

# Tarea: Análisis y Aplicación de Algoritmos de Optimización

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

12 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. Principales familias de algoritmos . . . . .	2
1.3.1. Métodos por direcciones (línea de búsqueda) . . . . .	2
1.4. Direcciones de búsqueda en los métodos quasi-Newton . . . . .	3
1.5. Selección del tamaño de paso: condiciones de Armijo y de Wolfe . . . . .	4
1.5.1. Métodos de región de confianza (trust-region) . . . . .	5
1.5.2. Métodos para problemas con restricciones . . . . .	6
1.6. Convergencia: órdenes y condiciones . . . . .	6
1.7. Criterios de parada y métricas numéricas . . . . .	6
1.8. Consideraciones numéricas y prácticas . . . . .	6
<b>2. Gradiente</b>	<b>7</b>
<b>3. Hessiana</b>	<b>7</b>
<b>4. Análisis de la función</b>	<b>9</b>
4.1. Regularidad . . . . .	9
4.2. Gradiente . . . . .	10
4.3. Puntos críticos . . . . .	10
4.4. Hessiano y clasificación del punto crítico . . . . .	10
4.5. Mínimo global y comportamiento en el infinito . . . . .	10
4.6. Convexidad . . . . .	11
4.7. Consecuencias prácticas para algoritmos . . . . .	11
4.8. Interpretación geométrica . . . . .	11
4.9. Cotas y comportamiento asintótico . . . . .	12
4.10. Campo de gradientes . . . . .	13
<b>5. Bibliografía</b>	<b>18</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, y que cumplen que 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, que existe mínimo global, o en problemas de mínimo no compactos cuando  $f(x) \rightarrow \infty$  junto a  $\|x\| \rightarrow \infty$ . 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. Principales familias de 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.

## 1.4. Direcciones de búsqueda en los métodos quasi-Newton

Los métodos *quasi-Newton* buscan aproximar la dirección de Newton sin calcular de forma explícita la matriz Hessiana ni su inversa en cada iteración. En el método de Newton puro,

$$d_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k).$$

Los métodos quasi-Newton reemplazan la inversa de la Hessiana por una matriz aproximada  $B_k$  o su inversa  $H_k \approx (\nabla^2 f(x_k))^{-1}$ , actualizada iterativamente a partir de información del gradiente.

**Forma general de la dirección** En todos los métodos quasi-Newton,

$$d_k = -H_k \nabla f(x_k),$$

donde  $H_k$  es una matriz simétrica definida positiva que se actualiza en cada iteración mediante la condición de *secante*:

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

La condición de secante garantiza que  $H_{k+1}$  aproxima correctamente la curvatura local del gradiente entre  $x_k$  y  $x_{k+1}$ .

### Actualización DFP (Davidon-Fletcher-Powell)

El método DFP actualiza directamente la aproximación de la *inversa de la Hessiana*  $H_k$  mediante:

$$H_{k+1} = H_k + \frac{s_k s_k^\top}{s_k^\top y_k} - \frac{H_k y_k y_k^\top H_k}{y_k^\top H_k y_k}.$$

Entonces la dirección de búsqueda se obtiene como:

$$d_k^{\text{DFP}} = -H_k \nabla f(x_k).$$

La actualización DFP mantiene  $H_k$  simétrica y definida positiva si  $s_k^\top y_k > 0$ .

### Actualización BFGS (Broyden-Fletcher-Goldfarb-Shanno)

El método BFGS es el más utilizado en la práctica por su mayor estabilidad numérica. Su actualización es:

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

La dirección de búsqueda se define igual:

$$d_k^{\text{BFGS}} = -H_k \nabla f(x_k).$$

También se puede expresar en términos de la aproximación de la Hessiana  $B_k = H_k^{-1}$ . En ese caso, la actualización equivalente 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:

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

## Resumen comparativo

Método	Actualiza	Dirección de búsqueda
DFP	$H_{k+1} = H_k + \frac{ss^\top}{s^\top y} - \frac{Hy y^\top H}{y^\top H y}$	$d_k = -H_k \nabla f(x_k)$
BFGS	$H_{k+1} = (I - \frac{sy^\top}{y^\top s}) H_k (I - \frac{ys^\top}{y^\top s}) + \frac{ss^\top}{y^\top s}$	$d_k = -H_k \nabla f(x_k)$

## Propiedades principales

- Si  $H_0$  es simétrica y definida positiva y  $s_k^\top y_k > 0$  para todo  $k$ , entonces  $H_k$  conserva esas propiedades.
- El método BFGS con búsqueda lineal que satisface las condiciones de Wolfe garantiza convergencia superlineal.
- El costo computacional por iteración es  $O(n^2)$ , mucho menor que el  $O(n^3)$  del método de Newton exacto.

## 1.5. 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.

### Condición de Armijo (descenso suficiente)

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

**Interpretación:** Si la condición no se cumple, el paso es demasiado grande; se reduce  $\alpha_k$  (por ejemplo multiplicándolo por un factor  $\beta \in (0, 1)$ ) hasta satisfacerla. Este procedimiento se conoce como *backtracking line search*.

## Condiciones de Wolfe

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  $c_1 = 10^{-4}$ ,  $c_2 = 0,9$ ). 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.

## Condiciones de Wolfe fuertes

Una variante frecuente son las **condiciones de Wolfe fuertes**, en las que la segunda desigualdad se reemplaza por:

$$\boxed{|\nabla f(x_k + \alpha_k d_k)^\top d_k| \leq c_2 |\nabla f(x_k)^\top d_k| .}$$

Esto impone una curvatura "moderada" en el punto nuevo y suele mejorar la estabilidad en métodos quasi-Newton.

## Propiedades y observaciones

- Las condiciones de Armijo garantizan *descenso global*, pero pueden generar pasos demasiado pequeños.
- Las condiciones de Wolfe (o Wolfe fuertes) 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).
- Bajo hipótesis estándar ( $f$  continuamente diferenciable y  $d_k$  de descenso), siempre existe un  $\alpha_k > 0$  que satisface las condiciones de Wolfe.

**Longitud de paso (line search):** Criterios Armijo y Wolfe para garantizar descenso suficiente y estabilidad.

### 1.5.1. 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. Ventajas: robustez en presencia de Hessianas indefinidas; buen comportamiento global.

### 1.5.2. Métodos para problemas con restricciones

Tres familias principales:

- **Métodos de penalidad:** transformar restricciones en términos de penalización  $f(x) + \rho P(x)$  con  $\rho \rightarrow \infty$ . Simples pero pueden inducir condicionamiento pobre y números grandes.
- **Métodos de barrera (interior-point):** aproximar restricciones mediante funciones-barrier y resolver secuencia de problemas suaves con parámetro  $\mu \rightarrow 0^+$ . 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. Orden de convergencia: cuadrático bajo condiciones regulares. Muy efectivos en restricciones suaves y de dimensión moderada.
- **Métodos de Lagrangianos aumentados (Augmented Lagrangian):** combinan penalidad y estimación de multiplicadores; robustos frente a problemas de condicionamiento.

### 1.6. Convergencia: órdenes y condiciones

- **Convergencia lineal:**  $\|x_{k+1} - x^*\| \leq c\|x_k - x^*\|$  con  $0 < c < 1$ .
- **Convergencia superlineal:** razón que tiende a 0.
- **Convergencia cuadrática:** rápida (p.ej. Newton cerca del óptimo cuando Hessiana es Lipschitz y definida positiva).
- La convergencia global suele requerir estrategias de globalización (line search o trust region) combinadas con condiciones de curvatura.

### 1.7. Criterios de parada y métricas numéricas

Criterios típicos:

- $\|\nabla f(x_k)\| < \varepsilon_{\text{grad}}$ .
- $\|x_{k+1} - x_k\| < \varepsilon_{\text{step}}$ .
- Violación de factibilidad (norma de restricciones violadas) por debajo de tolerancia.
- Número máximo de iteraciones o evaluaciones de función.

### 1.8. Consideraciones numéricas y prácticas

- **Escalado:** importante para estabilidad numérica. Reescalar variables y restricciones cuando las magnitudes difieren mucho.
- **Condicionamiento:** Hessianas mal condicionadas afectan métodos de Newton y Quasi-Newton; preconditionamiento o regularización pueden ser necesarios.
- **Cálculo de derivadas:** derivadas exactas (analíticas) cuando sean posibles; si se usan diferencias finitas, elegir paso apropiado y considerar el coste.
- **Tratamiento de restricciones activas:** detección de índices activos y manejo cuidadoso de la complementariedad para evitar oscilaciones numéricas.
- **Robustez:** combinar métodos (p.ej. pasar de método robusto de descenso a SQP o Newton local cerca del óptimo) suele ser la práctica recomendada.

## 2. Gradiente

Queremos calcular  $\nabla f = (\partial_x f, \partial_y f)^\top$ .

Empezamos observando la derivada de  $s$ :

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

pues  $s = \sqrt{x^2 + y^2 + 1}$  y aplicamos la regla de la cadena.

Ahora aplicamos la regla de la cadena a  $f(x, y) = -200e^{-0,02s}$ :

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

Calculamos el factor numérico  $(-200)(-0,02) = 4$ . Por tanto

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

De forma análoga,

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

Por lo tanto el gradiente queda

$$\boxed{\nabla f(x, y) = 4 e^{-0,02s} \frac{1}{s} \begin{pmatrix} x \\ y \end{pmatrix}}.$$

## 3. Hessiana

Denotaremos  $g(x, y) := 4 e^{-0,02s}$  y  $v(x, y) := \frac{1}{s} \begin{pmatrix} x \\ y \end{pmatrix}$ , de modo que  $\nabla f = g v$ . Para la Hessiana  $H = \nabla^2 f$  usamos la regla del producto matricial:

$$H = (\nabla g) v^\top + g \nabla v.$$

Calculamos los términos por separado.

### Derivadas de $g$

Recordando  $g = 4e^{-0,02s}$ ,

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

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 $v$

Recordemos  $v = (x/s, y/s)^\top$ . Calculemos las componentes de la matriz Jacobiana  $\nabla v$ , es decir  $(\partial_j v_i)_{i,j=1,2}$ .

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  $\nabla g$ ,  $v$  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  $H$ :

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

porque  $-0,08/4 = -0,02$ . 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  $2 \times 2$ .



## Componentes explícitos

Si se desea ver las componentes explícitas:

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

## Observaciones

- La Hessiana está definida en todo  $(x, y) \in \mathbb{R}^2$  ya que  $s = \sqrt{x^2 + y^2 + 1} \geq 1 > 0$ .
- Para análisis teórico y numérico, la forma factorizada

$$\nabla^2 f = 4e^{-0,02s} \left( \frac{1}{s} I - \frac{0,02s+1}{s^3} uu^\top \right), \quad u = \begin{pmatrix} x \\ y \end{pmatrix},$$

es conveniente: muestra explícitamente la estructura identidad menos un término de rango 1 escalado por  $uu^\top$ .

- Cerca del origen (pequeñas  $x, y$ ) se puede evaluar la Hessiana sustituyendo  $s \approx 1 + \frac{1}{2}(x^2 + y^2)$  si se desea una aproximación en serie.

## 4. Análisis de la función

Se considera la función

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

### 4.1. 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  $\mathbb{R}^2$ .

## 4.2. 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

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

## 4.3. 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.

## 4.4. Hessiano y clasificación del punto crítico

Evalutando 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**.

## 4.5. 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).

## 4.6. 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  $\mathbb{R}^2$ .

## 4.7. 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  $r > r_0$ .

## 4.8. 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.

## 4.9. 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.

## 4.10. 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 ( $\alpha$ ):

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

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 (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  $\alpha_k$ .

#### 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  $\alpha_k$  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  $\alpha_k$  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 $\alpha$ 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.

## 5. Bibliografía

### Referencias

- [1] G. Bouza Allende, *Notas de clase: Temas 1 y 2 — Optimización Matemática I*, February 2, 2021. (Documento proporcionado por el usuario).

- [2] Documento de la orden del proyecto (archivo `orden.pdf`) — enunciado del problema y requisitos específicos proporcionado por el usuario.