

Tarea 1 - Optimización

Daniel Vallejo Aldana
daniel.vallejo@cimat.mx

11 de febrero de 2020

Problema 1. Let $f(x, y) = x^2 - y^2$ and $g(x, y) = 2xy$. Represent the level sets associated with $f(x, y) = 12$ and $g(x, y) = 16$ on the same figure using python. Indicate on the figure the points $\mathbf{x} = [x, y]^t$ for which $f(\mathbf{x}) = [12, 16]^t$

Solución Se presentan a continuación los resultados obtenidos al graficar los conjuntos de nivel.

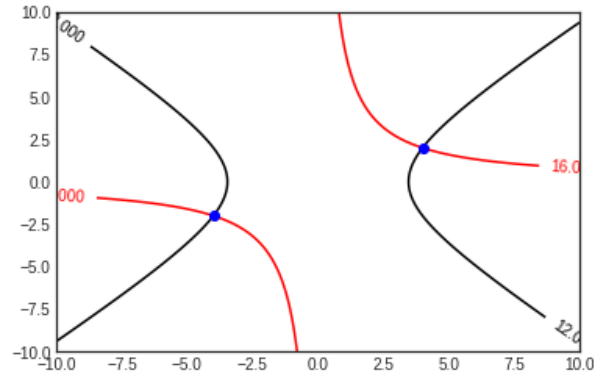


Figure 1: Resultado de la gráfica 1

Problema 2. Consider the function $f(\mathbf{x}) = (\mathbf{a}^t \mathbf{x})(\mathbf{b}^t \mathbf{x})$ where \mathbf{a}, \mathbf{b} and \mathbf{x} are n-dimentional vectors. Compute the gradient $\nabla f(\mathbf{x})$ and the Hessian $\nabla^2 f(\mathbf{x})$.

Solución Calcularemos primero el gradiente $\nabla f(\mathbf{x})$. Desarrollando $(\mathbf{a}^t \mathbf{x})(\mathbf{b}^t \mathbf{x})$ obtenemos lo siguiente.

$$\begin{aligned} f(\mathbf{x}) &= (\mathbf{a}^t \mathbf{x})(\mathbf{b}^t \mathbf{x}) \\ &= (a_1 x_1 + \dots + a_n x_n)(b_1 x_1 + \dots + b_n x_n) \\ &= \sum_{i=1}^n a_i x_i \sum_{i=1}^n b_i x_i \end{aligned}$$

Al calcular la derivada de $f(\mathbf{x})$ respecto a x_i se tiene lo siguiente utilizando la regla de la derivación para el producto.

$$\begin{aligned}\frac{\partial f(\mathbf{x})}{\partial x_j} &= \frac{\partial}{\partial x_j} \left(\sum_{i=1}^n a_i x_i \sum_{i=1}^n b_i x_i \right) \\ &= \left(\sum_{i=1}^n a_i x_i \right) b_j + \left(\sum_{i=1}^n b_i x_i \right) a_j \\ &= b_j(\mathbf{a}^t \mathbf{x}) + a_j(\mathbf{b}^t \mathbf{x})\end{aligned}$$

De lo anterior se tiene que el gradiente de $f(\mathbf{x})$ es $\nabla f(\mathbf{x}) = \mathbf{a}(\mathbf{b}^t \mathbf{x}) + \mathbf{b}(\mathbf{a}^t \mathbf{x})$

Calculemos ahora la matriz Hessiana de $f(\mathbf{x})$, para esto, consideremos la siguiente derivada.

$$\begin{aligned}\frac{\partial^2}{\partial x_j \partial x_i} f(\mathbf{x}) &= \frac{\partial}{\partial x_j} (b_i(\mathbf{a}^t \mathbf{x}) + a_i(\mathbf{b}^t \mathbf{x})) \\ &= b_i a_j + a_i b_j\end{aligned}$$

En particular si $i = j$ se tiene que $\frac{\partial^2}{\partial x_i^2} f(\mathbf{x}) = 2a_i b_i$ de modo que la matriz Hessiana para la función $f(\mathbf{x}) = (\mathbf{a}^t \mathbf{x})(\mathbf{b}^t \mathbf{x})$ es de la forma

$$\begin{bmatrix} 2a_1 b_1 & a_2 b_1 + b_2 a_1 & \dots & a_1 b_n + b_1 a_n \\ a_1 b_n + b_1 a_n & a_2 b_n + b_2 a_n & \dots & 2a_n b_n \end{bmatrix}$$

Problema 3. Let $f(x) = \frac{1}{1+e^{-x}}$ and $g(\mathbf{z}) = f(\mathbf{a}^t \mathbf{z} + b)$ with $\|\mathbf{a}\|_2 = 1$. Show that $D_{\mathbf{a}} g(\mathbf{z}) = g(\mathbf{z})(1 - g(\mathbf{z}))$

Demostración Utilizando el teorema de las diagonales se tiene que $D_{\mathbf{a}} g(\mathbf{z}) = Dg(\mathbf{z})\mathbf{a}$ para el vector unitario \mathbf{a} por lo que calcularemos $Dg(\mathbf{z})$.

Notemos entonces que

$$\begin{aligned}\frac{\partial}{\partial z_i} g(\mathbf{z}) &= \frac{\partial}{\partial z_i} f(\mathbf{a}^t \mathbf{z} + b) \\ &= \frac{\partial}{\partial z_i} \frac{1}{1 + e^{-a_1 z_1 - \dots - a_n z_n - b}} \\ &= \frac{a_i e^{-a_1 z_1 - \dots - a_n z_n - b}}{(1 + e^{-a_1 z_1 - \dots - a_n z_n - b})^2} \\ &= a_i g(\mathbf{z})(1 - g(\mathbf{z}))\end{aligned}$$

Por lo tanto

$$Dg(\mathbf{z}) = [a_1g(\mathbf{z})(1 - g(\mathbf{z})), \dots, a_ng(\mathbf{z})(1 - g(\mathbf{z}))]$$

De lo anterior se sigue que

$$\begin{aligned} D_{\mathbf{a}}g(\mathbf{z}) &= Dg(\mathbf{z})\mathbf{a} \\ &= a_1^2g(\mathbf{z})(1 - g(\mathbf{z})) + \dots + a_n^2g(\mathbf{z})(1 - g(\mathbf{z})) \\ &= g(\mathbf{z})(1 - g(\mathbf{z}))(a_1^2 + \dots + a_n^2) \end{aligned}$$

Sacando a $g(\mathbf{z})(1 - g(\mathbf{z}))$ como factor común

$$= g(\mathbf{z})(1 - g(\mathbf{z}))$$

ya que por hipótesis $\|\mathbf{a}\|_2 = 1$

De esta forma queda demostrado que $D_{\mathbf{a}}g(\mathbf{z}) = g(\mathbf{z})(1 - g(\mathbf{z}))$ ■

Problema 4. Compute the gradient of

$$f(\theta) = \frac{1}{2} \sum_{i=1}^n (g(\mathbf{x}_i) - g(\mathbf{A}\mathbf{x}_i + \mathbf{b}))^2$$

with respect to θ where $\theta = [a_{11}, a_{12}, a_{21}, a_{22}, b_1, b_2]^t$, $\mathbf{x}_i \in \mathbb{R}^2$ and \mathbf{A} and \mathbf{b} are defined as follows.

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\ \mathbf{b} &= [b_1 \quad b_2]^t \end{aligned}$$

and $g : \mathbb{R}^2 \rightarrow \mathbb{R} \in C^1$.

Solución Procederemos a calcular el gradiente de f utilizando la regla de la cadena para varias variables.

Para esto habrá que calcular la derivada parcial de la función f respecto a cada una de las variables de θ .

Notemos entonces lo siguiente

$$\begin{aligned} \frac{\partial}{\partial a_{11}} f(\theta) &= \frac{\partial}{\partial a_{11}} \left(\frac{1}{2} \sum_{i=1}^n (g(\mathbf{x}_i) - g(\mathbf{A}\mathbf{x}_i + \mathbf{b}))^2 \right) \\ &= \frac{1}{2} \sum_{i=1}^n \frac{\partial}{\partial a_{11}} ((g(\mathbf{x}_i) - g(\mathbf{A}\mathbf{x}_i + \mathbf{b}))^2) \end{aligned}$$

Por la linealidad de las derivadas

$$. = - \sum_{i=1}^n (g(\mathbf{x}_i) - g(\mathbf{Ax}_i + \mathbf{b})) \nabla^t g(\mathbf{Ax}_i + \mathbf{b}) \frac{\partial}{\partial a_{11}} (\mathbf{Ax}_i + \mathbf{b})$$

Utilizando la regla de la cadena para varias variables.

Vemos además que

$$\begin{aligned} \frac{\partial}{\partial a_{11}} (\mathbf{Ax}_i + \mathbf{b}) &= \frac{\partial}{\partial a_{11}} \begin{bmatrix} a_{11}x_{i1} + a_{12}x_{i2} + b_1 \\ a_{21}x_{i1} + a_{22}x_{i2} + b_2 \end{bmatrix} \\ &= \begin{bmatrix} x_{i1} \\ 0 \end{bmatrix} \end{aligned}$$

Por lo que

$$\frac{\partial}{\partial a_{11}} f(\theta) = - \sum_{i=1}^n (g(\mathbf{x}_i) - g(\mathbf{Ax}_i + \mathbf{b})) \nabla^t g(\mathbf{Ax}_i + \mathbf{b}) \begin{bmatrix} x_{i1} \\ 0 \end{bmatrix}$$

Análogamente se tiene lo siguiente

$$\begin{aligned} \frac{\partial}{\partial a_{12}} f(\theta) &= - \sum_{i=1}^n (g(\mathbf{x}_i) - g(\mathbf{Ax}_i + \mathbf{b})) \nabla^t g(\mathbf{Ax}_i + \mathbf{b}) \begin{bmatrix} x_{i2} \\ 0 \end{bmatrix} \\ \frac{\partial}{\partial a_{21}} f(\theta) &= - \sum_{i=1}^n (g(\mathbf{x}_i) - g(\mathbf{Ax}_i + \mathbf{b})) \nabla^t g(\mathbf{Ax}_i + \mathbf{b}) \begin{bmatrix} 0 \\ x_{i1} \end{bmatrix} \\ \frac{\partial}{\partial a_{22}} f(\theta) &= - \sum_{i=1}^n (g(\mathbf{x}_i) - g(\mathbf{Ax}_i + \mathbf{b})) \nabla^t g(\mathbf{Ax}_i + \mathbf{b}) \begin{bmatrix} 0 \\ x_{i2} \end{bmatrix} \\ \frac{\partial}{\partial b_1} f(\theta) &= - \sum_{i=1}^n (g(\mathbf{x}_i) - g(\mathbf{Ax}_i + \mathbf{b})) \nabla^t g(\mathbf{Ax}_i + \mathbf{b}) \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \frac{\partial}{\partial b_2} f(\theta) &= - \sum_{i=1}^n (g(\mathbf{x}_i) - g(\mathbf{Ax}_i + \mathbf{b})) \nabla^t g(\mathbf{Ax}_i + \mathbf{b}) \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned}$$

Por lo que el gradiente de $f(\theta)$ queda de la forma

$$\nabla f(\theta) = \left(- \sum_{i=1}^n (g(\mathbf{x}_i) - g(\mathbf{Ax}_i + \mathbf{b})) \nabla^t g(\mathbf{Ax}_i + \mathbf{b}) \right) \begin{bmatrix} x_{i1} & x_{i2} & 0 & 0 & 1 & 0 \\ 0 & 0 & x_{i1} & x_{i2} & 0 & 1 \end{bmatrix}$$

Problema 5. Show that $k(\mathbf{A}) \geq 1$ where $\|\mathbf{A}\| = \max_{\mathbf{x}} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|}$. Hint: Show that $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$

Proposición 1. Sea $\|\cdot\|$ una norma matricial y \mathbf{A} una matriz, entonces $\|\mathbf{Ax}\| \leq \|\mathbf{A}\| \|\mathbf{x}\|$ para \mathbf{x} un vector.

Demostración Sabemos que $\|\mathbf{A}\| = \max_{\mathbf{x}} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|}$ entonces para cualquier vector \mathbf{x} se tiene la siguiente desigualdad.

$$\|\mathbf{A}\| \geq \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|}$$

Por lo tanto $\|\mathbf{A}\| \|\mathbf{x}\| \geq \|\mathbf{Ax}\|$ que es lo que se buscaba probar ■

Proposición 2. Sea $\|\cdot\|$ una norma matricial y \mathbf{A} y \mathbf{B} matrices entonces $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$.

Demostración Por definición de norma sabemos que

$$\begin{aligned} \|\mathbf{AB}\| &= \max_{\mathbf{x}} \frac{\|\mathbf{ABx}\|}{\|\mathbf{x}\|} \\ &\leq \max_{\mathbf{x}} \frac{\|\mathbf{A}\| \|\mathbf{Bx}\|}{\|\mathbf{x}\|} \end{aligned}$$

Por lo demostrado en la proposición 1

$$\begin{aligned} &\leq \|\mathbf{A}\| \max_{\mathbf{x}} \frac{\|\mathbf{Bx}\|}{\|\mathbf{x}\|} \\ &= \|\mathbf{A}\| \|\mathbf{B}\| \end{aligned}$$

Por la definición de norma

De esta forma se tiene que $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$ ■.

Solución Utilizando la proposición 2 sabemos que $1 = \|\mathbf{AA}^{-1}\| \leq \|\mathbf{A}\| \|\mathbf{A}^{-1}\| = k(\mathbf{A})$ ■

Problema 6. Show that $x - \sin x = o(x^2)$, as $x \rightarrow 0$.

Demostración Definamos $f : \mathbb{R} \rightarrow \mathbb{R}$ como $f(x) = x - \sin x$ y $g : \mathbb{R} \rightarrow \mathbb{R}$ como $g(x) = x^2$. Notemos que f y g son clase C^∞ .

Por definición sabemos que $f(x) = o(g(x))$ cuando $x \rightarrow a$ si

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = 0$$

Por lo que para este caso se tiene lo siguiente

$$\begin{aligned}\lim_{x \rightarrow 0} \frac{f(x)}{g(x)} &= \lim_{x \rightarrow 0} \frac{x - \sin x}{x^2} \\ &= \lim_{x \rightarrow 0} \frac{1 - \cos x}{2x}\end{aligned}$$

Al aplicar la regla de L'Hospital

$$= \lim_{x \rightarrow 0} \frac{\sin x}{2}$$

Aplicando nuevamente la regla de L'Hospital

$$= 0$$

De lo anteriormente probado se concluye que $x - \sin x = o(x^2)$ cuando $x \rightarrow 0$ con lo que se completa la demostración ■.

Problema 7. Suppose that $f(\mathbf{x}) = o(g(\mathbf{x}))$. Show that for any given $\epsilon > 0$, there exists a $\delta > 0$ such that if $0 < \|\mathbf{x}\| < \delta$, then $|f(\mathbf{x})| < \epsilon|g(\mathbf{x})|$, i.e $f(\mathbf{x}) = O(g(\mathbf{x}))$ for $0 < \|\mathbf{x}\| < \delta$

Demostración Por hipótesis del problema sabemos que $f(\mathbf{x}) = o(g(\mathbf{x}))$, así mismo supondremos que $\mathbf{x} \rightarrow 0$. Como $f(\mathbf{x}) = o(g(\mathbf{x}))$ entonces se tiene que

$$\lim_{\mathbf{x} \rightarrow 0} \frac{f(\mathbf{x})}{g(\mathbf{x})} = 0$$

Y por la definición de límite se tiene que para cada $\epsilon > 0$ existe $\delta > 0$ tal que si $0 < \|\mathbf{x}\| < \delta$ entonces $|\frac{f(\mathbf{x})}{g(\mathbf{x})}| < \epsilon$. Por lo tanto para $\delta > 0$ se tiene que $|f(\mathbf{x})| < \epsilon|g(\mathbf{x})|$ que es lo que se buscaba demostrar ■.

Problema 8. Show that if functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfy $f(\mathbf{x}) = -g(\mathbf{x}) + o(g(\mathbf{x}))$ and $g(\mathbf{x}) > 0$ for all $\mathbf{x} \neq 0$, then for all $\mathbf{x} \neq 0$ sufficiently small, we have $f(\mathbf{x}) < 0$

Demostración Sea $f(\mathbf{x}) = -g(\mathbf{x}) + o(g(\mathbf{x}))$ entonces consideremos $h : \mathbb{R}^n \rightarrow \mathbb{R}$ una función tal que $f(\mathbf{x}) = -g(\mathbf{x}) + h(\mathbf{x})$. Notamos que $h(\mathbf{x}) = o(g(\mathbf{x}))$, lo que implica que

$$\lim_{\mathbf{x} \rightarrow 0} \frac{h(\mathbf{x})}{g(\mathbf{x})} = 0$$

De acuerdo al problema (7) para $\epsilon > 0$ existe $\delta > 0$ tal que si $0 < \|\mathbf{x}\| < \delta$ entonces $|h(\mathbf{x})| < \epsilon|g(\mathbf{x})|$, como $g(\mathbf{x}) > 0$ se tiene que $|h(\mathbf{x})| < \epsilon g(\mathbf{x})$. Si $h(\mathbf{x}) < 0$ para $\mathbf{x} \neq 0$ entonces se completa la demostración. Si $h(\mathbf{x}) > 0$ entonces se tiene lo siguiente

$$\begin{aligned}f(\mathbf{x}) &= -g(\mathbf{x}) + h(\mathbf{x}) \\ &< -g(\mathbf{x}) + \epsilon g(\mathbf{x})\end{aligned}$$

Al suponer $h(\mathbf{x}) > 0$

$$= g(\mathbf{x})(\epsilon - 1)$$

Como $g(\mathbf{x}) > 0$, al tomar $\epsilon < 1$ se tiene que $f(\mathbf{x}) < 0$ por lo que para $\mathbf{x} \neq 0$ suficientemente pequeño $f(\mathbf{x}) < 0$ ■