

# Tarea: Análisis y Aplicación de Algoritmos de Optimización

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

13 de noviembre de 2025

## Índice

<b>1. Marco Teórico [1]</b>	<b>2</b>
1.1. Notación y definición general . . . . .	2
1.2. Clases de problemas y propiedades básicas . . . . .	2
1.3. Algoritmos . . . . .	2
1.3.1. Métodos por direcciones (línea de búsqueda) . . . . .	2
1.3.2. Selección del tamaño de paso: condiciones de Armijo y de Wolfe . . . . .	3
1.3.3. Métodos de región de confianza (trust-region) . . . . .	4
1.3.4. Métodos para problemas con restricciones . . . . .	4
1.3.5. Criterios de parada típicos y métricas numéricas . . . . .	4
<b>2. Análisis de la función</b>	<b>5</b>
2.1. Puntos críticos . . . . .	5
2.2. Hessiano y clasificación del punto crítico . . . . .	5
2.3. Interpretación geométrica . . . . .	8
2.4. Cotas y comportamiento asintótico . . . . .	9
2.5. Implicaciones para algoritmos de optimización . . . . .	10
<b>3. Selección y análisis de algoritmos de optimización</b>	<b>11</b>
3.1. Método de máximo descenso . . . . .	11
3.2. Método de Newton . . . . .	12
3.3. Comparación . . . . .	13
<b>4. Discusión de resultados</b>	<b>13</b>
<b>5. Conclusiones</b>	<b>16</b>

# 1. Marco Teórico [1]

## 1.1. Notación y definición general

Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  la función objetivo y sea  $M \subset \mathbb{R}^n$  el conjunto factible de soluciones. Consideramos el problema general:

$$(P) \quad \min_{x \in M} f(x).$$

Con frecuencia  $M$  viene descrito mediante restricciones de igualdad y desigualdad:

$$M = \{x \in \mathbb{R}^n \mid h_i(x) = 0, i = 1, \dots, m; g_j(x) \leq 0, j = 1, \dots, s\}.$$

Denotamos por  $\nabla f(x)$  el gradiente y por  $\nabla^2 f(x)$  la matriz Hessiana cuando existan.

## 1.2. Clases de problemas y propiedades básicas

Los problemas suelen clasificarse en clases para su análisis y solución. Los tipos más comunes son los irrestrictos (con  $M = \mathbb{R}^n$ ), los continuos (con  $f \in C^1$ ) y los convexos (con  $f$  convexa y  $M$  convexo, donde cualquier mínimo local es también mínimo global). En general, para  $f \in C^2$ ,  $f$  convexa  $\iff \nabla^2 f(x) \succeq 0$  para todo  $x$ .

Puede garantizarse la existencia de soluciones en algunos casos específicos, como cuando  $M$  es compacto y  $f$  es continua (existe mínimo global), o en problemas no compactos cuando  $f(x) \rightarrow \infty$  al tiempo que  $\|x\| \rightarrow \infty$ . En el caso irrestricto, si  $f \in C^1$  y  $x^*$  es un mínimo local interior, entonces  $\nabla f(x^*) = 0$  (condición necesaria de primer orden). Otra condición viene dada por el **Teorema de KKT**: Sea  $x^*$  mínimo local factible bajo condiciones regulares (LICQ). Entonces existen multiplicadores  $\lambda \in \mathbb{R}^m$ ,  $\mu \in \mathbb{R}^s$  con  $\mu_j \geq 0$  tales que:

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^s \mu_j \nabla g_j(x^*) = 0,$$

sueto a  $h_i(x^*) = 0$ ,  $g_j(x^*) \leq 0$  y  $\mu_j g_j(x^*) = 0$ .

Puede hablarse a su vez de condiciones suficientes de optimalidad; o sea, condiciones que de cumplirse por un punto sería óptimo. En el caso irrestricto, si  $\nabla f(x^*) = 0$  y  $\nabla^2 f(x^*)$  es definida positiva, entonces  $x^*$  es mínimo local. En problemas convexos, puede afirmarse además que cualquier punto que satisfaga la condición de KKT será óptimo global.

## 1.3. Algoritmos

### 1.3.1. Métodos por direcciones (línea de búsqueda)

El esquema general está dado por:

$$x_{k+1} = x_k + \alpha_k d_k, \quad d_k \text{ dirección de descenso, } \alpha_k > 0 \text{ longitud de paso.}$$

Las direcciones tomadas varían en función del algoritmo a utilizar:

- *Gradiente descendente (steepest descent)*:  $d_k = -\nabla f(x_k)$ . Orden de convergencia lineal en general.
- *Newton*:  $d_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$ . Requiere Hessiana y su orden de convergencia es cuasi-cuadrático local si la Hessiana es definida positiva cerca del óptimo. Debido a la necesidad del cálculo de la inversa de la Hessiana, no es recomendable para matrices mal condicionadas.

- *Quasi-Newton (BFGS, DFP)*: Aproximan la inversa de la Hessiana con convergencia superlineal típica, lo que los hace buena práctica numérica para problemas grandes. La forma general del vector director es  $d_k = -H_k \nabla f(x_k)$  donde  $H_k$  es una matriz simétrica definida positiva que se actualiza en cada iteración y garantiza que se aproxime correctamente la curvatura local del gradiente entre  $x_k$  y  $x_{k+1}$ .

El método DFP (Davidon-Fletcher-Powell) actualiza directamente la aproximación de la inversa de la Hessiana  $H_k$  mediante:

$$H_{k+1} = H_k + \frac{s_k s_k^\top}{s_k^\top y_k} - \frac{H_k y_k y_k^\top H_k}{y_k^\top H_k y_k}, \quad s_k = x_{k+1} - x_k, \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

Esta actualización mantiene  $H_k$  simétrica y definida positiva siempre que  $s_k^\top y_k > 0$ .

El método BFGS (Broyden-Fletcher-Goldfarb-Shanno) es el más utilizado en la práctica por su mayor estabilidad numérica. Su actualización es:

$$H_{k+1} = \left( I - \frac{s_k y_k^\top}{y_k^\top s_k} \right) H_k \left( I - \frac{y_k s_k^\top}{y_k^\top s_k} \right) + \frac{s_k s_k^\top}{y_k^\top s_k}$$

También se puede expresar en términos de la aproximación de la Hessiana  $B_k = H_k^{-1}$ . En ese caso, la actualización equivalente, y nótese su similitud con la forma de  $H_k$  en DFP, es:

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

y la dirección de búsqueda resulta de resolver:

$$B_k d_k = -\nabla f(x_k), \quad \text{o equivalentemente} \quad d_k = -B_k^{-1} \nabla f(x_k) = -H_k \nabla f(x_k).$$

En estos métodos quasi-newton, si  $H_0$  es simétrica y definida positiva y  $s_k^\top y_k > 0$  para todo  $k$ , entonces  $H_k$  conserva esas propiedades. Además, el costo computacional por iteración es  $O(n^2)$ , mucho menor que el  $O(n^3)$  del método de Newton exacto.

### 1.3.2. Selección del tamaño de paso: condiciones de Armijo y de Wolfe

En los métodos de optimización de tipo búsqueda en línea, la iteración tiene la forma

$$x_{k+1} = x_k + \alpha_k d_k,$$

donde  $d_k$  es una dirección de descenso ( $\nabla f(x_k)^\top d_k < 0$ ) y  $\alpha_k > 0$  es el tamaño de paso. La elección adecuada de  $\alpha_k$  es fundamental para garantizar la *convergencia global* y la eficiencia del algoritmo.

La **condición de Armijo**, también llamada *criterio de descenso suficiente*, establece que el nuevo punto debe producir una reducción suficiente en la función objetivo. Se define un parámetro  $c_1 \in (0, 1)$  (típicamente  $c_1 \approx 10^{-4}$ ) y se busca un  $\alpha_k > 0$  tal que:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k \nabla f(x_k)^\top d_k.$$

Esta desigualdad garantiza que el decremento real en  $f$  es al menos una fracción  $c_1$  del decremento predicho por el modelo lineal  $f(x_k) + \alpha_k \nabla f(x_k)^\top d_k$ . Si la condición no se

cumple, el paso se reduce (por ejemplo multiplicándolo por un factor  $\beta \in (0, 1)$ ) hasta satisfacerla. Este procedimiento se conoce como *backtracking line search*.

Sin embargo, para obtener convergencia más robusta, especialmente en métodos de Newton y quasi-Newton, se utilizan las **condiciones de Wolfe**, que añaden un control de curvatura además del descenso suficiente. Se definen dos parámetros  $0 < c_1 < c_2 < 1$  (por ejemplo  $c_1 = 10^{-4}$ ,  $c_2 = 0,9$ ) tales que el valor de  $\alpha_k$  debe satisfacer simultáneamente:

$$(\text{Condición de Armijo}) \quad f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k \nabla f(x_k)^\top d_k, \quad (1)$$

$$(\text{Condición de curvatura}) \quad \nabla f(x_k + \alpha_k d_k)^\top d_k \geq c_2 \nabla f(x_k)^\top d_k. \quad (2)$$

La primera (1) es idéntica a la de Armijo; la segunda (2) asegura que el gradiente se haya reducido suficientemente en la dirección  $d_k$ , evitando pasos demasiado pequeños.

Las condiciones de Armijo garantizan *descenso global*, pero pueden generar pasos demasiado pequeños, mientras que las de Wolfe balancean descenso y curvatura, promoviendo convergencia superlineal cuando se combinan con BFGS o DFP. Existen algoritmos eficientes para encontrar  $\alpha_k$  que cumpla ambas condiciones (por búsqueda por intervalos o interpolación cuadrática/cúbica).

### 1.3.3. Métodos de región de confianza (trust-region)

En vez de buscar en una dirección y hacer line search, se resuelve en cada iteración un subproblema:

$$\min_{\|d\| \leq \Delta_k} m_k(d) = f(x_k) + \nabla f(x_k)^\top d + \frac{1}{2} d^\top B_k d,$$

donde  $B_k$  es una aproximación de la Hessiana. Este método es robusto en presencia de Hessianas indefinidas y mantiene un buen comportamiento global.

### 1.3.4. Métodos para problemas con restricciones

- *Métodos de penalidad*: Transforman restricciones en términos de penalización  $f(x) + \rho P(x)$  con  $\rho \rightarrow \infty$ . Son simples y fáciles de implementar, pero pueden inducir condicionamiento pobre y números grandes.
- *Métodos de barrera*: Aproximan restricciones mediante funciones barrera y resuelven una secuencia de problemas suaves con parámetro  $\mu \rightarrow 0^+$ . Son muy usados en problemas grandes y convexos.
- *Métodos de tipo SQP (Sequential Quadratic Programming)*: en cada iteración se resuelve un QP que aproxima el problema original. El orden de convergencia es cuadrático bajo condiciones regulares y son muy efectivos en restricciones suaves y de dimensión moderada.

### 1.3.5. Criterios de parada típicos y métricas numéricas

- $\|\nabla f(x_k)\| < \varepsilon_{\text{grad}}$ .
- $\|x_{k+1} - x_k\| < \varepsilon_{\text{step}}$ .
- Número máximo de iteraciones o evaluaciones de función.

## 2. Análisis de la función

Se considera la función

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

Definiendo  $s(x, y) := \sqrt{x^2 + y^2 + 1}$ , la función  $s$  es de clase  $C^2$  en  $\mathbb{R}^2$  (no hay singularidades porque  $x^2 + y^2 + 1 \geq 1$ , manteniendo siempre definidas  $s$  y sus derivadas). 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  $\mathbb{R}^2$ .

Si quisiéramos calcular  $\nabla f = (\partial_x f, \partial_y f)^\top$ . Podemos empezar observando que la derivada de  $s$  cumple al aplicar la regla de la cadena:

$$\frac{\partial s}{\partial x} = \frac{x}{s}, \quad \frac{\partial s}{\partial y} = \frac{y}{s},$$

Pasando a  $f$  y aplicando nuevamente la regla de la cadena tenemos:

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

y análogamente, dada la simetría de la función:

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

Por tanto, podemos escribir el gradiente de forma compacta como

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

### 2.1. Puntos críticos y estacionarios

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

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

con  $s = \sqrt{x^2 + y^2 + 1}$ . Dado que  $\forall (x, y) \in \mathbb{R}^2 : \frac{4e^{-0,02s}}{s} > 0$ , la única solución posible es:

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

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

### 2.2. Hessiano y clasificación del punto crítico

Si denotamos  $g(x, y) := 4 e^{-0,02s}$  y  $v(x, y) := \frac{1}{s} \begin{pmatrix} x \\ y \end{pmatrix}$ , tendremos que  $\nabla f = g v$ . Para la Hessiana  $H = \nabla^2 f$  usamos la regla del producto matricial:  $H = (\nabla g) v^\top + g \nabla v$ . Calcularemos los términos por separado.

## Derivadas de $g$

$$\frac{\partial g}{\partial x} = 4 \cdot e^{-0,02s} \cdot (-0,02) \cdot \frac{\partial s}{\partial x} = -0,08 e^{-0,02s} \frac{x}{s},$$

y de forma análoga

$$\frac{\partial g}{\partial y} = -0,08 e^{-0,02s} \frac{y}{s}.$$

Así,

$$\nabla g = -0,08 e^{-0,02s} \frac{1}{s} \begin{pmatrix} x \\ y \end{pmatrix}.$$

## Derivadas de $v$

Primero derivamos  $v_1 = x/s$  respecto a  $x$  e  $y$ :

$$\frac{\partial}{\partial x} \left( \frac{x}{s} \right) = \frac{1}{s} - \frac{x}{s^2} \frac{\partial s}{\partial x} = \frac{1}{s} - \frac{x}{s^2} \cdot \frac{x}{s} = \frac{1}{s} - \frac{x^2}{s^3} = \frac{s^2 - x^2}{s^3}.$$

Observando que  $s^2 = x^2 + y^2 + 1$ , se puede escribir  $s^2 - x^2 = y^2 + 1$ , por tanto

$$\frac{\partial}{\partial x} \left( \frac{x}{s} \right) = \frac{y^2 + 1}{s^3}.$$

La derivada de  $v_1 = x/s$  respecto a  $y$  es

$$\frac{\partial}{\partial y} \left( \frac{x}{s} \right) = -\frac{x}{s^2} \frac{\partial s}{\partial y} = -\frac{x}{s^2} \cdot \frac{y}{s} = -\frac{xy}{s^3}.$$

Análogamente para  $v_2 = y/s$ :

$$\frac{\partial}{\partial y} \left( \frac{y}{s} \right) = \frac{x^2 + 1}{s^3}, \quad \frac{\partial}{\partial x} \left( \frac{y}{s} \right) = -\frac{xy}{s^3}.$$

Por tanto la matriz  $\nabla v$  (Jacobiano de  $v$ ) es

$$\nabla v = \frac{1}{s^3} \begin{pmatrix} y^2 + 1 & -xy \\ -xy & x^2 + 1 \end{pmatrix}.$$

## Composición final

Recordando  $H = (\nabla g) v^\top + g \nabla v$ , sustituimos  $\nabla g$ ,  $\nabla v$ ,  $v$  y  $g$ .

Calculemos primero  $(\nabla g) v^\top$ . Dado que  $\nabla g = -0,08 e^{-0,02s} \frac{1}{s} \begin{pmatrix} x \\ y \end{pmatrix}$  y  $v^\top = \frac{1}{s}(x \ y)$ , tenemos

$$(\nabla g) v^\top = -0,08 e^{-0,02s} \frac{1}{s} \begin{pmatrix} x \\ y \end{pmatrix} \cdot \frac{1}{s}(x \ y) = -0,08 e^{-0,02s} \frac{1}{s^2} \begin{pmatrix} x^2 & xy \\ xy & y^2 \end{pmatrix}.$$

$$\text{Ahora } g \nabla v = 4e^{-0,02s} \cdot \frac{1}{s^3} \begin{pmatrix} y^2 + 1 & -xy \\ -xy & x^2 + 1 \end{pmatrix}.$$

Sumando ambos términos obtenemos  $H$ :

$$H = 4e^{-0,02s} \left[ -0,02 \frac{1}{s^2} \begin{pmatrix} x^2 & xy \\ xy & y^2 \end{pmatrix} + \frac{1}{s^3} \begin{pmatrix} y^2 + 1 & -xy \\ -xy & x^2 + 1 \end{pmatrix} \right],$$

Si factorizamos  $4e^{-0,02s}$  y combinamos términos, una forma compacta y útil es:

$$\boxed{\nabla^2 f(x, y) = 4e^{-0,02s} \left( \frac{1}{s} I_2 - \frac{0,02s + 1}{s^3} \begin{pmatrix} x^2 & xy \\ xy & y^2 \end{pmatrix} \right)}$$

donde  $I_2$  es la matriz identidad  $2 \times 2$ .

Si se desea ver las componentes explícitas:

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &= 4e^{-0,02s} \left( \frac{1}{s} - \frac{x^2(0,02s + 1)}{s^3} \right), \\ \frac{\partial^2 f}{\partial y^2} &= 4e^{-0,02s} \left( \frac{1}{s} - \frac{y^2(0,02s + 1)}{s^3} \right), \\ \frac{\partial^2 f}{\partial x \partial y} &= -4e^{-0,02s} \frac{xy(0,02s + 1)}{s^3}. \end{aligned}$$

Si analizamos la convexidad de la función comprobando que  $\nabla^2 f(x, y)$  es semidefinida positiva veremos que debe cumplirse en:

$$\frac{1}{s} I_2 - \frac{0,02s + 1}{s^3} \begin{pmatrix} x^2 & xy \\ xy & y^2 \end{pmatrix},$$

de donde puede comprobarse que esta propiedad solo se cumple para una bola específica centrada en  $(0, 0)$ , por lo que la convexidad será local y no puede extenderse a todo el dominio, lo cual tiene sentido viendo el comportamiento de la función cuando  $\|(x, y)\| \rightarrow \infty$ .

### Clasificación del punto

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

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

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

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

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

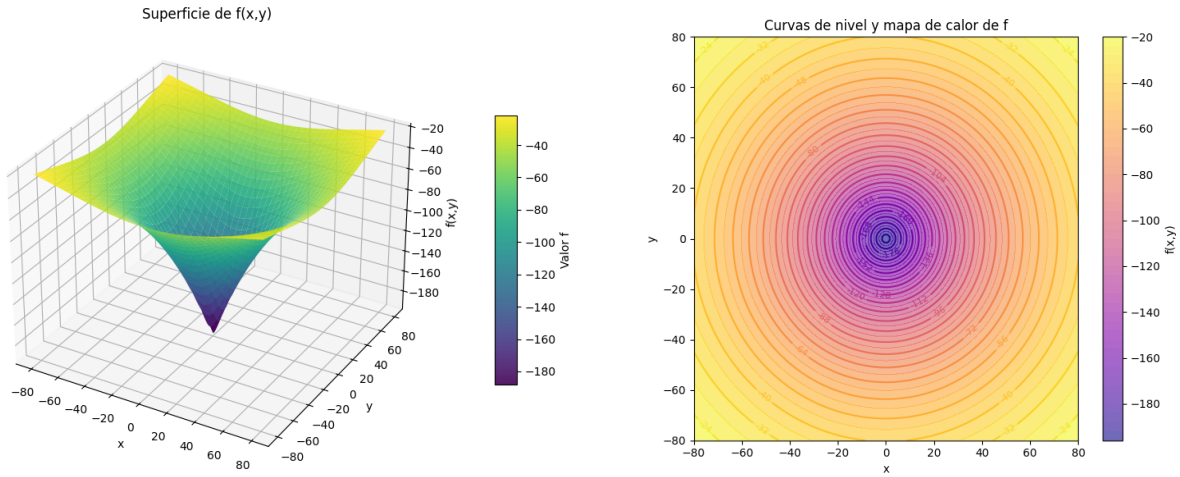
$$f(x, y) = -200e^{-0,02s} \rightarrow 0^-.$$

Esto implica:

- $\inf_{(x,y) \in \mathbb{R}^2} f(x,y) = f(0,0) \approx -196,0397$ , 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).

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 al  $(0,0)$ . No obstante, la no convexidad global 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).

## 2.3. Interpretación geométrica



(a) Gráfico de la función.

(b) Curvas de nivel de  $f$ .

Figura 1: Representaciones de la función para  $(x,y) \in [-80, 80]^2$ .

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. 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 (Figura 1b), lo cual evidencia la estructura perfectamente simétrica del modelo.



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 (Figura 1a).

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.

Geoméricamente, 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.

## 2.4. 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.

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

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

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  $f(0, 0)$ .

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.

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.

## 2.5. 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 ( $\alpha$ ):

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

Si  $\alpha$  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.

**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 aleatoriamente distribuidos que justifiquen la aleatoriedad, por lo que los métodos deterministas estudiados son la mejor opción para realizar el análisis.

A partir de las propiedades demostradas, pueden establecerse las siguientes expectativas y objetivos de validación empírica:

- 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.

### 3. 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:

- Método del gradiente o máximo descenso como técnica de primer orden.
- 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.

#### 3.1. Método de máximo descenso

El *método del máximo descenso*, también denominado método del gradiente descendente, 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.

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  $\alpha_k$ .

Este método tiene una implementación sencilla y conlleva un bajo costo computacional por iteración; requiere únicamente el cálculo del gradiente, sin necesidad del Hessiano, y garantiza convergencia global para funciones suavemente decrecientes como la analizada. Sin embargo, puede mostrar convergencia lenta en regiones planas (donde  $\|\nabla f\|$  es pequeño) y es altamente sensible a la elección del tamaño de paso: un  $\alpha_k$  inadecuado puede causar divergencia o estancamiento. Para mitigar sus limitaciones, se empleó una búsqueda de paso adaptativa con la regla de Armijo.

### 3.2. Método de Newton

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.

Para la función en estudio, el Hessiano presenta la forma general:

$$\nabla^2 f(x, y) = 4e^{-0,02s} \left( \frac{1}{s} I_2 - \frac{0,02s+1}{s^3} \begin{pmatrix} x^2 & xy \\ xy & y^2 \end{pmatrix} \right),$$

donde  $s = \sqrt{x^2 + y^2 + 1}$ . Esta expresión es simétrica y definida positiva en un entorno del origen, garantizando que el método converja cuadráticamente hacia el mínimo global.

Tiene una convergencia rápida en la vecindad del óptimo; aprovecha la información de curvatura, adaptando la dirección de búsqueda según la topología local, y no requiere ajuste manual del tamaño de paso si el Hessiano es bien condicionado. Su principal problemática radica en el costo computacional elevado que implica, dado que requiere el cálculo y la inversión del Hessiano en cada iteración, lo que también puede desencadenar errores 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, pues puede fallar si el punto inicial está demasiado alejado del óptimo.

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. 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.

### 3.3. Comparación

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 $\alpha$ adecuado)	Sensible al Hessiano
Convergencia global	Sí	No garantizada
Eficiencia cerca del óptimo	Moderada	Excelente

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.

## 4. Discusión de resultados

Con el propósito de evaluar comparativamente los métodos de optimización implementados sobre la función

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

se ejecutaron múltiples experimentos a partir de puntos iniciales distribuidos aleatoriamente en el plano, aplicando tres estrategias distintas:

- **Gradiente Descendente (GD)** con paso adaptativo;
- **BFGS**, método quasi-Newton de segundo orden;
- **Método Híbrido**, que inicia con GD y conmuta a BFGS al alcanzar un radio de confianza ( $R_{\text{switch}} = 10$  por defecto).

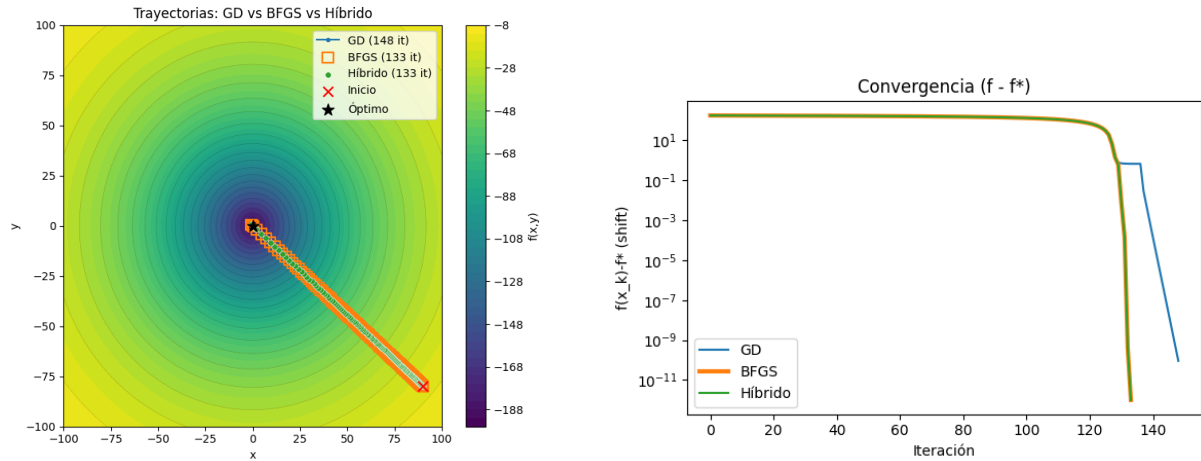
En todos los casos, los tres métodos convergieron al mismo punto crítico  $(x^*, y^*) \approx (0, 0)$ , alcanzando el valor

$$f(x^*, y^*) = -196,03973466135,$$

lo cual coincide con el mínimo global teórico identificado previamente en el análisis del gradiente y la matriz Hessiana. Las normas finales del gradiente fueron del orden de:

$$\|\nabla f\|_{\text{GD}} \approx 10^{-5}, \quad \|\nabla f\|_{\text{BFGS}} \approx 10^{-10}, \quad \|\nabla f\|_{\text{Híbrido}} \approx 10^{-13},$$

lo que evidencia una convergencia numérica completa (Figura 2).



(a) Camino seguido por los métodos en su búsqueda del mínimo. (b) Convergencia del valor obtenido en cada iteración.

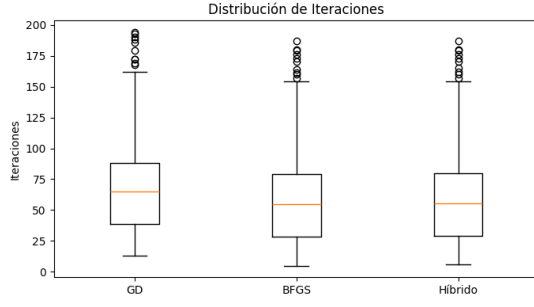
Figura 2: Análisis en el punto  $(90, -80)$ .

A partir de los datos experimentales, el número de iteraciones requerido por cada método varió según el punto inicial, pero siguió un patrón claro. Los resultados pueden revisarse en el Cuadro 1.

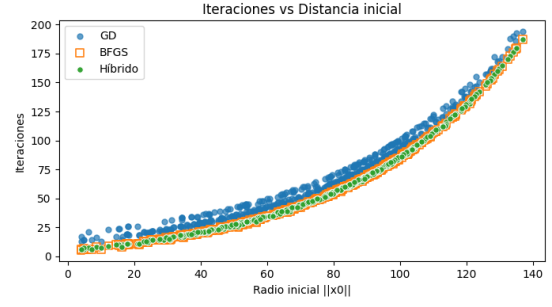
Método	Iteraciones promedio	Norma final del gradiente	Precisión
Gradiente Descendente (GD)	70–100	$10^{-5}$	Media
BFGS	50–90	$10^{-10}$	Alta
Método Híbrido	50–90	$10^{-13}$	Muy alta

Cuadro 1: Comparación de desempeño entre métodos de optimización.

El método de gradiente descendente logró converger en todos los casos, pero con un número mayor de iteraciones y una tasa de convergencia lineal. El método BFGS mostró una convergencia superlineal, reduciendo el número de iteraciones a casi la mitad y alcanzando precisiones varias órdenes de magnitud mayores. El método híbrido reprodujo de forma consistente los resultados de BFGS, conmutando hacia este en torno a la iteración 60–70, momento en el cual la magnitud del gradiente ya había disminuido lo suficiente como para garantizar un comportamiento local cuadrático (Figura 3).



(a) Cajas y bigotes.



(b) Iteraciones vs. Radio inicial

Figura 3: Gráficos producidos tras realizar 500 experimentos con puntos aleatorios en  $[-100, 100]^2$ .

Los experimentos se ejecutaron desde posiciones iniciales con magnitudes superiores a  $R_0 = 100$ , y en ningún caso se observó divergencia. Esto confirma que la función está bien condicionada en esta zona y es suavemente decreciente hacia el mínimo global. Sin embargo, la distancia inicial afectó directamente la cantidad de iteraciones necesarias, aumentándolas en cada método a medida que nos alejábamos del origen.

El análisis comparativo permite establecer las siguientes observaciones:

1. El **gradiente descendente** garantiza robustez global y estabilidad numérica, pero presenta una tasa de convergencia lineal, sensible a la escala del gradiente, y necesita un alto número de iteraciones para puntos alejados del mínimo.
2. El **BFGS** alcanza el equilibrio ideal entre precisión y velocidad, mostrando convergencia superlineal e independencia casi total del punto inicial.
3. El **método híbrido** combina la robustez del GD en las primeras fases con la eficiencia del BFGS en la etapa local. Esta combinación elimina el riesgo de oscilaciones iniciales y mantiene la tasa de convergencia alta en la vecindad del mínimo. Sin embargo, en los casos analizados, no representó una mejoría evidente sobre el BFGS.

La estrategia híbrida se confirma así como la más eficiente: requiere un número de iteraciones similar al BFGS, pero con menor sensibilidad a las condiciones iniciales y sin riesgo de inestabilidad.

Los resultados experimentales confirman que los tres métodos convergen al mismo mínimo global con gran precisión. BFGS e híbrido reducen el número de iteraciones en un 40-60 % respecto al gradiente descendente y en términos de desempeño, el híbrido logra una precisión de  $10^{-13}$  con un costo computacional equivalente al de BFGS puro. Se respalda la conclusión de que la combinación adaptativa de métodos de primer y segundo orden es la alternativa más equilibrada para problemas de optimización no lineal suaves, proporcionando tanto robustez global como eficiencia local.

## 5. Conclusiones

1. La función  $f(x, y) = -200e^{-0,02\sqrt{x^2+y^2+1}}$  es de clase  $C^2$ , acotada, radialmente simétrica y estrictamente creciente. Presenta un único punto crítico en  $(0, 0)$ , el cual es un mínimo global, y supremo 0 no alcanzado para  $\|(x, y)\| \rightarrow \infty$ .
2. El análisis del Hessiano evidencia convexidad local pero no global, por lo que la superficie es cuasi-convexa. Esta propiedad garantiza que los métodos de descenso convergen al mínimo desde cualquier punto inicial, aunque la velocidad de convergencia depende de la magnitud del gradiente.
3. El método de gradiente descendente ofrece robustez y simplicidad con convergencia global, aunque lenta en regiones planas. El método de Newton y sus variantes quasi-Newton (BFGS, DFP) proporcionan convergencia superlineal o cuadrática en la vecindad del mínimo, siendo más eficientes localmente pero más costosos computacionalmente.
4. La función analizada constituye un caso ideal para validar experimentalmente el comportamiento teórico de los algoritmos clásicos de optimización no lineal, ya que combina suavidad, simetría y un mínimo global único sin irregularidades locales.
5. Se confirmó empíricamente que la elección híbrida —gradiente descendente en las fases iniciales y quasi-Newton en el refinamiento local— representa la estrategia más eficiente para la función estudiada y para superficies de estructura similar, aunque en el intervalo analizado no mostró una reducción evidente del número de iteraciones del método quasi-Newton implementado (BFGS), por lo que para nuestro problema la utilización de este último es igual de válida.

## Referencias

- [1] G. Bouza Allende, *Notas de clase: Tems 1 y 2 — Optimización Matemática I*, February 2, 2021.
- [2] Documento de la orden del proyecto (archivo `orden.pdf`) — enunciado del problema y requisitos específicos.