Problema 4: Gradiente Descendente en Optimizaciónl

Marco Alejandro Ramirez - Juan Sebastian Sanchez

21 de marzo de 2025

1. Formulación Matemática de los Problemas

Parte A: Implementación de Gradiente descendente en 3-D

Consideramos la función de pérdida:

$$L(x, y) = (x - 2)^2 + (y + 1)^2.$$

El objetivo es encontrar el mínimo de la función, el cual se conoce teóricamente en el punto $(2, -1)$.

Gradiente:

$$\nabla L(x, y) = \begin{bmatrix} 2(x - 2) \\ 2(y + 1) \end{bmatrix}.$$

Parte B: Comparación entre Newton-Raphson y Gradiente Descendente

1. **Definición de la función, gradiente y Hessiana:**
   La función en $\mathbb{R}^4$ se define como:
   $$f(x, y) = (x - 2)^2(y + 2)^2 + (x + 1)^2 + (y - 1)^2.$$
   Su gradiente es:
   $$\nabla f(x, y) = \begin{bmatrix} 2(x - 2)(y + 2)^2 + 2(x + 1) \\ 2(x - 2)^2(y + 2) + 2(y - 1) \end{bmatrix},$$

2. **Formulación del algoritmo de Newton-Raphson y Gradiente Descendente en $\mathbb{R}^3$:**
   Se implementó el método iterativo para Newton Raphson:
   $$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha\, H(\mathbf{x}_k)^{-1} \nabla f(\mathbf{x}_k),$$
   utilizando operaciones básicas de NumPy.También se implementó el método iterativo para Gradiente Descendiente:
   $$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha \mathbf{g_k},$$

3. **Criterio de parada:**
   La iteración se detiene cuando la norma del gradiente es menor que una tolerancia definida, es decir, $\|\nabla f(\mathbf{x}_k)\| < \epsilon$.

4. **Almacenamiento y análisis de la convergencia:**
   Se almacenó la trayectoria iterativa y se representó la convergencia mediante una gráfica que muestra la norma del gradiente versus el número de iteraciones.

2. Análisis de Resultados

Parte A: Implementación de Gradiente descendente en 3-D

El método de Gradiente Descendente, aplicado con el punto inicial $(0, 0)$, logró converger hacia una solución aproximada cercana al mínimo teórico en $(2, -1)$. La trayectoria iterativa, no fue posible graficarla sobre una superficie de la función de pérdida debido a la gran cantidad de trayectorias para los diferentes alpha. Por esto, se graficó en un plano mostrando la convergencia del método y en cuantas iteraciones para los diferentes alpha.

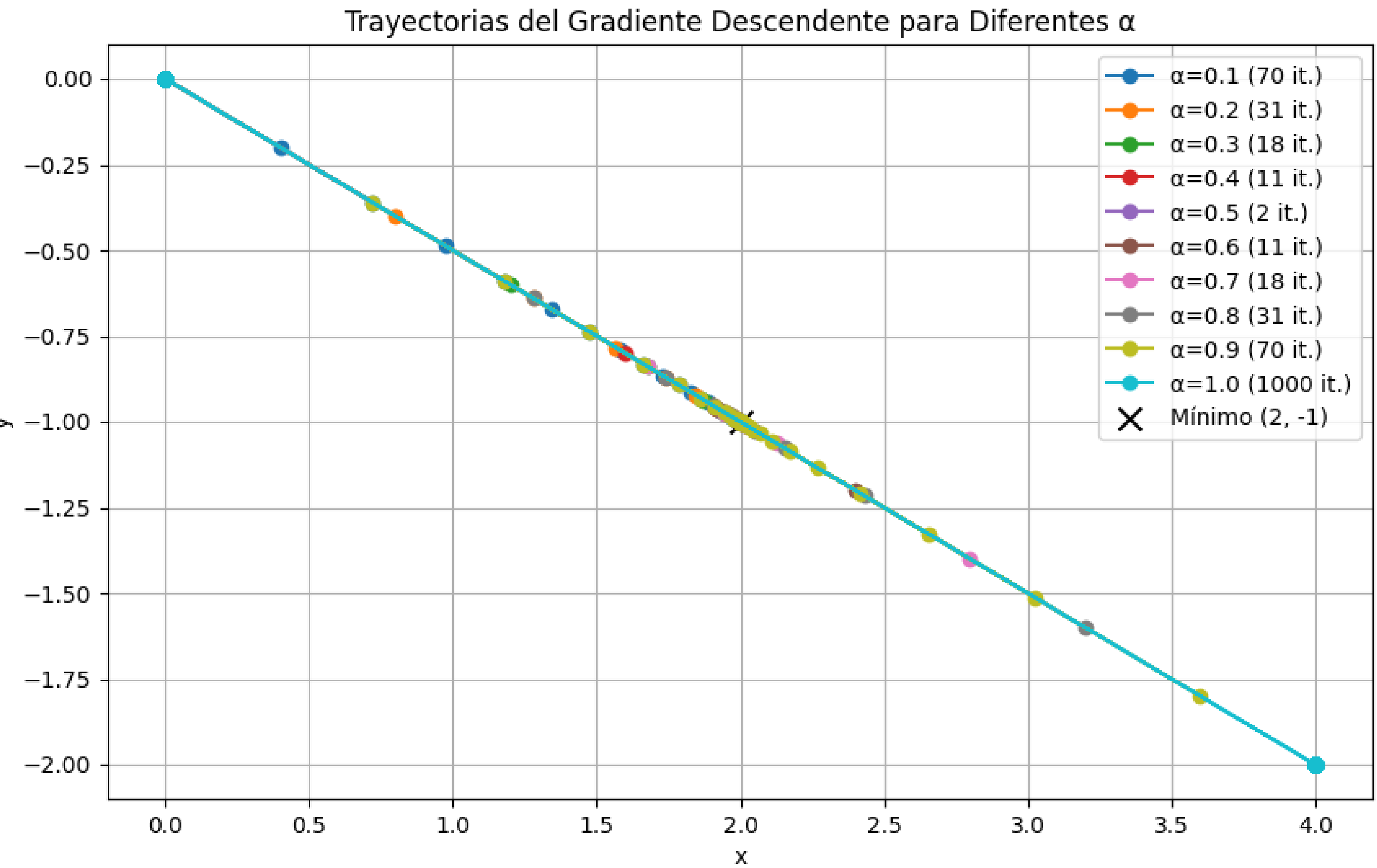

Figura 1: Función de pérdida y trayectoria iterativa para diferentes alpha

Parte B: Comparación entre Newton-Raphson y Gradiente Descendente

Para la función partiendo del punto inicial $(-2, -3)$, el método converge al mínimo global dependiendo del alpha. En el caso del descenso del gradiente, este valor nunca converge debido a que el gradiente crece demasiado rápido al ser una función cuadrática y partir de un punto distinto a (0,0). Encontramos que el valor optimo de alpha es 0.7 para el método de Newton Raphson en este caso específico ya que encuentra el mínimo global en el menor número de iteraciones. También notamos que en este caso no hay un valor optimo. Dependiendo del algoritmo que usemos y el alpha, va a cambiar nuestro resultado.Esto se debe a que nuestra funcion presenta un punto de silla. Además, en descenso del gradiente para alphas mayores a 0.1, el gradiente es demasiado grande y en cada iteración sigue creciendo, no convergiendo a nada. En la siguiente gráfica podemos ver las trayectorias de ambos algoritmos y la convergencia al error.

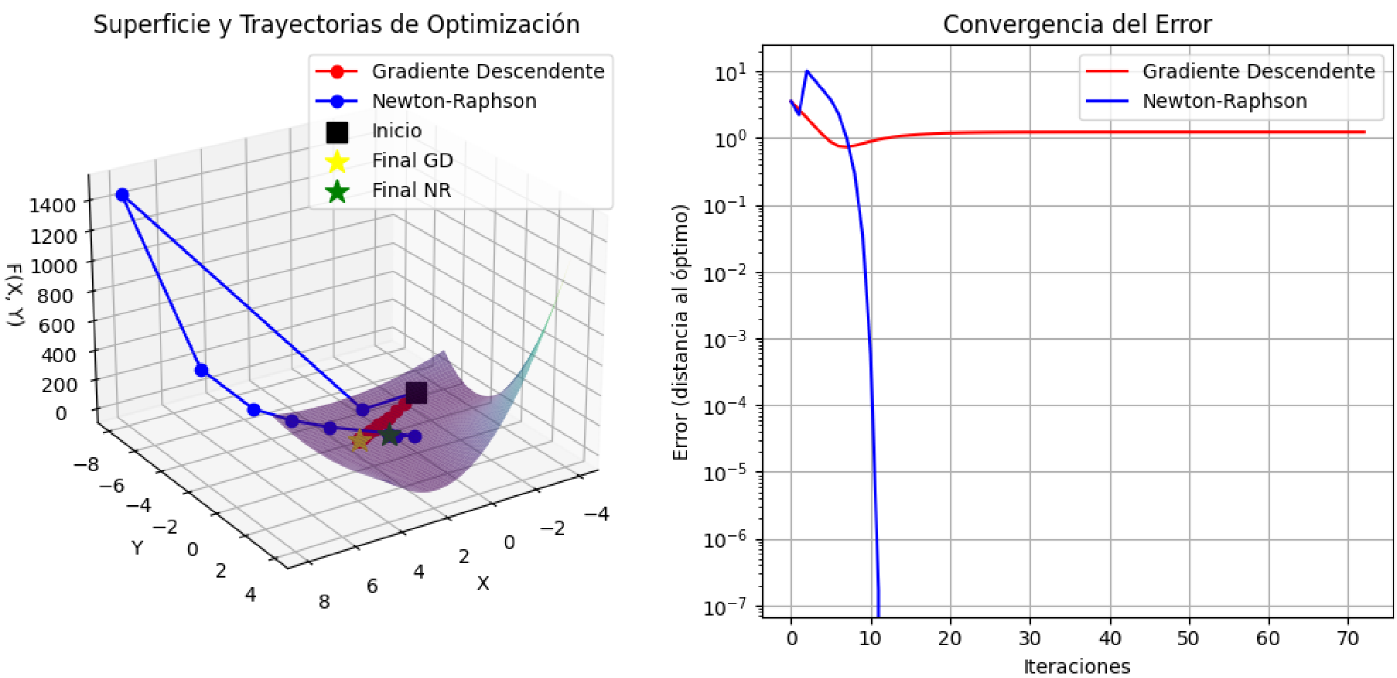

Figura 2: Trayectoria de ambos algoritmos y gráfica de convergencia del error

3. Conclusiones y Observaciones

- La implementación del método de Newton-Raphson y Descenso del Gradiente en espacios multidimensionales (3D) fue exitosa para encontrar mínimos de funciones no lineales.
- En la Parte A, la función de pérdida presentó una trayectoria iterativa que converge hacia el mínimo teórico $(2, -1)$, lo que se evidenció claramente en la gráfica 2D para los diferentes valores de alpha.
- En la Parte A, encontramos que el valor de alpha optimo es 0.5, encontrando la solución exacta en tan solo 2 iteraciones
- **Análisis de sensibilidad:**
  - Para valores lejanos al α óptimo ($> \alpha + 0{,}3$) hacen que la convergencia sea lenta.
  - Valores moderados de α ($\alpha + 0{,}1$ a $\alpha + 0{,}2$) permiten una convergencia rápida y estable.
  - Valores grandes de α ($\alpha > 1{,}0$) pueden causar oscilaciones o divergencia. Nunca converge.

  Para seleccionar óptimamente el α, podemos experimentar con diferentes valores y escoger el que nos dé la menor distancia al valor analítico en el menor número de iteraciones.

- Para la parte B, la función presenta un punto de silla, por lo cual, dependiendo del alpha y algoritmo que escojamos vamos a encontrar el mínimo global o nos quedaremos estancados en el punto de silla.
- Para el punto B, el alpha optimo fue de 0.7, encontrando la solución en tan solo 26 iteraciones con el algoritmo de Newton-Raphson.
- Para el caso del algoritmo del Descenso del Gradiente, para alphas mayores a 1, el gradiente es demasiado grande y el metodo no converge.
- La representación gráfica tanto de la trayectoria iterativa sobre la superficie como de la convergencia del gradiente es fundamental para analizar y comprender el comportamiento del algoritmo.