

Tarea: Análisis y Aplicación de Algoritmos de Optimización

Sebastian González Alfonso
Ciencia de la Computación
Tercer Año
Modelos de Optimización

11 de noviembre de 2025

Índice

1. Marco Teórico	2
2. Análisis de la función	2
2.1. Regularidad	2
2.2. Gradiente	2
2.3. Puntos críticos	2
2.4. Hessiano y clasificación del punto crítico	3
2.5. Mínimo global y comportamiento en el infinito	3
2.6. Convexidad	3
2.7. Consecuencias prácticas para algoritmos	3
2.8. Interpretación geométrica	4
2.9. Cotas y comportamiento asintótico	5
2.10. Campo de gradientes	6

1. Marco Teórico

hi

2. Análisis de la función

Se considera la función

$$f(x, y) = -200 e^{-0,02\sqrt{x^2+y^2+1}}, \quad (x, y) \in \mathbb{R}^2.$$

2.1. Regularidad

Definiendo

$$s_1(x, y) := \sqrt{x^2 + y^2 + 1},$$

la función s_1 es de clase C^2 en \mathbb{R}^2 (no hay singularidades porque $x^2 + y^2 + 1 \geq 1$, manteniendo siempre definidas s_1 y s'_1). Como la composición de funciones C^2 es C^2 , sabiendo que $s_2(x) := e^x$ y $\forall a \in \mathbb{R} : s_3(x) := ax$ son C^2 , se tiene

$$f \in C^2(\mathbb{R}^2).$$

En particular, f es continua y diferenciable de todas las órdenes en todo \mathbb{R}^2 .

2.2. Gradiente

Sea $r := s_1(x, y) = \sqrt{x^2 + y^2 + 1}$. Aplicando la regla de la cadena,

$$\frac{\partial f}{\partial x} = -200 \cdot (-0,02) e^{-0,02r} \cdot \frac{\partial r}{\partial x} = 4 e^{-0,02r} \frac{x}{r},$$

y análogamente, dada la simetría de la función:

$$\frac{\partial f}{\partial y} = 4 e^{-0,02r} \frac{y}{r}.$$

Por tanto, podemos escribir el gradiente de forma compacta como

$$\nabla f(x, y) = \frac{4e^{-0,02r}}{r} \begin{pmatrix} x \\ y \end{pmatrix}.$$

2.3. Puntos críticos

Las ecuaciones críticas $\nabla f(x, y) = 0_2$ conducen a

$$\frac{4e^{-0,02r}}{r} x = 0, \quad \frac{4e^{-0,02r}}{r} y = 0.$$

Dado que $\forall (x, y) \in \mathbb{R}^2 : \frac{4e^{-0,02r}}{r} > 0$, la única solución posible es:

$$(x, y) = (0, 0).$$

Por tanto, el único punto crítico existente en \mathbb{R}^2 es el origen.

2.4. Hessiano y clasificación del punto crítico

Evaluable en el origen $(x, y) = (0, 0)$ (donde $r = 1$) se obtiene el Hessiano

$$\nabla^2 f(0, 0) = 4e^{-1/50} I \quad \Rightarrow \quad f_{xx}(0, 0) = f_{yy}(0, 0) = 4e^{-1/50}, \quad f_{xy}(0, 0) = 0,$$

con autovalores positivos iguales: $4e^{-1/50} > 0$. Por tanto el punto crítico $(0, 0)$ es un **mínimo local**.

2.5. Mínimo global y comportamiento en el infinito

Calculemos valores extremos:

$$f(0, 0) = -200e^{-0,02\sqrt{0+0+1}} = -200e^{-0,02} \approx -200 \cdot 0,98019867 \approx -196,0397.$$

Para cualquier (x, y) se cumple $r = \sqrt{x^2 + y^2 + 1} \geq 1$, y la función $r \mapsto e^{-0,02r}$ es estrictamente decreciente en r . Por tanto f alcanza su valor mínimo en r mínimo, es decir en $(0, 0)$. Además, para $\|(x, y)\| \rightarrow \infty$ se tiene $r \rightarrow \infty$ y

$$f(x, y) = -200e^{-0,02r} \rightarrow 0^-.$$

Esto implica:

- $\inf_{(x,y) \in \mathbb{R}^2} f(x, y) = f(0, 0)$, y el mínimo es **global y único**.
- $\sup_{(x,y) \in \mathbb{R}^2} f(x, y) = 0$, pero *no* se alcanza: $f(x, y) < 0$ para todo (x, y) y $\lim_{\|(x,y)\| \rightarrow \infty} f(x, y) = 0$. Por tanto no existe máximo global alcanzado en \mathbb{R}^2 (el supremo es 0).

2.6. Convexidad

Para estudiar la convexidad global debe comprobarse la positividad semidefinida del Hessiano en todo \mathbb{R}^2 . Como la autovalor tangencial $\lambda_{\text{tang}}(r) > 0$ para todo $r \geq 1$ pero la autovalor radial $\lambda_{\text{rad}}(r)$ cambia de signo en $r_0 \approx 3,7745$, concluimos que:

- En la bola centrada en el origen de radio $u_{\text{crit}} = \sqrt{r_0^2 - 1} \approx 3,6396$ (es decir para $\|(x, y)\| < 3,6396$) el Hessiano es definido positivo y f es localmente convexa en esa región.
- Para $\|(x, y)\| > 3,6396$ la autovalor radial es negativa, por lo que el Hessiano es indefinido y la función *no* es convexa globalmente.

En particular, f no es ni convexa ni cóncava en todo \mathbb{R}^2 .

2.7. Consecuencias prácticas para algoritmos

- Dado que existe un único mínimo global en $(0, 0)$ y la función es suave, los métodos locales bien diseñados (gradiente descendente con decaimiento apropiado del paso, métodos de Newton con buen manejo de Hessiano, etc.) tienen una buena probabilidad de converger hacia el mínimo si el punto inicial está dentro de la cuenca de atracción que incluye $(0, 0)$.
- No obstante, la no convexidad global (cambio de signo de la curvatura radial para $\|(x, y)\| > 3,6396$) indica que:

- Métodos que suponen convexidad global pueden comportarse mal o fallar fuera de la región convexa (por ejemplo, estimaciones de paso óptimo basadas en convexidad).
- Es importante ensayar distintos puntos iniciales y tamaños de paso. También es aconsejable monitorear la curvatura (autovalores del Hessiano) en iteraciones de segundo orden.

Resumen: f es C^2 , tiene un único punto crítico en $(0, 0)$ que es mínimo global, no posee máximo alcanzado (supremo 0), y es localmente convexa cerca del origen (bola de radio aproximado 3,6396) pero no convexa globalmente debido a la curvatura radial negativa para $r > r_0$.

2.8. Interpretación geométrica

La función analizada

$$f(x, y) = -200e^{-0,02\sqrt{x^2+y^2+1}}$$

depende únicamente del término $\sqrt{x^2 + y^2}$, es decir, de la *distancia euclidiana al origen*. Por lo tanto, puede clasificarse como una función *radialmente simétrica*, en la que los valores de f permanecen constantes para todos los puntos equidistantes del origen.

Curvas de nivel. Las curvas de nivel o *iso-contornos* están dadas por el conjunto de puntos (x, y) que satisfacen

$$f(x, y) = c \quad \Rightarrow \quad -200e^{-0,02\sqrt{x^2+y^2+1}} = c.$$

De esta expresión se deduce que cada nivel c corresponde a una circunferencia de radio

$$r(c) = \sqrt{\left[-50 \ln\left(-\frac{c}{200}\right)\right]^2 - 1}, \quad c \in (-200, 0).$$

Así, todas las curvas de nivel son *circunferencias concéntricas* centradas en el origen, lo cual evidencia la estructura perfectamente simétrica del modelo.

Forma tridimensional. En el espacio tridimensional $(x, y, f(x, y))$, la superficie adopta la forma de un *cuenco suave* o *superficie de potencial radial decreciente*. El valor máximo (menos negativo) se alcanza cuando $r \rightarrow \infty$, donde la función tiende asintóticamente a 0, mientras que el mínimo absoluto ocurre en el origen, con

$$f(0, 0) = -200e^{-0,02} \approx -196,04.$$

Esta forma implica que la superficie desciende de manera monótona desde el exterior hacia el centro, sin presentar valles secundarios ni crestas locales, lo que anticipa la existencia de un único mínimo global.

Comportamiento de la pendiente. La pendiente de la función, representada por la magnitud del gradiente,

$$\|\nabla f(x, y)\| = 4e^{-0,02\sqrt{x^2+y^2+1}},$$

disminuye de forma exponencial conforme aumenta la distancia al origen. Esto significa que la superficie es empinada cerca del centro y cada vez más plana en las regiones periféricas. Desde el punto de vista geométrico, este comportamiento genera una *zona de alta curvatura* en el entorno del mínimo, seguida de un *amplio altiplano* donde la función varía lentamente. Tal estructura tiene implicaciones directas en la dinámica de los algoritmos de optimización: el gradiente descendente experimentará convergencia rápida cerca del origen y muy lenta a grandes distancias, debido a la suavidad creciente de la función.

Interpretación física. Geométricamente, puede interpretarse como un modelo de *campo de potencial negativo* cuya intensidad disminuye exponencialmente con la distancia al centro. El término -200 actúa como factor de escala del potencial, mientras que el coeficiente $0,02$ regula la rapidez con la que dicho potencial se disipa. Por analogía con la física, el origen representa una *zona de atracción estable*, hacia la cual convergen las trayectorias definidas por el campo de gradiente.

En síntesis, la función define una superficie radialmente simétrica, suavemente decreciente y sin irregularidades topológicas.

2.9. Cotas y comportamiento asintótico

Dado que la función está definida por un término exponencial negativo, su rango de valores es finito y puede determinarse directamente a partir de las propiedades de la función exponencial.

Cotas globales. Para todo $(x, y) \in \mathbb{R}^2$,

$$0 < e^{-0,02\sqrt{x^2+y^2+1}} \leq e^{-0,02},$$

ya que el exponente $-0,02\sqrt{x^2+y^2+1}$ toma valores en el intervalo $(-\infty, -0,02]$. Multiplicando por el factor -200 , se obtiene:

$$-200e^{-0,02} \leq f(x, y) < 0.$$

Por lo tanto, el rango de la función es:

$$f(\mathbb{R}^2) = (-200e^{-0,02}, 0) \approx (-196,04, 0).$$

La función está así acotada superiormente por 0 y posee un mínimo global en $f(0, 0)$.

Comportamiento asintótico. Al analizar el límite cuando la distancia al origen tiende a infinito, se tiene:

$$\lim_{\sqrt{x^2+y^2} \rightarrow \infty} f(x, y) = -200 \lim_{r \rightarrow \infty} e^{-0,02\sqrt{r^2+1}} = 0^-.$$

Esto implica que la superficie se aproxima al plano $f = 0$ sin llegar a tocarlo, generando un *decaimiento exponencial* de la magnitud de f .

En el extremo opuesto, cuando $r = \sqrt{x^2 + y^2} \rightarrow 0$:

$$\lim_{r \rightarrow 0} f(x, y) = -200e^{-0,02} \approx -196,04,$$

confirmando que el mínimo absoluto ocurre en el origen.

Conclusión del análisis asintótico. El término exponencial impone una disminución controlada que asegura la existencia de un mínimo global y la ausencia de oscilaciones. En términos prácticos, esto significa que la función es *bien condicionada*, pues su crecimiento y curvatura son moderados, evitando problemas numéricos asociados a gradientes excesivamente grandes o abruptos.

En consecuencia, $f(x, y)$ constituye un caso representativo de superficie suave, acotada y radialmente decreciente, que permite un análisis teórico exhaustivo y una exploración experimental clara de los métodos de optimización aplicados.

2.10. Campo de gradientes

Para estudiar el comportamiento dinámico de los métodos de optimización sobre la función

$$f(x, y) = -200e^{-0,02\sqrt{x^2+y^2+1}},$$

es fundamental analizar su campo de gradientes, el cual determina las direcciones de ascenso y descenso más pronunciadas.

Gradiente analítico. Sea $r = \sqrt{x^2 + y^2 + 1}$. Aplicando la regla de la cadena, se obtiene:

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) = \left(4xe^{-0,02r} \frac{1}{r}, 4ye^{-0,02r} \frac{1}{r} \right) = \frac{4e^{-0,02r}}{r}(x, y).$$

El gradiente es proporcional al vector de posición (x, y) , lo cual significa que su dirección coincide con el radio vector. Por tanto, el campo de gradientes presenta *simetría radial perfecta*: todos los vectores son colineales con el eje que une el punto (x, y) con el origen.

Interpretación geométrica. Cada vector gradiente apunta hacia fuera del origen, pues f es una función decreciente con respecto a la distancia radial. En consecuencia:

- Los algoritmos de **ascenso del gradiente** divergerán del origen siguiendo trayectorias radiales.
- Los algoritmos de **descenso del gradiente**, por el contrario, seguirán trayectorias de convergencia directa hacia el origen.

Magnitud y comportamiento. La norma del gradiente está dada por:

$$\|\nabla f(x, y)\| = 4e^{-0,02r},$$

la cual decrece exponencialmente con r . Esto implica que el campo vectorial es fuerte en la región central (valores pequeños de r) y débil en la periferia. Geométricamente, el flujo se ralentiza a medida que los puntos se alejan del origen, generando zonas de

movimiento casi nulo donde el gradiente se aproxima a cero. Este patrón es característico de superficies con *curvatura decreciente* y explica por qué los métodos de optimización pueden experimentar convergencia lenta en regiones lejanas al mínimo.

Curvatura y estabilidad local. El Hessiano de f resulta ser definido positivo en un entorno del origen, lo cual confirma la convexidad local. Esta propiedad garantiza que las trayectorias del gradiente descendente no oscilen ni diverjan en las proximidades del mínimo global.

Implicaciones para algoritmos de optimización

El comportamiento teórico del gradiente y la forma global de la superficie permiten anticipar el rendimiento y las posibles dificultades de los algoritmos de optimización aplicados a $f(x, y)$.

Método del gradiente descendente. Dado que el gradiente apunta radialmente hacia el origen y su magnitud decrece suavemente, el método de gradiente descendente garantiza convergencia global para cualquier punto inicial (x_0, y_0) . No obstante, la velocidad de convergencia dependerá del tamaño del paso (α):

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

Si α es demasiado grande, el algoritmo puede sobrepasar el mínimo (debido a la pendiente suave en la periferia); si es demasiado pequeño, el avance se vuelve extremadamente lento. Por ello, resulta recomendable emplear técnicas adaptativas de selección de paso (por ejemplo, *backtracking line search* o *Armijo rule*).

Método de Newton. El método de Newton se beneficia de la suavidad y la positividad local del Hessiano:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [\nabla^2 f(\mathbf{x}_k)]^{-1} \nabla f(\mathbf{x}_k).$$

Cerca del origen, la curvatura es pronunciada y el Hessiano está bien condicionado, lo que garantiza convergencia cuadrática. Sin embargo, en regiones alejadas (donde la curvatura tiende a cero), el Hessiano se aproxima a la matriz nula y puede causar inestabilidad numérica. Por tanto, el método de Newton debe iniciarse en puntos relativamente cercanos al mínimo o combinarse con estrategias híbridas de primer orden.

Métodos estocásticos y variantes. Los algoritmos estocásticos o metaheurísticos (como *simulated annealing* o *particle swarm optimization*) no se benefician particularmente de la estructura de f , dado que no existen óptimos locales que justifiquen la aleatoriedad. Sin embargo, estos métodos pueden emplearse como contraste experimental para observar la rapidez con que identifican el mínimo en comparación con los métodos deterministas.

Conclusión metodológica. En resumen:

- Los algoritmos de primer orden son apropiados y garantizan convergencia global.
- Los métodos de segundo orden presentan mejor rendimiento local, pero requieren condiciones iniciales más favorables.

- La ausencia de irregularidades topológicas evita estancamientos o bifurcaciones.

Esta combinación convierte a $f(x, y)$ en una superficie ideal para estudiar propiedades de convergencia, estabilidad y sensibilidad numérica en entornos continuos.

Relación con el análisis experimental

El estudio teórico anterior proporciona un marco de referencia preciso para la fase experimental de la tarea. A partir de las propiedades demostradas, pueden establecerse las siguientes expectativas y objetivos de validación empírica:

Predicciones basadas en el análisis teórico.

- El punto $(0, 0)$ debe ser identificado consistentemente como el *único mínimo global*.
- Las trayectorias de convergencia de los algoritmos deben ser aproximadamente radiales, reflejando la simetría del campo de gradiente.
- Los valores de $f(x, y)$ deben aproximarse a cero para puntos alejados del origen, validando el comportamiento asintótico descrito.
- La magnitud del gradiente debe decrecer exponencialmente con la distancia, produciendo una convergencia lenta en la periferia.

Diseño de experimentos. Para corroborar las conclusiones teóricas, se recomienda:

1. Realizar simulaciones con distintos puntos iniciales distribuidos uniformemente en el dominio $[-100, 100]^2$.
2. Comparar el número de iteraciones necesarias para alcanzar un error predefinido con diferentes tamaños de paso.
3. Visualizar los campos de gradiente y las trayectorias de los algoritmos mediante gráficas de contorno y superficies 3D.
4. Registrar la evolución temporal del valor de la función y de la norma del gradiente.

Validación esperada. Los resultados experimentales deberán reproducir fielmente las propiedades teóricas: un mínimo global bien definido, convergencia radial, ausencia de oscilaciones y comportamiento exponencial del gradiente. Las discrepancias que puedan aparecer estarán asociadas a la discretización numérica o a la elección de parámetros de los métodos, lo que constituye un aspecto de análisis relevante en la comparación entre algoritmos.

Conclusión. La conexión entre el marco teórico y la experimentación computacional permite verificar empíricamente las propiedades de convexidad, suavidad y estabilidad derivadas del análisis matemático. De este modo, el problema sirve como un modelo de referencia ideal para evaluar la eficiencia, robustez y precisión de los algoritmos de optimización continua en espacios bidimensionales.

Selección y análisis de algoritmos de optimización

Con base en las propiedades teóricas del modelo

$$f(x, y) = -200e^{-0,02\sqrt{x^2+y^2+1}},$$

se seleccionan algoritmos adecuados considerando la suavidad, continuidad y ausencia de restricciones del problema. Dado que la función es diferenciable en todo \mathbb{R}^2 , con un único mínimo global y sin discontinuidades, los métodos de optimización continua sin restricciones son los más apropiados.

Entre las opciones posibles (métodos de barrera, penalización, máximo descenso, búsqueda direccional, Newton y quasi-Newton), se proponen dos enfoques representativos: 1. **Método del gradiente o máximo descenso**, como técnica de primer orden. 2. **Método de Newton o una de sus variantes quasi-Newton**, como técnica de segundo orden.

Ambos permitirán comparar la eficiencia y estabilidad de los métodos según el orden de derivadas empleadas y su sensibilidad a la curvatura del modelo.

1. Método del gradiente descendente (máximo descenso)

Descripción general. El método del gradiente descendente, también denominado *método del máximo descenso*, se basa en la iteración:

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

donde $\alpha_k > 0$ es el tamaño de paso o *learning rate*. En cada iteración, el algoritmo se mueve en la dirección opuesta al gradiente, buscando el punto donde la función disminuye más rápidamente.

Aplicación al modelo. Para la función $f(x, y)$, el gradiente analítico es:

$$\nabla f(x, y) = \frac{4e^{-0,02\sqrt{x^2+y^2+1}}}{\sqrt{x^2 + y^2 + 1}}(x, y),$$

por lo que la iteración queda explícitamente definida como:

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} - \alpha_k \frac{4e^{-0,02\sqrt{x_k^2+y_k^2+1}}}{\sqrt{x_k^2 + y_k^2 + 1}} \begin{pmatrix} x_k \\ y_k \end{pmatrix}.$$

Dada la simetría radial de f , las trayectorias de convergencia serán líneas rectas hacia el origen, y la velocidad de convergencia dependerá únicamente del valor de α_k .

Ventajas.

- Implementación sencilla y bajo costo computacional por iteración.
- Requiere únicamente el cálculo del gradiente, sin necesidad del Hessiano.
- Garantiza convergencia global para funciones suavemente decrecientes, como la analizada.

Desventajas.

- Convergencia lenta en regiones planas (donde $\|\nabla f\|$ es pequeño).
- Alta sensibilidad a la elección del tamaño de paso: un α_k inadecuado puede causar divergencia o estancamiento.
- No aprovecha información de curvatura, por lo que su desempeño empeora en problemas mal condicionados.

Recomendación práctica. Para mitigar sus limitaciones, se sugiere emplear una búsqueda de paso adaptativa (*line search*) o estrategias de descenso con paso variable, ajustando α_k en función de la norma del gradiente.

2. Método de Newton

Descripción general. El método de Newton es un procedimiento de segundo orden que incorpora información de la curvatura del modelo mediante el Hessiano de la función. Su fórmula iterativa es:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [\nabla^2 f(\mathbf{x}_k)]^{-1} \nabla f(\mathbf{x}_k),$$

donde $\nabla^2 f(\mathbf{x}_k)$ es la matriz Hessiana evaluada en el punto actual.

Aplicación al modelo. Para la función en estudio, el Hessiano presenta la forma general:

$$\nabla^2 f(x, y) = \frac{4e^{-0,02r}}{r^3} \begin{pmatrix} r^2 - 0,02x^2(r+1) & -0,02xy(r+1) \\ -0,02xy(r+1) & r^2 - 0,02y^2(r+1) \end{pmatrix},$$

donde $r = \sqrt{x^2 + y^2 + 1}$. Este tensor es simétrico y definido positivo en un entorno del origen, garantizando que el método converja cuadráticamente hacia el mínimo global.

Ventajas.

- Convergencia rápida (cuadrática) en la vecindad del óptimo.
- Aprovecha la información de curvatura, adaptando la dirección de búsqueda según la topología local.
- No requiere ajuste manual del tamaño de paso si el Hessiano es bien condicionado.

Desventajas.

- Costo computacional elevado: requiere el cálculo y la inversión del Hessiano en cada iteración.
- Posibles problemas numéricos si el Hessiano se aproxima a la singularidad (lo que ocurre lejos del mínimo, donde $\|\nabla^2 f\| \rightarrow 0$).
- No garantiza convergencia global; puede fallar si el punto inicial está demasiado alejado del óptimo.

Variante quasi-Newton (BFGS). Para evitar los inconvenientes anteriores, se puede utilizar una versión quasi-Newton, como el algoritmo BFGS, que construye una aproximación del Hessiano a partir de información del gradiente:

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

donde $s_k = x_{k+1} - x_k$ y $y_k = \nabla f_{k+1} - \nabla f_k$. Este enfoque reduce significativamente el costo computacional sin sacrificar la estabilidad, logrando convergencia superlineal.

Recomendación práctica. El método de Newton (o su versión BFGS) resulta ideal para refinar la solución una vez que el gradiente descendente ha aproximado la región del mínimo, combinando así convergencia global y eficiencia local.

3. Comparación general de los algoritmos propuestos

Criterio	Gradiente descendente	Newton / quasi-Newton
Tipo de método	Primer orden	Segundo orden
Requisitos de derivadas	Solo gradiente	Gradiente + Hessiano (o aproximación)
Costo por iteración	Bajo	Alto / Medio (BFGS)
Velocidad de convergencia	Lineal	Cuadrática (o superlineal)
Estabilidad numérica	Alta (si α adecuado)	Sensible al Hessiano
Convergencia global	Sí	No garantizada
Eficiencia cerca del óptimo	Moderada	Excelente

Conclusión comparativa. Ambos métodos se complementan: el gradiente descendente ofrece robustez global y simplicidad, mientras que Newton o quasi-Newton proporcionan refinamiento rápido y precisión local. Una estrategia combinada —utilizando primero el gradiente descendente y luego Newton o BFGS al acercarse al óptimo— representa la solución más eficiente para el problema planteado.