

# **FUNDAMENTOS DE OPTIMIZACIÓN II**

ALAN REYES-FIGUEROA  
MÉTODOS NUMÉRICOS II

(AULA 15) 30.AGOSTO.2022

# Conjuntos de Nivel

## Definición

Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  una función y  $c \in \mathbb{R}$ . El **conjunto de nivel**  $c$  de la función  $f$  es el conjunto de puntos

$$S_c = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = c\}.$$

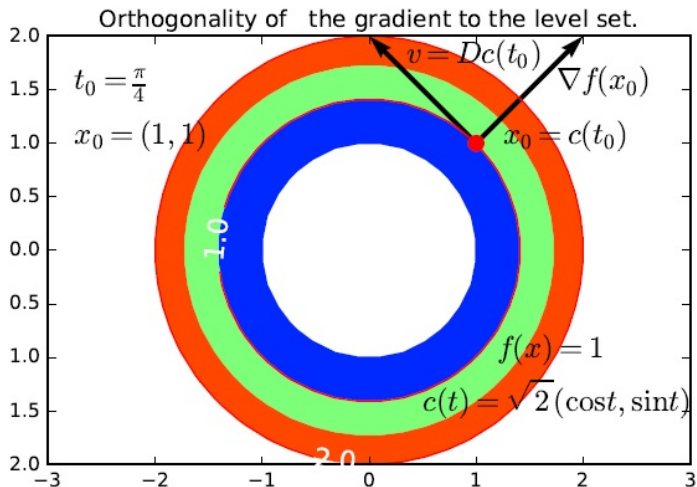
Típicamente,  $S_c$  o es vacío, o  $S_c$  induce una hipersuperficie de codimensión 1 (esto es de dimensión  $n - 1$  dentro de  $\mathbb{R}^n$ ), aunque en ocasiones,  $S_c$  se degenera en un objeto de menor dimensión.

Por ejemplo

- Si  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , entonces  $S_c$  es una curva.
- Si  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ , entonces  $S_c$  es una superficie 2-dimensional.
- En general, Si  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , entonces  $S_c$  es una hipersuperficie  $(n - 1)$ -dimensional.

**Ejemplo:**  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  dada por  $f(x, y) = \frac{1}{2}(x^2 + y^2)$ . El conjunto de nivel  $S_1 = \{(x, y) \in \mathbb{R}^2 : f(x, y) = 1\}$  corresponde al círculo  $x^2 + y^2 = 2$ . Una parametrización de  $S_c$  se obtiene al hacer  $\gamma(t) = (2 \cos t, 2 \sin t)$ ,  $t \in \mathbb{R}$ .

# Conjuntos de Nivel



# Conjuntos de Nivel

## Teorema

Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  diferenciable. Entonces, el vector gradiente  $\nabla_{\mathbf{x}} f(\mathbf{p})$  es ortogonal al vector tangente a cualquier curva suave que pasa por  $\mathbf{p}$ , contenida en el conjunto de nivel  $S_c$  de  $f$ , donde  $c = f(\mathbf{p})$ .

0.2cm

Prueba: Sea  $\gamma : (a, b) \subseteq \mathbb{R} \rightarrow \mathbb{R}^n$  una parametrización diferenciable de la curva suave, tal que  $\gamma(o) = \mathbf{p}$ , y sea  $\gamma'(o) = \mathbf{v}$  el vector tangente a esta curva en  $\mathbf{p}$ . Consideramos la función  $h : (a, b) \subseteq \mathbb{R} \rightarrow \mathbb{R}$ , dada por  $h = f \circ \gamma$ .

Como  $\gamma(t)$  está contenida dentro del conjunto de nivel  $S_c$ , entonces  $f(\gamma(t)) = c$ , para todo  $t \in (a, b)$ . Luego,  $h = f \circ \gamma$  es constante.

Aplicando la regla de la cadena a la función  $h(t) = (f \circ \gamma)(t)$ , resulta

$$0 = \frac{dh}{dt}(o) = Dh(o) = D(f \circ \gamma)(o) = Df(\gamma(o)) \cdot \gamma'(o) = \nabla_{\mathbf{x}} f(\mathbf{p}) \cdot \mathbf{v},$$

de modo que  $\nabla_{\mathbf{x}} f(\mathbf{p}) \perp \mathbf{v}$ , como se quería demostrar.  $\square$

# Gradiente

Recordemos que la derivada direccional de  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , en el punto  $\mathbf{p} \in \mathbb{R}^n$ , en la dirección del vector unitario  $\mathbf{u} \in \mathbb{R}^n$  es

$$D_{\mathbf{u}}f(\mathbf{p}) = Df(\mathbf{p}) \cdot \mathbf{u} = \nabla_{\mathbf{x}}f(\mathbf{p})^T \mathbf{u}.$$

## Propiedad

Suponga que  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es de clase  $C^1$  en un disco abierto que contiene al punto  $\mathbf{p}$ . Entonces, para cualquier vector unitario  $\mathbf{u} \in \mathbb{R}^n$ ,  $D_{\mathbf{u}}f(\mathbf{p})$  existe y

$$D_{\mathbf{u}}f(\mathbf{p}) = Df(\mathbf{p}) \cdot \mathbf{u} = \nabla_{\mathbf{x}}f(\mathbf{p})^T \mathbf{u}.$$

De la desigualdad de Cauchy-Schwarz, tenemos

$$\|D_{\mathbf{u}}f(\mathbf{p})\| = \|\nabla_{\mathbf{x}}f(\mathbf{p})^T \mathbf{u}\| \leq \|\nabla_{\mathbf{x}}f(\mathbf{p})\| \cdot \|\mathbf{u}\| = \|\nabla_{\mathbf{x}}f(\mathbf{p})\|.$$

Si  $\nabla_{\mathbf{x}}f(\mathbf{p}) \neq \mathbf{0}$ , tomando  $\mathbf{u} = \frac{\nabla_{\mathbf{x}}f(\mathbf{p})}{\|\nabla_{\mathbf{x}}f(\mathbf{p})\|}$ , obtenemos

$$D_{\mathbf{u}}f(\mathbf{p}) = \|\nabla_{\mathbf{x}}f(\mathbf{p})\|, \quad D_{-\mathbf{u}}f(\mathbf{p}) = -\|\nabla_{\mathbf{x}}f(\mathbf{p})\|.$$

## Teorema

Supongamos que  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es de clase  $C^1$  en una bola abierta que contiene al punto  $\mathbf{p}$ . Entonces,  $D_{\mathbf{u}}f(\mathbf{p})$  alcanza un valor máximo de  $\|\nabla_{\mathbf{x}}f(\mathbf{p})\|$  cuando  $\mathbf{u}$  es la dirección de  $\nabla_{\mathbf{x}}f(\mathbf{p})$  y alcanza un valor mínimo  $-\|\nabla_{\mathbf{x}}f(\mathbf{p})\|$  cuando  $\mathbf{u}$  es la dirección de  $-\nabla_{\mathbf{x}}f(\mathbf{p})$ .

Prueba: Como

$$\begin{aligned} D_{\mathbf{u}}f(\mathbf{p}) &= \nabla_{\mathbf{x}}f(\mathbf{p})^T \mathbf{u} = \|\nabla_{\mathbf{x}}f(\mathbf{p})\| \cdot \|\mathbf{u}\| \cos \angle(\nabla_{\mathbf{x}}f(\mathbf{p}), \mathbf{u}) \\ &= \|\nabla_{\mathbf{x}}f(\mathbf{p})\| \cos \angle(\nabla_{\mathbf{x}}f(\mathbf{p}), \mathbf{u}). \end{aligned}$$

El máximo y el mínimo de  $D_{\mathbf{u}}f(\mathbf{p})$  se alcanzan, respectivamente, cuando  $\cos \angle(\nabla_{\mathbf{x}}f(\mathbf{p}), \mathbf{u}) = 1$  y  $\cos \angle(\nabla_{\mathbf{x}}f(\mathbf{p}), \mathbf{u}) = -1$ .

Pero esto ocurre precisamente cuando  $\mathbf{u} = \frac{\nabla_{\mathbf{x}}f(\mathbf{p})}{\|\nabla_{\mathbf{x}}f(\mathbf{p})\|}$  y cuando  $\mathbf{u} = -\frac{\nabla_{\mathbf{x}}f(\mathbf{p})}{\|\nabla_{\mathbf{x}}f(\mathbf{p})\|}$ , respectivamente.

En particular, en tales casos,  $D_{\mathbf{u}}f(\mathbf{p}) = \|\nabla_{\mathbf{x}}f(\mathbf{p})\|$  y  $D_{\mathbf{u}}f(\mathbf{p}) = -\|\nabla_{\mathbf{x}}f(\mathbf{p})\|$ , resp.  $\square$

# Gradiente

## Propiedades del gradiente:

- El gradiente,  $\nabla_{\mathbf{x}}f(\mathbf{p})$ , de una función diferenciable, en el punto  $\mathbf{p}$ , es ortogonal al conjunto de nivel de la función  $f$  en ese punto.
- El vector de gradiente apunta en la dirección de máxima tasa de aumento de la función y el negativo del gradiente apunta en la dirección de la tasa máximo descenso de la función.
- La longitud del vector de gradiente nos dice la tasa de aumento en la dirección de aumento máximo y su negativo nos dice la tasa de disminución en la dirección de la disminución máxima.
- Similarmente, la magnitud de la derivada direccional  $|\nabla_{\mathbf{x}}f(\mathbf{p})^T \mathbf{u}|$  indica la tasa de aumento/reducción de  $f$  en la dirección de  $\mathbf{u}$ .

# Big O y Little o

## Definición

Decimos que  $f(\mathbf{x}) = O(g(\mathbf{x}))$ ,  $f$  es **O-grande** respecto de  $g$ , cuando  $\mathbf{x} \rightarrow \mathbf{a}$ , si existe una constante  $C$  tal que

$$|f(\mathbf{x})| \leq C|g(\mathbf{x})|, \quad \text{para todo } \mathbf{x} \in \mathbb{D}_\delta(\mathbf{a}).$$

Decimos que  $f(\mathbf{x}) = O(g(\mathbf{x}))$  cuando  $\mathbf{x} \rightarrow \infty$  si existen constantes positivas  $r$  y  $C$  tales que  $|f(\mathbf{x})| \leq C|g(\mathbf{x})|$ , para todo  $\mathbf{x}$  con  $\|\mathbf{x}\| \geq r$ .

Equivalentemente,  $f(\mathbf{x}) = O(g(\mathbf{x}))$  cuando  $\mathbf{x} \rightarrow \mathbf{a}$  si  $\lim_{\mathbf{x} \rightarrow \mathbf{a}} \left| \frac{f(\mathbf{x})}{g(\mathbf{x})} \right| = C$ , para alguna constante  $C \neq 0$ .

## Definición

Decimos que  $f(\mathbf{x}) = o(g(\mathbf{x}))$ ,  $f$  es **o-pequeña** respecto de  $g$ , cuando  $\mathbf{x} \rightarrow \mathbf{a}$ , si

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} \left| \frac{f(\mathbf{x})}{g(\mathbf{x})} \right| = 0.$$



# Big O y Little o

**Ejemplo:**  $f(x) = 5x^3 - 2x + 1$  es  $O(x^3)$ , cuando  $x \rightarrow \infty$ .

Basta ver que  $\lim_{x \rightarrow \infty} \left| \frac{5x^3 - 2x + 1}{x^3} \right| = 5$ .

Esto muestra que  $f(x) = O(x^3)$ .

**Ejemplo:**  $f(x) = 5x^3 - 2x + 1$  es  $o(x^4)$ , cuando  $x \rightarrow \infty$ .

Basta ver que  $\lim_{x \rightarrow \infty} \left| \frac{5x^3 - 2x + 1}{x^4} \right| = 0$ .

Esto muestra que  $f(x) = o(x^4)$ .

**Ejemplo:**  $f(x) = x - \sin x$  es  $o(x)$ , cuando  $x \rightarrow 0$ .

Basta ver que  $\lim_{x \rightarrow 0} \left| \frac{x - \sin x}{x} \right| = \lim_{x \rightarrow 0} \left| \frac{x - (x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots)}{x} \right| = 0$ .

# Big O y Little o

## Ejemplos de Big O:

- $x = O(x)$ , cuando  $x \rightarrow \infty$ ,
- $x = O(x^2)$ , cuando  $x \rightarrow \infty$ ,
- $ax^n = O(x^m)$ , para  $m \geq n$ , cuando  $x \rightarrow \infty$ ,
- $ax^n \neq O(x^m)$ , para  $m < n$ , cuando  $x \rightarrow \infty$ ,

## Ejemplos de little o:

- $x^2 = o(x)$ , cuando  $x \rightarrow 0$ ,
- $x \neq o(x^2)$ , cuando  $x \rightarrow 0$ ,
- $x - \sin x = o(x)$ , cuando  $x \rightarrow 0$ ,
- $x - \sin x = o(x^2)$ , cuando  $x \rightarrow 0$ ,

**Obs!** Importante!, la notaciones  $O$  y  $o$  dependen del punto donde se toma el límite.

Ejemplo:  $x^2 = o(x^3)$  cuando  $x \rightarrow \infty$ , pero  $x^2 \neq o(x^3)$  cuando  $x \rightarrow 0$ .

# Teorema de Taylor

## Propiedades:

- $f(x) = O(f(x))$ .
- Si  $f(x) = O(g(x))$ , entonces  $cf(x) = O(g(x))$ , para toda  $c \in \mathbb{R}$ ,  $c \neq 0$ .
- Si  $f_1(x), f_2(x)$  son  $O(g(x))$ , entonces  $f_1(x) + f_2(x) = O(g(x))$ .
- Si  $f(x) = o(g(x))$ , entonces  $f(x) = O(g(x))$ .
- Si  $f(x) = O(g(x))$ , entonces  $O(f(x)) + O(g(x)) = O(g(x))$ .
- Si  $f(x) = O(g(x))$ , entonces  $o(f(x)) + o(g(x)) = o(g(x))$ .
- Si  $f_1(x) = O(g(x))$ , pero  $f_2(x) = o(g(x))$ , entonces  $f_1(x) + f_2(x) = O(g(x))$ .
- Si  $f(x) = O(g(x))$  y  $g(x) = o(h(x))$ , entonces  $f(x) = o(h(x))$ .
- Para  $c \in \mathbb{R}$ ,  $c \neq 0$ ,  $cO(g(x)) = O(g(x))$  y  $co(g(x)) = o(g(x))$ .
- $O(f(x))O(g(x)) = O(f(x)g(x))$ .
- $o(f(x))O(g(x)) = o(f(x)g(x))$ .
- $o(f(x))o(g(x)) = o(f(x)g(x))$ .

# Teorema de Taylor

## Teorema (Fórmula de Taylor en $\mathbb{R}$ )

Suponga que  $f : \mathbb{R} \rightarrow \mathbb{R}$  es de clase  $C^{m+1}$  sobre  $\mathbb{R}$ , y sea  $x_0 \in \mathbb{R}$ . Denotemos,  $h = x - x_0$ . Entonces

$$f(x) = f(x_0) + hf'(x_0) + \frac{h^2}{2}f''(x_0) + \dots + \frac{h^m}{m!}f^{(m)}(x_0) + R_{m+1},$$

donde

$$R_{m+1} = \frac{h^{m+1}}{(m+1)!}f^{(m+1)}(x_0 + th), \quad \text{para algún } t \in (0, 1).$$

Usando la notación Big O, observe que  $R_{m+1} = O(h^{m+1})$ , si  $h \rightarrow 0$ . Así, la Fórmula de Taylor resulta

$$f(x) = f(x_0) + hf'(x_0) + \frac{h^2}{2}f''(x_0) + \dots + \frac{h^m}{m!}f^{(m)}(x_0) + O(h^{m+1}).$$

Usando la notación o pequeña, observe que  $R_{m+1} = o(h^m)$ , si  $h \rightarrow 0$ . Así, la fórmula es

$$f(x) = f(x_0) + hf'(x_0) + \frac{h^2}{2}f''(x_0) + \dots + \frac{h^m}{m!}f^{(m)}(x_0) + o(h^m).$$

# Teorema de Taylor

## Teorema (Fórmula de Taylor en $\mathbb{R}^n$ )

Suponga que  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es de clase  $C^{m+1}$  sobre  $\mathbb{R}^n$ , y sea  $\mathbf{x}_0 \in \mathbb{R}^n$ . Denotemos,  $\mathbf{h} = \mathbf{x} - \mathbf{x}_0$ . Entonces

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\mathbf{x}_0) h_j + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0) h_i h_j + \dots + \frac{1}{m!} \sum_{|I|=m} \frac{\partial^m f}{\partial \mathbf{x}_I}(\mathbf{x}_0) \mathbf{h}_I + R_{m+1},$$

donde

$$R_{m+1} = \frac{1}{(m+1)!} \sum_{|I|=m+1} \frac{\partial^{m+1} f}{\partial \mathbf{x}_I}(\mathbf{x}_0 + t\mathbf{h}) \mathbf{h}_I, \text{ para algún } t \in (0, 1).$$

Aquí,  $\mathbf{h} = (h_1, \dots, h_n)$ , y si  $I = (i_1, \dots, i_n) \in \mathbb{N}^n$  es tal que  $|I| = \sum_j i_j = m$ , entonces denotamos  $\mathbf{x}_I = (x_1^{i_1}, \dots, x_n^{i_n})$ ,  $\mathbf{h}_I = (h_1^{i_1}, \dots, h_n^{i_n})$  y  $\frac{\partial^m f}{\partial \mathbf{x}_I} = \frac{\partial^m f}{\partial x_1^{i_1} \dots \partial x_n^{i_n}}$ .

Al igual que en el caso unidimensional, podemos escribir  $R_{m+1} = O(\|\mathbf{h}\|^{m+1})$  y  $R_{m+1} = o(\|\mathbf{h}\|^m)$ , cuando  $\mathbf{h} \rightarrow \mathbf{0}$ .

# Teorema de Taylor

Dos casos particulares:

## Teorema (Aproximación de Taylor de primer orden)

Suponga que  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es de clase  $C^2$  sobre  $\mathbb{R}^n$ , y sea  $\mathbf{x}_0 \in \mathbb{R}$ ,  $\mathbf{h} = \mathbf{x} - \mathbf{x}_0$ . Entonces

$$f(\mathbf{x}) = f(\mathbf{x}_0) + Df(\mathbf{x}_0) \cdot \mathbf{h} + R_2 = f(\mathbf{x}_0) + Df(\mathbf{x}_0) \cdot \mathbf{h} + O(\|\mathbf{h}\|^2),$$

donde

$$R_2 = \frac{1}{2} \sum_{|I|=2} \frac{\partial^2 f}{\partial \mathbf{x}_I}(\mathbf{x}_0 + t\mathbf{h}) \mathbf{h}_I, \text{ para algún } t \in (0, 1).$$

## Teorema (Aproximación de Taylor de segundo orden)

Suponga que  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es de clase  $C^3$  sobre  $\mathbb{R}^n$ , y sea  $\mathbf{x}_0 \in \mathbb{R}$ ,  $\mathbf{h} = \mathbf{x} - \mathbf{x}_0$ . Entonces

$$f(\mathbf{x}) = f(\mathbf{x}_0) + Df(\mathbf{x}_0) \cdot \mathbf{h} + \frac{1}{2} \mathbf{h}^T D^2 f(\mathbf{x}_0) \mathbf{h} + R_3,$$

donde

$$R_3 = \frac{1}{3!} \sum_{|I|=3} \frac{\partial^3 f}{\partial \mathbf{x}_I}(\mathbf{x}_0 + t\mathbf{h}) \mathbf{h}_I, \text{ para algún } t \in (0, 1).$$

# Teorema de Taylor

En resumen, Si  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es de clase  $C^2$ , podemos escribir

$$f(\mathbf{x}) = f(\mathbf{x}_0) + Df(\mathbf{x}_0 + t\mathbf{h}), \quad t \in (0, 1).$$

$$f(\mathbf{x}) = f(\mathbf{x}_0) + Df(\mathbf{x}_0) \cdot \mathbf{h} + \frac{1}{2} \mathbf{h}^T D^2 f(\mathbf{x}_0 + t\mathbf{h}) \mathbf{h}, \quad t \in (0, 1).$$