Laboratorio 3 Newton–Raphson y Gradiente Descendente

Marcos España, Jorge Bustamante

Marzo 2025

1 Problema 1

1.1 Formulación Matemática

La función a analizar es:

$$f(x) = 3x^3 - 10x^2 - 56x + 50 \tag{1}$$

1.2 Primera Derivada

La primera derivada, o gradiente en una dimensión, se calcula derivando cada término de $f(x)$ respecto a x:

$$\begin{aligned} f'(x) &= \frac{d}{dx}(3x^3) + \frac{d}{dx}(-10x^2) + \frac{d}{dx}(-56x) + \frac{d}{dx}(50) \\ &= 9x^2 - 20x - 56 \end{aligned}$$

Entonces:

$$f'(x) = 9x^2 - 20x - 56$$

1.3 Segunda Derivada

La segunda derivada, o Hessiana en una dimensión, se obtiene derivando $f'(x)$ respecto a x:

$$\begin{aligned} f''(x) &= \frac{d}{dx}(9x^2) + \frac{d}{dx}(-20x) + \frac{d}{dx}(-56) \\ &= 18x - 20 \end{aligned}$$

Entonces:

$$f''(x) = 18x - 20$$

1.4 Análisis Comportamiento

Table 1: Análisis de convergencia del método de Newton-Raphson

x_0	α	x^*	**Iteraciones**	$\|\nabla f(x^*)\|$	**Tipo de extremo**
-6	1.0	-1.619601	6	0.000000	Máximo
-6	0.6	-1.619601	24	0.000001	Máximo
-6	0.3	-1.619601	57	0.000001	Máximo
-2	1.0	-1.619601	4	0.000000	Máximo
-2	0.6	-1.619601	20	0.000001	Máximo
-2	0.3	-1.619601	49	0.000001	Máximo
0	1.0	-1.619601	6	0.000000	Máximo
0	0.6	-1.619601	19	0.000001	Máximo
0	0.3	-1.619601	51	0.000001	Máximo
2	1.0	3.841823	6	0.000000	Mínimo
2	0.6	3.841824	21	0.000001	Mínimo
2	0.3	3.841823	51	0.000001	Mínimo
6	1.0	3.841824	5	0.000000	Mínimo
6	0.6	3.841824	22	0.000001	Mínimo
6	0.3	3.841824	55	0.000001	Mínimo

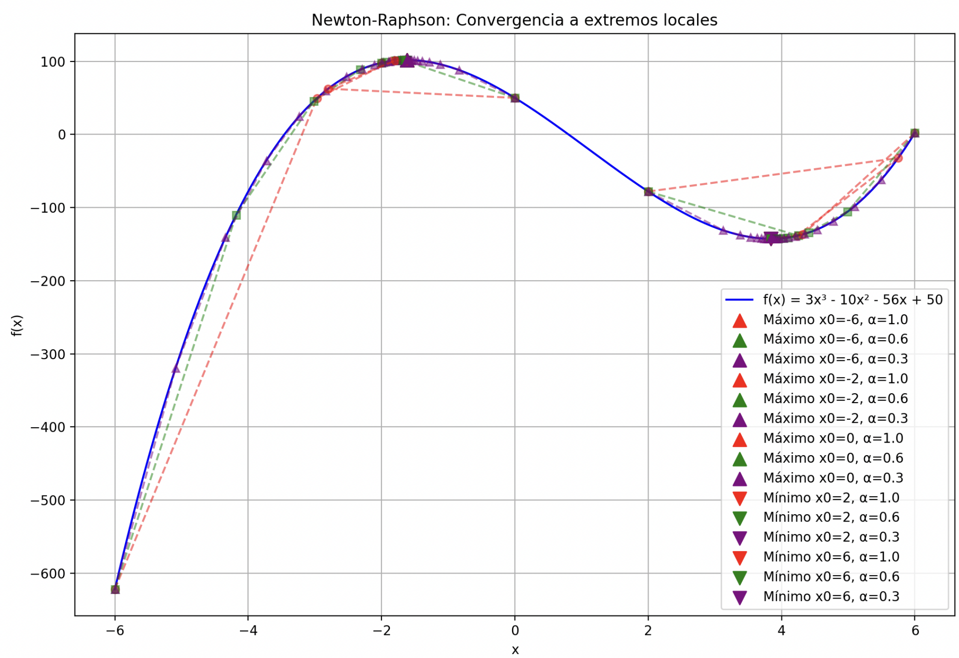
El análisis de convergencia del método de Newton-Raphson muestra que el punto inicial x_0 determina el tipo de extremo al que converge el algoritmo. Desde valores iniciales menores a $x = 2$, el método converge consistentemente a un máximo local en $x \approx -1.6196$, mientras que desde valores iniciales mayores, converge a un mínimo local en $x \approx 3.8418$. Además, se observa que el valor del parámetro de paso α influye directamente en la cantidad de iteraciones necesarias: valores mayores como $\alpha = 1.0$ permiten una convergencia rápida, mientras que valores menores como $\alpha = 0.3$ requieren más pasos, aunque mantienen la estabilidad. En todos los casos, la norma del gradiente al finalizar fue cercana a cero, lo que indica que el criterio de parada basado en $\|$
$derivada f(x)\| <$
e se cumple adecuadamente. Esto demuestra que el método de Newton-Raphson es eficiente, pero también altamente sensible al punto de inicio, lo cual debe considerarse cuidadosamente al buscar mínimos globales ya que un punto de inicio lejano hara que este tenga aun mas iteraciones.

2 Problema 2: Análisis de extremos locales y globales

2.1 Formulación Matemática

La función a analizar es:

$$f(x) = x^5 - 8x^3 + 10x + 6 \tag{2}$$



Newton-Raphson: Convergencia a extremos locales
f(x)
x
f(x) = 3x³ - 10x² - 56x + 50
Máximo x0=-6, α=1.0
Máximo x0=-6, α=0.6
Máximo x0=-6, α=0.3
Máximo x0=-2, α=1.0
Máximo x0=-2, α=0.6
Máximo x0=-2, α=0.3
Máximo x0=0, α=1.0
Máximo x0=0, α=0.6
Máximo x0=0, α=0.3
Mínimo x0=2, α=1.0
Mínimo x0=2, α=0.6
Mínimo x0=2, α=0.3
Mínimo x0=6, α=1.0
Mínimo x0=6, α=0.6
Mínimo x0=6, α=0.3

Su derivada primera es:

$$f'(x) = 5x^4 - 24x^2 + 10 \tag{3}$$

Y su derivada segunda es:

$$f''(x) = 20x^3 - 48x \tag{4}$$

El método de Newton-Raphson se define como:

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)} \tag{5}$$

2.2 Resultados y Gráficas

En la siguiente figura se muestran los extremos locales encontrados por el método de Newton-Raphson, junto con el máximo y mínimo global destacados en color rojo.

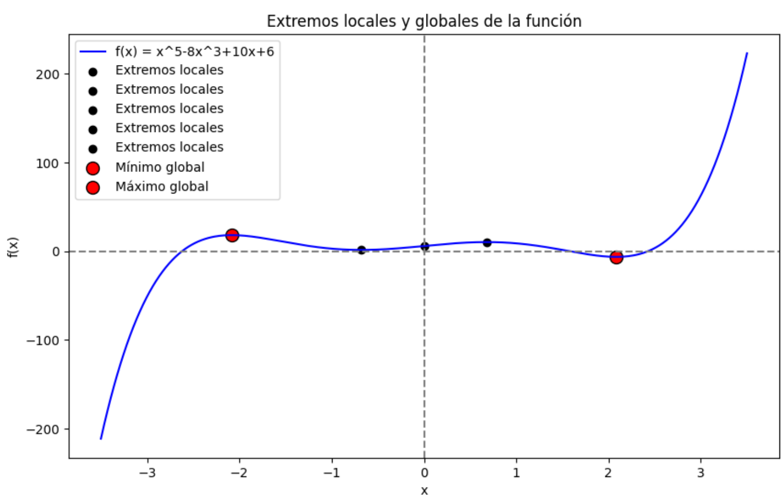


Figure 1: Gráfica de la función con los extremos locales y globales.

2.3 Análisis de Resultados

El método de Newton-Raphson ha permitido encontrar los puntos críticos de la función y clasificarlos en mínimos y máximos locales. El análisis de la segunda derivada ha permitido confirmar la concavidad en dichos puntos.

El comportamiento del método muestra buena convergencia para los valores iniciales elegidos. Sin embargo, en ciertos casos, la función puede tener puntos donde la segunda derivada es cero, lo que requiere un análisis especial.

2.4 Conclusiones

- Se identificaron los extremos locales de la función correctamente.
- Se verificó que el método converge bien para los valores iniciales seleccionados.
- Se observó que en ciertos casos puede ser necesario usar métodos adicionales cuando $f''(x) = 0$.
- La representación gráfica ayuda a visualizar claramente los extremos encontrados.

3 Problema 3: Newton-Raphson Multidimensional

3.1 Formulación Matemática Parte A

$$f(x, y) = (x - 1)^2 + 100(y - x^2)^2$$

El gradiente es:

$$\nabla f(x, y) = \begin{bmatrix} -2(1 - x) - 400x(y - x^2) \\ \\ 200(y - x^2) \end{bmatrix}$$

La matriz hessiana es:

$$H(x, y) = \begin{bmatrix} 2 - 400(y - x^2) + 800x^2 & -400x \\ \\ -400x & 200 \end{bmatrix}$$

3.2 Analisis de convergencia Parte A

A partir del punto inicial $[0, 10]$, se aplicó Newton-Raphson para minimizar la función de Rosenbrock, La trayectoria tiene una curva que se aproxima al fondo del valle de la función, pero no alcanza aconverger en 100 iteraciones, deteniéndose en el punto $\mathbf{x}^{=}[0.507, 0.250]$ con un valor de función $f(\mathbf{x}^*) \approx 0.2477$. Aunque el valor disminuyó significativamente desde el inicio ($f(\mathbf{x}_0) \gg 0$), no se alcanzó la tolerancia requerida para el gradiente, lo cual indica que el método quedó atrapado en una zona de curvatura fuerte. Esto nos demuestra que deberia haberse probado con distintos puntos iniciales y parametros α viendo como el algoritmo desciente por el valle pero aun se encuentra un poco lejos del minimo real

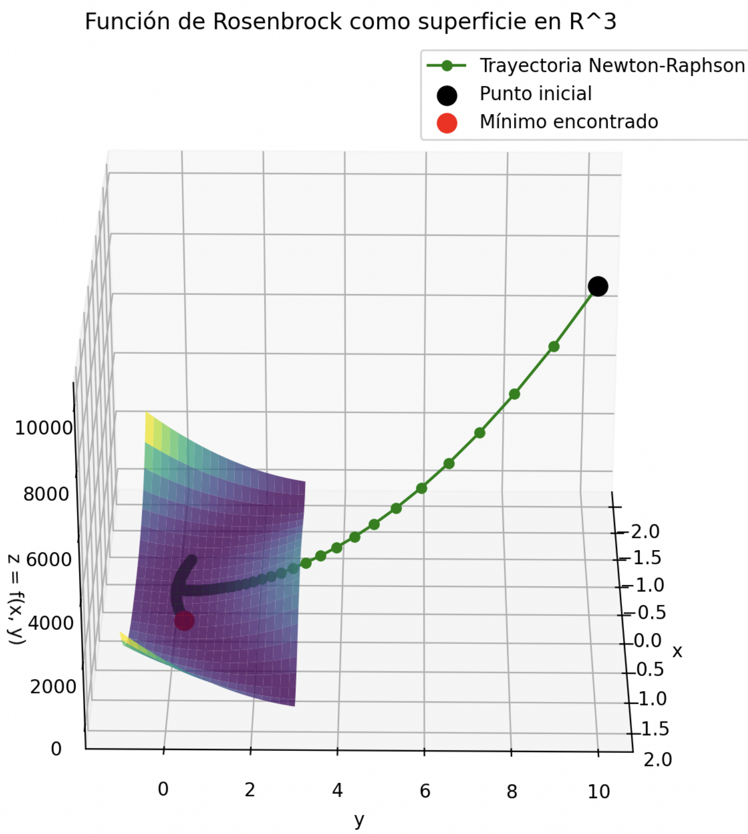


Figure 2: Trayectoria del método de Newton-Raphson en la función de Rosenbrock

3.3 Formulación Matemática Parte B

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

Funcion escogida:

$$f(x, y, z, w) = (x-1)^2 + 2(y-2)^2 + 3(z-3)^2 + 4(w-4)^2$$

Gradiente

$$\begin{bmatrix} \frac{\partial f}{\partial x} \\ \\ \frac{\partial f}{\partial y} \\ \\ \frac{\partial f}{\partial z} \\ \\ \frac{\partial f}{\partial w} \end{bmatrix} = \begin{bmatrix} 2(x-1) \\ \\ 4(y-2) \\ \\ 6(z-3) \\ \\ 8(w-4) \end{bmatrix}$$

Hessiana:

$$H(x,y,z,w) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & 0 & 0 & 0 \\ \\ 0 & \frac{\partial^2 f}{\partial y^2} & 0 & 0 \\ \\ 0 & 0 & \frac{\partial^2 f}{\partial z^2} & 0 \\ \\ 0 & 0 & 0 & \frac{\partial^2 f}{\partial w^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ \\ 0 & 4 & 0 & 0 \\ \\ 0 & 0 & 6 & 0 \\ \\ 0 & 0 & 0 & 8 \end{bmatrix}$$

3.4 Analisis de convergencia Parte B

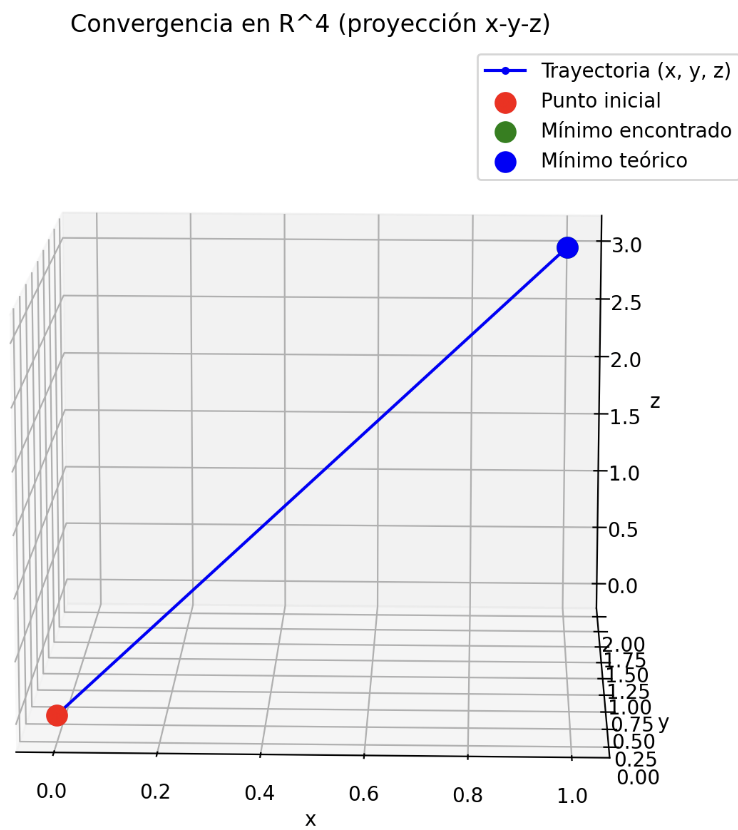


Figure 3: Enter Caption

En base al punto inicial $[0, 0, 0, 0]$ el algoritmo tan solo necesito 2 iteraciones para converger al minimo global teorico de la funcion usada $[1, 2, 3, 4]$, en la imagen se muestra la trayectoria tridimensional donde se observa un trayecto directo hacia el optimo. Este comportamiento es esperable ya que la función es estrictamente convexa, y su Hessiana es una matriz diagonal constante, bien condicionada y no singular. La estructura simple de la función permite que Newton-Raphson aproveche al máximo su convergencia cuadrática.

Todo esto implica un costo computacional de orden $\mathcal{O}(n^3)$, el cual, aunque manejable en cuatro dimensiones, se vuelve complejo en espacios de mayor dimensionalidad. Las representaciones gráficas mediante proyecciones en 3D resultan limitadas, pues inevitablemente pierden información al reducir la dimensionalidad del espacio original. El almacenamiento del historial completo de iteraciones genera un overhead de memoria que crece significativamente con la dimensionalidad del problema.

4 Problema 4: Gradiente Descendente en Optimización

4.1 Formulación Matemática Parte A

La función de pérdida utilizada es:

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

Gradiente:

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

Se implementó el algoritmo de gradiente descendente clásico, que actualiza los parámetros iterativamente según la regla:

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

4.2 Análisis de Resultados Parte A

Los experimentos con el método de gradiente descendente aplicado a $L(x, y) = (x - 2)^2 + (y + 1)^2$ revelan:

- **Convergencia óptima**: $\alpha = 0.50$ alcanza el mínimo global $(2, -1)$ en 1 iteración ($L = 0$)
- **Convergencia lenta**: $\alpha \in \{0.10, 0.90\}$ requieren más iteraciones
- **No convergencia**: Casos extremos ($\alpha = 0.01$ muy pequeño o $\alpha = 1.10$ muy grande)

La sensibilidad del parametro tiene 3 regiemens caracteristicos

Table 2: Comportamiento según valores de α

Rango	**Iteraciones**	**Estabilidad**
$\alpha < 0.1$	Excesivas	Estable
$0.1 \leq \alpha \leq 0.9$	Óptimas	Estable
$\alpha > 1.0$	Divergencia	Inestable

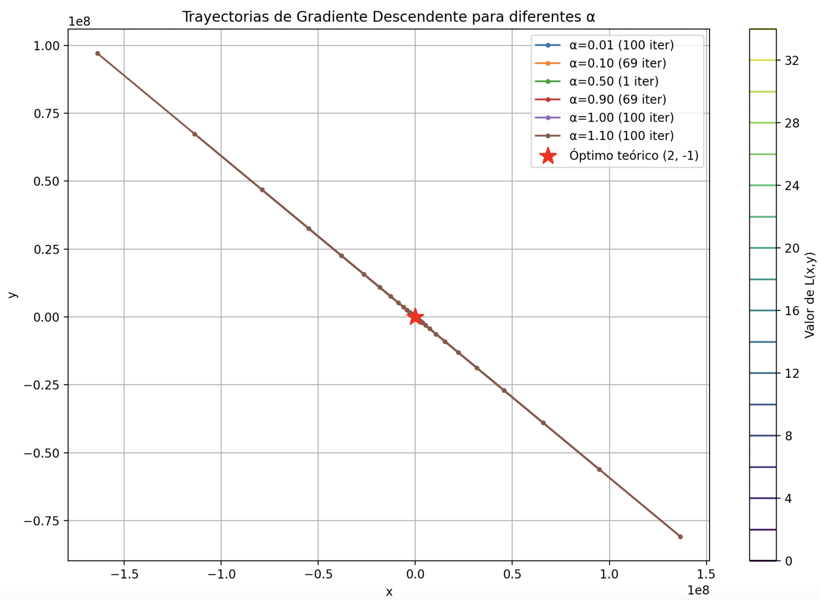
Estas son las estrategias recomenadas para escoger el α

1. **Búsqueda en rejilla**: Evaluar $\alpha \in [0.1, 0.9]$ con paso 0.1

2. **Decaimiento adaptativo**: $\alpha_k = \alpha_0/(1+k)$ combina rapidez inicial con precisión

3. **Métodos avanzados**: Adam/RMSProp para problemas no convexos

Conclusión: El valor $\alpha = 0.5$ demostró ser óptimo para esta función cuadrática, aunque métodos adaptativos son preferibles en casos generales.



4.3 Formulación Matemática Parte B

La funcion es esta:

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

A partir del punto inicial $(-2,-3)$, ambos métodos fueron evaluados con los mejores valores de α encontrados experimentalmente. Los resultados fueron:

- **Gradiente Descendente:** $\alpha = 0.050$, convergió en 217 iteraciones, con punto final $(-0.618, -1.618)$ y valor final $f = 8.000$.

- **Newton-Raphson:** $\alpha = 1.0$, convergió en solo 12 iteraciones, con punto final $(0.787, -0.787)$ y valor final $f = 8.551$.

Analisis de desempeño:

Criterio	**Gradiente Descendente**	**Newton-Raphson**
Iteraciones	217	12
Tiempo por iteración	0.000004 s	0.000011 s
Precisión final (error)	2.65×10^0	2.53×10^0
Robustez al parámetro α	Baja (muy sensible)	Alta (poco sensible)
Costo computacional por iteración	Bajo (solo gradiente)	Alto (requiere Hessiana)

Table 3: Comparación cuantitativa entre Gradiente Descendente y Newton-Raphson

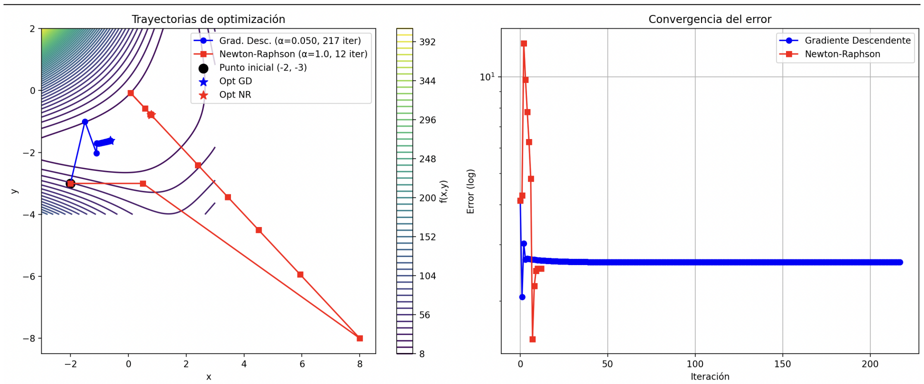


Figure 4: Enter Caption

Conclusion general:

- Newton-Raphson converge en menos iteraciones y más rápido en funciones bien comportadas, aunque cada iteración tiene mayor costo por requerir el cálculo y la inversión de la Hessiana.
- Gradiente Descendente es más estable con tasas de aprendizaje pequeñas, pero requiere muchas más iteraciones.
- Para problemas a gran escala, Gradiente Descendente puede ser preferible por su bajo costo computacional.
- Newton-Raphson es preferible cuando se dispone de la Hessiana y se requiere rapidez en convergencia.

5 Problema 5: Descenso de Gradiente y Descenso de Gradiente Basado en Momento

5.1 Análisis y Comparación

En esta sección se comparan los resultados de los dos métodos de entrenamiento implementados: **descenso de gradiente básico** (gradient) y **descenso de gradiente con momento** (moment). El objetivo es analizar la **velocidad de convergencia** y la **calidad de la aproximación a la función seno**.

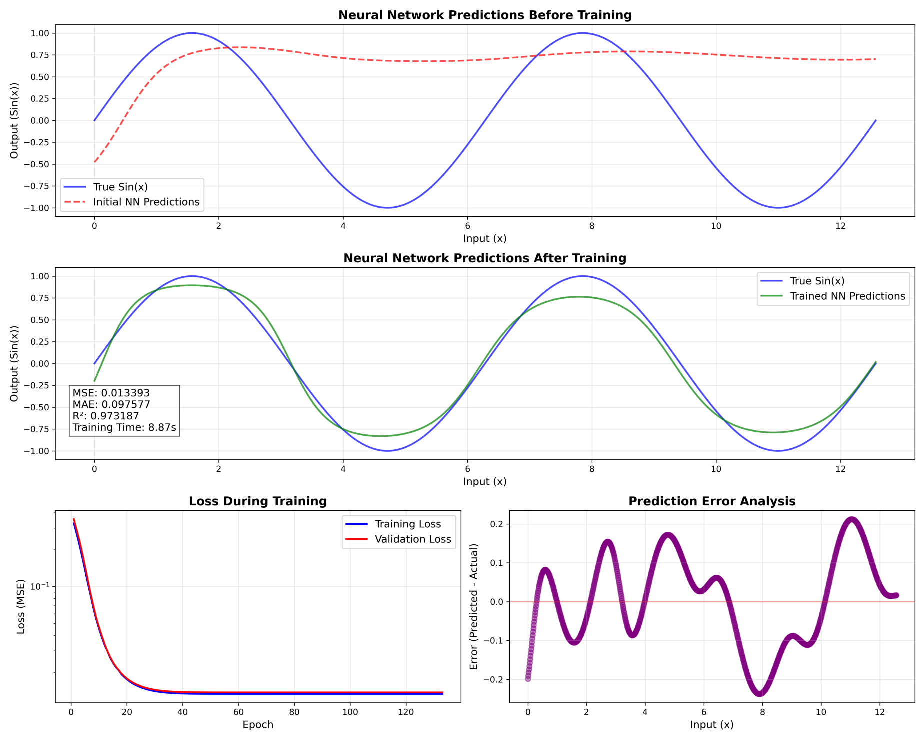


Figure 5: Arquitectura de la red neuronal usando descenso de gradiente básico.

En las Figuras 5 y 6 se muestran las redes neuronales diseñadas para cada uno de los métodos. Aunque la arquitectura base puede ser la misma (mismas capas y neuronas), la diferencia radica en el mecanismo de actualización de los pesos y sesgos.

Las Figuras 7 y 8 muestran la convergencia de la red hacia la función seno. Se pueden destacar las siguientes observaciones:

- **Velocidad de convergencia:** El método con momento (moment) suele requerir menos iteraciones para lograr un error similar o menor al método

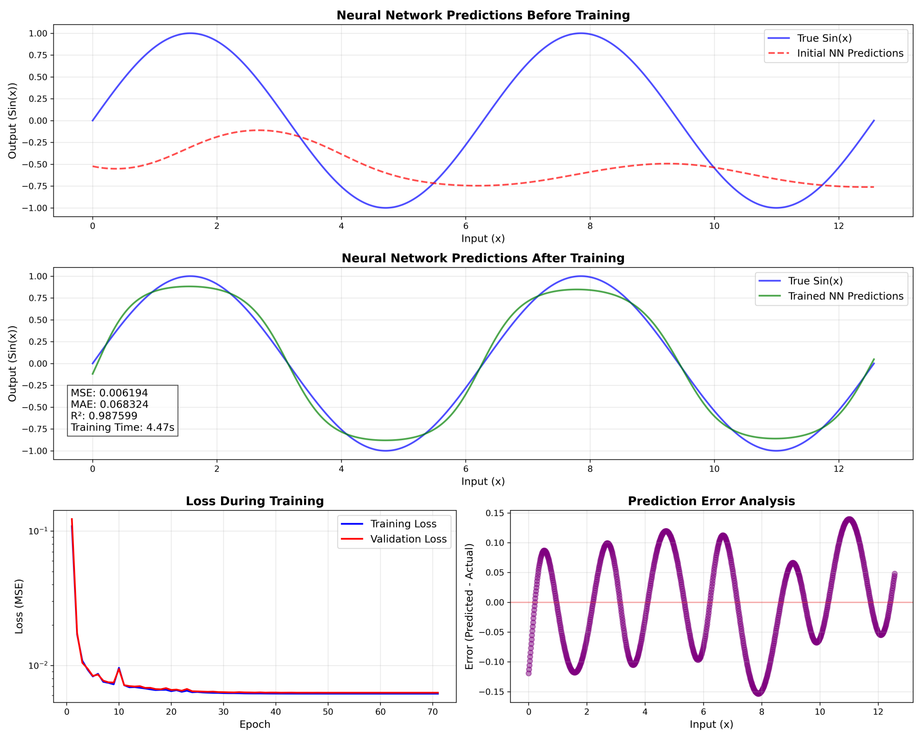


Figure 6: Arquitectura de la red neuronal usando descenso de gradiente con momento.

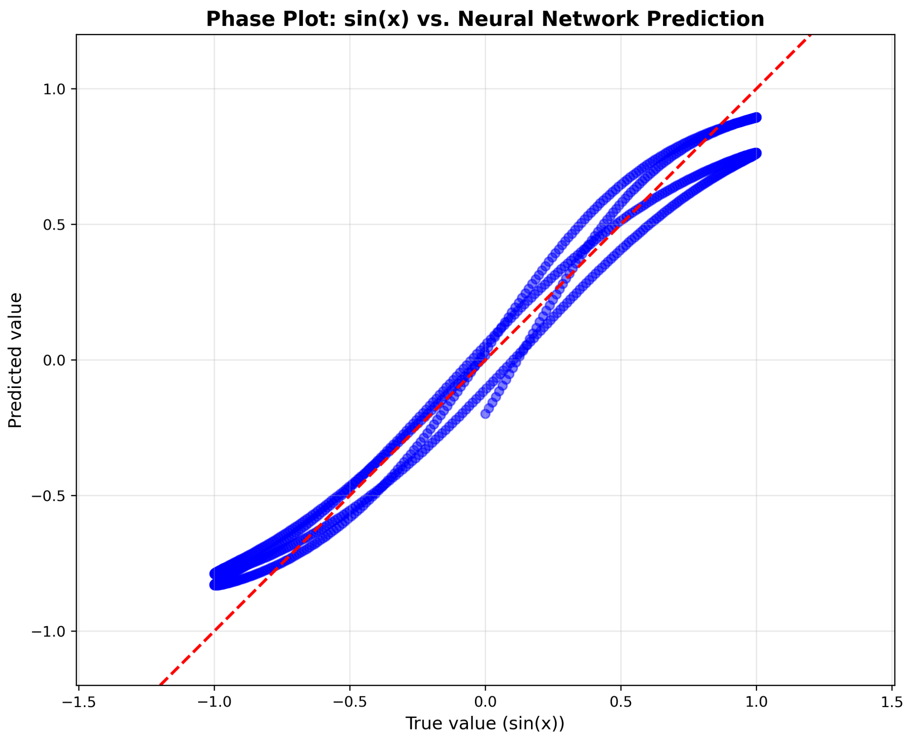


Figure 7: Evolución de la aproximación a la función seno con descenso de gradiente básico.

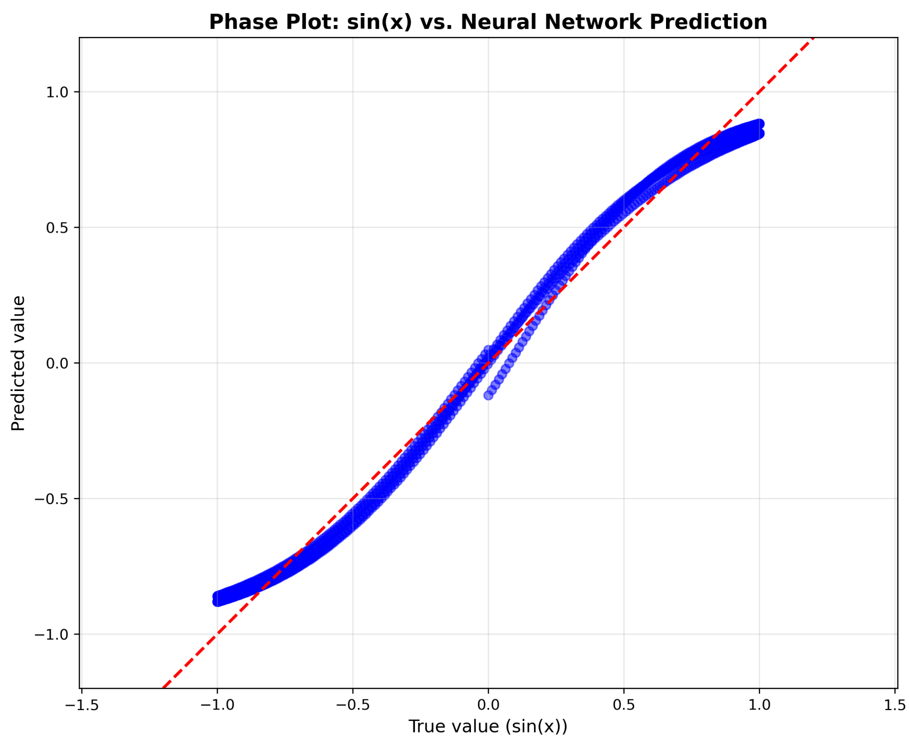


Figure 8: Evolución de la aproximación a la función seno con descenso de gradiente con momento.

básico (gradient). Esto se observa en la curva de entrenamiento, donde el método con momento desciende más rápidamente.

- **Calidad de la aproximación:** Ambos métodos convergen a una aproximación razonable de la función seno, pero el método con momento tiende a ser más estable frente a oscilaciones, lo que le permite llegar a mejores soluciones en menos tiempo.
- **Sensibilidad a hiperparámetros:** La tasa de aprendizaje (learning rate) y el coeficiente de momento (μ) pueden afectar significativamente la convergencia. Un valor de μ demasiado alto puede provocar oscilaciones, mientras que uno demasiado bajo se acerca al comportamiento del descenso de gradiente básico.

En conclusión, **el descenso de gradiente con momento** muestra un desempeño superior en términos de **velocidad de convergencia** y **estabilidad** en la aproximación a la función seno, aunque el método básico también alcanza resultados aceptables si se sintonizan correctamente los hiperparámetros.