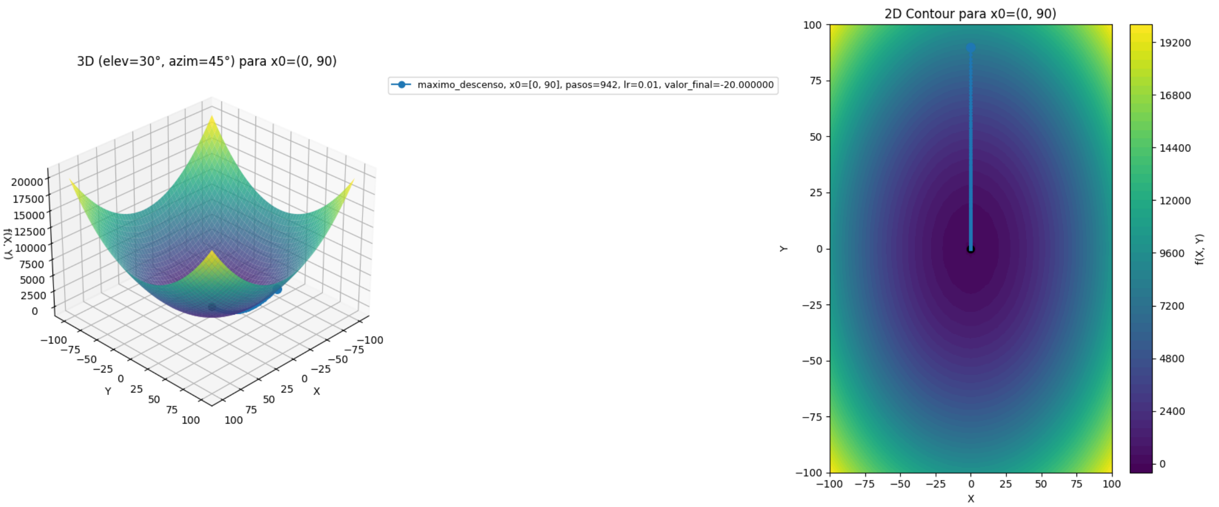

Figura 4: Gráfico de resultados del algoritmo de Máximo Descenso con lr=0.01

10.5. Máximo Descenso (lr=0.001): Avance insuficiente

Con un lr demasiado pequeño, aunque se realizan muchos pasos, el algoritmo no avanza lo suficiente y queda lejos del óptimo, evidenciando la importancia crítica de ajustar el lr.

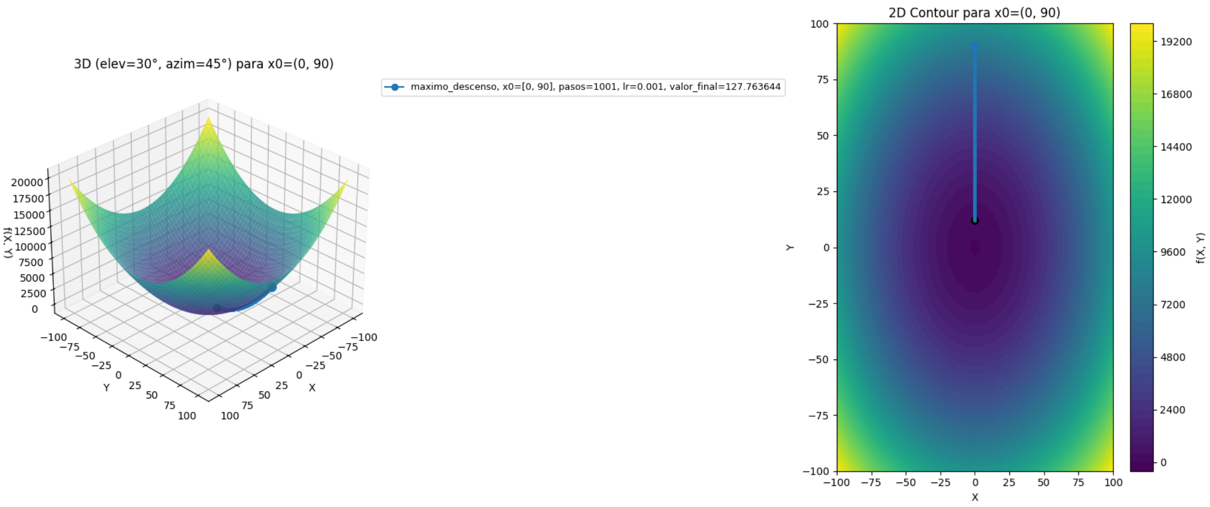

Figura 5: Gráfico de resultados del algoritmo de Máximo Descenso con lr=0.001