Tarea de Optimización

Frank Piz

November 2025

1 Análisis Teórico del Modelo

Función Objetivo

$$f(x,y) = 0.1\left(e^{(x^2+y^2)} - \tan(1.5\sin(x+y))\right)$$

Definimos para simplificar:

$$r(x,y) := x^2 + y^2, \qquad u(x,y) := 1.5\sin(x+y).$$

1. Tipo de Variables y Dominio

Las variables x y y son **continuas reales**, $x, y \in \mathbb{R}$.

El argumento de la tangente $u(x,y)$ está acotado en $[-1.5, 1.5]$. Como $|1.5| < \pi/2 \approx 1.5708$, la tangente nunca se indetermina. Por tanto, la función f está bien definida y es **suave e infinitamente diferenciable** (C^∞) en todo $\mathbb{R}^2$.

2. Existencia de Solución y de Óptimo Global

El término $e^{x^2+y^2}$ domina para valores grandes. Dado que f es continua y $\lim_{\|(x,y)\|\to\infty} f(x,y) = +\infty$ (coercitividad), se garantiza la existencia de **al menos un mínimo global** (Teorema de Weierstrass, extendido a funciones coercitivas en $\mathbb{R}^n$). No existen máximos globales.

3. Gradiente (∇f) y Puntos Críticos

Gradiente (Condición Necesaria de Primer Orden)

$$\nabla f(x,y) = 0.1\begin{pmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{pmatrix} = 0.1\begin{pmatrix} 2x\,e^r - 1.5\cos(x+y)\,\sec^2(u) \\ 2y\,e^r - 1.5\cos(x+y)\,\sec^2(u) \end{pmatrix}.$$

Deducción de Puntos Críticos Para que (x^*, y^*) sea un punto crítico, $\nabla f(x^*, y^*) = \mathbf{0}$, lo que implica igualar a cero ambas componentes:

$$\frac{\partial f}{\partial x} = 0 \quad \text{y} \quad \frac{\partial f}{\partial y} = 0.$$

Restando las dos ecuaciones, se obtiene:

$$0.1\left(2x^* e^{r^*} - 1.5\cos(\dots)\right) = 0.1\left(2y^* e^{r^*} - 1.5\cos(\dots)\right)$$

Esto se simplifica a $2x^* e^{r^*} = 2y^* e^{r^*}$, lo cual implica que $**x^* = y^{***}$ (dado que $e^{r^*} > 0$). La condición unidimensional para los puntos críticos es:

$$2x\, e^{2x^2} = 1.5\cos(2x)\, \sec^2\big(1.5\sin(2x)\big).$$

4. Matriz Hessiana (H_f) y Test de Segundo Orden

Hessiana La matriz Hessiana H_f es simétrica por el Teorema de Clairaut (dado que f es C^∞):

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}.$$

Las componentes no idénticas son:

$$\begin{aligned}
\frac{\partial^2 f}{\partial x^2} &= 0.1\Big((2+4x^2)e^r - [\,4.5\cos^2(x+y)\sec^2(u)\tan(u) - 1.5\sin(x+y)\sec^2(u)\,]\Big),\\
\frac{\partial^2 f}{\partial y^2} &= 0.1\Big((2+4y^2)e^r - [\,4.5\cos^2(x+y)\sec^2(u)\tan(u) - 1.5\sin(x+y)\sec^2(u)\,]\Big),\\
\frac{\partial^2 f}{\partial x \partial y} &= 0.1\Big(4xy\, e^r - [\,4.5\cos^2(x+y)\sec^2(u)\tan(u) - 1.5\sin(x+y)\sec^2(u)\,]\Big).
\end{aligned}$$

Condiciones Suficientes de Segundo Orden En un punto crítico (x^*, y^*) donde $\nabla f(x^*, y^*) = \mathbf{0}$:

- Si H_f es **definida positiva** $\Rightarrow$ mínimo local estricto.
- Si H_f es **indefinida** $\Rightarrow$ punto de silla.

5. Conclusiones sobre Óptimos y Región de Prueba

Comportamiento General La función es no convexa. La **coercitividad** garantiza la existencia de un mínimo global. La simetría y la naturaleza oscilante implican que los puntos críticos se encuentran en la recta $x = y$ y que es probable que existan **múltiples mínimos locales** cerca del origen.

Análisis en la Región $[-100, 100]^2$ En este dominio, el término exponencial $e^{x^2+y^2}$ domina rápidamente fuera de una pequeña región alrededor del origen (por ejemplo, $|x|, |y| < 2$). Esto implica:

- Los **mínimos** (locales y globales) se ubican en el **interior** de esta región, cerca de $(0, 0)$.
- El resto de la región presenta un crecimiento monótono hacia $+\infty$.

El problema de optimización se reduce a una búsqueda local eficiente a lo largo de la diagonal $x = y$ cerca del origen.

2 Descripción de los Algoritmos Utilizados

En esta sección se describen los dos métodos de optimización aplicados a la función

$$f(x, y) = 0.1\big(e^{x^2+y^2} - \tan(1.5 \sin(x + y))\big),$$

dentro del dominio de prueba $[-100, 100]\times[-100, 100]$. Ambos algoritmos buscan minimizar $f(x, y)$, aprovechando su suavidad y la existencia de gradiente e información de curvatura.

4.1. Método de Descenso del Gradiente

Descripción general El método de **descenso del gradiente** es un algoritmo iterativo de optimización que actualiza las variables en la dirección opuesta al gradiente, buscando la región de menor valor de la función. En cada iteración:

$$(x_{k+1}, y_{k+1}) = (x_k, y_k) - \alpha_k \nabla f(x_k, y_k),$$

donde $\alpha_k > 0$ es la **tasa de aprendizaje** o **tamaño de paso**.

Criterio de parada El algoritmo se detiene cuando se cumple alguna de las siguientes condiciones:

- $\|\nabla f(x_k, y_k)\| < \varepsilon$ (el gradiente es casi nulo),
- el número máximo de iteraciones es alcanzado,
- o los cambios en $f(x, y)$ entre iteraciones son despreciables.

Uso en la función dada Para la función $f(x, y)$, el gradiente es continuo y fácilmente computable:

$$\nabla f(x, y) = 0.1 \begin{pmatrix} 2xe^{x^2+y^2} - 1.5 \cos(x + y) \sec^2(1.5 \sin(x + y)) \\ 2ye^{x^2+y^2} - 1.5 \cos(x + y) \sec^2(1.5 \sin(x + y)) \end{pmatrix}.$$

En la práctica:

- Se elegirá un punto inicial (x_0, y_0) dentro de $[-2, 2]$, pues fuera de esa zona la función crece rápidamente.

- El tamaño de paso α_k puede ser fijo (por ejemplo, 10^{-3}) o adaptativo mediante búsqueda en línea (*line search*) tipo Armijo.
- Se espera convergencia hacia el mínimo global cercano al origen.
- Sin embargo, debido al término oscilante $-\tan(1.5\sin(x+y))$, el método puede estancarse en mínimos locales si el punto inicial no está suficientemente cerca del mínimo global.

Ventajas y desventajas

- Ventajas: fácil de implementar, requiere solo el gradiente, bajo costo computacional por iteración.
- Desventajas: puede converger lentamente o quedar atrapado en mínimos locales; sensible al tamaño de paso.

4.2. Método Quasi-Newton (BFGS)

Descripción general El método **Quasi-Newton** (en particular el algoritmo BFGS) mejora el descenso del gradiente aproximando la información de la Hessiana sin calcularla explícitamente. La dirección de búsqueda p_k se obtiene como:

$$p_k = -B_k^{-1}\nabla f(x_k, y_k),$$

donde B_k es una matriz que aproxima la Hessiana $H_f(x_k, y_k)$. La actualización de B_k se realiza mediante:

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k},$$

con

$$s_k = x_{k+1} - x_k, \qquad y_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

Uso en la función dada Para $f(x, y)$:

- El gradiente $\nabla f(x, y)$ se utiliza directamente en cada iteración.
- No se requiere calcular la Hessiana exacta, lo que reduce el costo computacional frente al método de Newton puro.
- La función es suave y coercitiva, por lo que BFGS converge localmente de manera superlineal cerca de los mínimos.
- Debido a su actualización adaptativa de curvatura, puede escapar más fácilmente de mínimos locales débiles que el descenso del gradiente.

Criterio de parada Mismos criterios que en el descenso del gradiente:

$$\|\nabla f(x_k, y_k)\| < \varepsilon, \quad \text{o bien un número máximo de iteraciones.}$$

Ventajas y desventajas

- Ventajas: converge más rápido que el descenso del gradiente, no requiere Hessiana exacta, robusto ante curvaturas pronunciadas.
- Desventajas: mayor costo por iteración, necesita almacenamiento y operaciones matriciales (aunque en este caso $n = 2$).

4.3. Comparación esperada entre ambos métodos

- El **descenso del gradiente** alcanzará el mínimo global si parte cerca del origen, pero puede requerir muchos pasos para valores de α pequeños.
- El **método Quasi-Newton (BFGS)** convergerá en menos iteraciones gracias a su aproximación de la curvatura local.
- En la región $[-100, 100]^2$, ambos métodos deben detenerse antes de alcanzar los bordes, pues $f(x, y)$ crece rápidamente y no hay óptimos fuera del entorno del origen.

Conclusión general Ambos algoritmos son adecuados para el problema:

- **Descenso del Gradiente:** simple, estable, pero más lento.
- **Quasi-Newton (BFGS):** más eficiente, aprovecha información de curvatura y tiende a converger más rápido hacia el mínimo global de f.