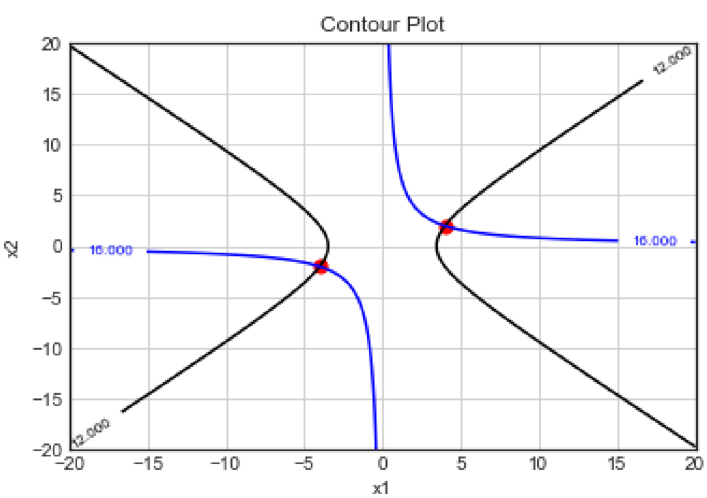
Tarea Uno

Optimización I

Esteban Reyes Saldaña

2 de febrero de 2021

Problema 1. Sea $f_1(x_1, x_2) = x_1^2 - x_2^2$, $f_2(x_1, x_2) = 2x_1x_2$. Represente los conjuntos de nivel asociados con $f_1(x_1, x_2) = 12$ y $f_2(x_1, x_2) = 16$ sobre la misma figura usando Python. Indique sobre la figura los puntos $\mathbf{x} = [x_1, x_2]^T$ para los cuales $f(\mathbf{x}) = [f_1(x_1, x_2), f_2(x_1, x_2)]^T = [12, 16]^T$.
Solución la gráfica obtenida fue



(Parte analítica). Dadas las ecuaciones

$$\begin{cases} x_1^2 - x_2^2 & = & 12 \\ 2x_1x_2 & = & 16. \end{cases} \tag{1}$$

Como $x_1 x_2 = 8$, entonces tanto x_1 como x_2 son distintos de cero. Luego, de la segunda ecuación tenemos que

$$x_2 = \frac{8}{x_1}. \tag{2}$$

Así que la primera ecuación se puede reescribir como

$$\begin{aligned} x_1^2 - \left(\frac{8}{x_1}\right)^2 &= 12 \\ x_1^2 - \frac{64}{x_1^2} &= 12. \end{aligned}$$

Sea $u = x_1^2$, entonces

$$\begin{aligned} u - \frac{64}{u} &= 12 \\ \frac{u^2 - 64}{u} &= 12 \\ u^2 - 64 &= 12u \\ u^2 - 12u - 64 &= 0. \end{aligned}$$

Resolviendo esta ecuación cuadrática obtenemos

$$u = \frac{12 \pm \sqrt{144 + 256}}{2} = \frac{12 \pm 20}{2}$$

Así que las soluciones para u son

$$u = 16 \text{ o } u = -4$$

Como $x_1^2 = u \in \mathbb{R}^+$ entonces obtenemos que $x_1^2 = 16$, de donde

$$x_1 = -4 \text{ o } x_1 = 4.$$

Finalmente, de la ecuación (2) obtenemos que las soluciones a (1) son

$$(-4, -2) \text{ o } (4, 2).$$

Problema 2. Considere la función $f(x) = (a^T x)(b^T x)$, donde a, b y x son vectores n- dimensionales. Calcule el gradiente $\nabla f(x)$ y el Hessiano $\nabla^2 f(x)$.
Solución. Sabemos que

$$Df(x) = \nabla f(x)^T.$$

y que para dos vectores $a, b \in \mathbb{R}^n$ se tiene que

$$< a, b >= ab^T = ba^T \tag{3}$$

Así que usando la regla del producto para derivadas obtenemos

$$\begin{aligned} D_x(a^T x)(b^T x) &= (a^T x)D_x(b^T x) + (b^T x)D_x(a^T x) \\ &= (a^T x)b^T + (b^T x)a^T \\ &= a^T x b^T + b^T x a^T \\ &= \underbrace{x^T ab^T + x^T ba^T}_{(3)} \\ &= x^T(ab^T + ba^T) \\ &= \underbrace{(ab^T + ba^T)x^T}_{\text{conmutatividad escalar-vector}} \end{aligned}$$

Por lo que

$$\begin{aligned} \nabla f(x) &= D_x^T f(x) \\ &= \left[(ab^T + ba^T)x^T\right]^T \\ &= (ab^T + ba^T)x. \end{aligned}$$

Ahora, dado que el gradiente de f resultó ser el producto de un escalar por el vector x tenemos que

$$\begin{aligned} \nabla^2 f(x) &= \nabla(\nabla f(x)) \\ &= \nabla((ab^T + ba^T)x) \\ &= D_x((ab^T + ba^T)x)^T \\ &= (ab^T + ba^T)D_x(x)^T \\ &= (ab^T + ba^T). \end{aligned}$$

Problema 3. Sea $f(x) = \dfrac{1}{1 + e^{-x}}$ y $g(z) = f(a^T z + b)$ con $||a||_2 = 1$. Demuestre que $D_a g(z) = g(z)(1 - g(z))$.

Demostración. Por un lado, tenemos que $g : \mathbb{R}^n \mapsto \mathbb{R}$ y en clase se vió que

$$D_a g(z) = \nabla g(x)^T a \tag{4}$$

Por otro lado, sea $h(z) = a^T z + b$, entonces

$$g(z) = f(h(z)). \tag{5}$$

Observemos que

$$\begin{aligned}
f'(x) &= -\frac{1}{(1+e^{-x})^2}(-e^{-x}) \\
&= \frac{1}{(1+e^{-x})^2}(1+e^{-x}-1) \\
&= \frac{1+e^{-x}}{(1+e^{-x})^2} - \frac{1}{(1+e^{-x})^2} \\
&= \frac{1}{1+e^{-x}} - \left(\frac{1}{1+e^{-x}}\right)^2 \\
&= f(x) - f^2(x) \\
&= f(x)(1-f(x)).
\end{aligned}$$

y

$$\begin{aligned}
D_z(h(z)) &= D_z(a^T z + b) \\
&= D_z(a^T z) + D_z(b) \\
&= a^T.
\end{aligned}$$

Ahora, de (4) y por regla de la cadena tenemos que

$$\begin{aligned}
D_a g(z) &= Df(h(z))D(h(z))a \\
&= f(h(z))(1-f(h(z)))a^T a.
\end{aligned}$$

Por la forma en la que se definió $h(z)$ en (5) y dado que $a^T a = ||a||_2 = 1$, se concluye que

$$\begin{aligned}
D_a g(z) &= f(h(z))(1-g(h(z)))a^T a \\
&= g(z)(1-g(z)).
\end{aligned}$$

□

Problema 4. Calcule el gradiente de

$$f(\theta) = \frac{1}{2}\sum_{t=1}^{n}[g(x_i) - g(Ax_i+b)]^2$$

con respecto a θ, donde $\theta = [a_{11}, a_{12}, a_{21}, a_{22}, b_1, b_2]^T$, $x_i \in \mathbb{R}^2$, $A \in \mathbb{R}^{2\times 2}$, $b \in \mathbb{R}^2$ están definidas como sigue

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

y $g : \mathbb{R}^2 \to \mathbb{R} \in \mathcal{C}^1$.
Solución. Usando la relación entre gradiente-derivada y por regla de la cadena tenemos que

$$\begin{aligned} \nabla f(\theta)^T &= D_\theta \left(\frac{1}{2} \sum_{t=1}^{n} [g(x_i) - g(Ax_i + b)]^2 \right) \\ &= \frac{1}{2} 2 \sum_{t=1}^{n} [g(x_i) - g(Ax_i + b)] D[g(x_i) - g(Ax_i + b)] \\ &= \frac{1}{2} 2 \sum_{t=1}^{n} [g(x_i) - g(Ax_i + b)][Dg(x_i) - Dg(Ax_i + b)] \\ &= \sum_{t=1}^{n} [g(x_i) - g(Ax_i + b)][Dg(x_i) - Dg(Ax_i + b)] \\ &= \sum_{t=1}^{n} [g(x_i) - g(Ax_i + b)][0 - Dg(Ax_i + b) D(Ax_i + b)] \\ &= -\sum_{t=1}^{n} [g(x_i) - g(Ax_i + b)] Dg(Ax_i + b) D(Ax_i + b) \\ &= -\sum_{t=1}^{n} [g(x_i) - g(Ax_i + b)] Dg(Ax_i + b)(D(Ax_i) + D(b)). \end{aligned}$$

Problema 5. Demuestre que $\kappa(A) \geq 1$ donde $||A|| = \text{máx}_x \dfrac{||Ax||}{||x||}$ (Sugerencia: demuestre que $||AB|| \leq ||A||||B||$).

Demostración. Primero se verá que $||AB|| \leq ||A||||B||$. En efecto, para $x \neq 0$ tenemos que

$$||A|| = \text{máx}_x \frac{||Ax||}{||x||}$$

y por la definición de máximo sobre x no nulo,

$$\begin{aligned} \frac{||Ax||}{||x||} &\leq ||A|| \\ ||Ax|| &\leq ||A||||x||. \end{aligned}$$

Ahora, por la propiedad anterior se tiene que

$$\begin{aligned} ||AB|| &= \max_x \frac{||ABx||}{||x||} \\ &\leq \max_x \frac{||A||||Bx||}{||x||} \\ &\leq \max_x \frac{||A||||B||||x||}{||x||} \\ &= \max_x ||A||||B|| \\ &= ||A||||B||. \end{aligned}$$

□

Problema 6. Demuestre que $x - \sin(x) = o(x^2)$ cuando $x \to 0$.

Demostración. Hay que demostrar que

$$\lim_{x \to 0} \frac{x - sen(x)}{x^2} = 0.$$

En efecto, procedemos aplicando la regla de L'Hospital. Si

$$\begin{aligned} f(x) &= x - sen(x). \\ g(x) &= x^2 \end{aligned}$$

tenemos que $f(x), g(x) \in \mathcal{C}^\infty$ sobre $\mathbb{R}$ con $f(0) = 0$ y $g(0) = 0$. Además,

$$\begin{aligned} f'(x) &= 1 + cos(x) \\ g'(x) &= 2x \end{aligned}$$

y

$$\begin{aligned} f''(x) &= -sen(x). \\ g'(x) &= 2 \end{aligned}$$

donde $f''(0) = 0$ y $g'''(0) = 2 \neq 0$, así que, por la regla de $L'Hospital$, tenemos que

$$\begin{aligned} \lim_{x\to 0} \frac{x - sen(x)}{x^2} &= \lim_{x\to 0} \frac{1 + cos(x)}{2x} \\ &= \lim_{x\to 0} \frac{-sen(x)}{2} \\ &= 0. \end{aligned}$$

De lo anterior se concluye que $x - sen(x) \in 0(x^2)$. □

Problema 7. Suponga que $f(x) = o(g(x))$. Demuestre que para todo $\epsilon > 0$, existe $\delta > 0$ tal que si $0 < ||x|| < \delta$, entonces $|f(x)| < \epsilon |g(x)|$, i.e, $f(x) = O(g(x))$ para $0 < ||x|| < \delta$.

Demostración. Sin pérdida de generalidad, suponga que $x \to 0$ (de otro modo se considera una translación a cero). Supongamos que $f(x) = o(g(x))$ cuando $x \to 0$. Entonces

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

Por definición se tiene que, para todo $\epsilon > 0$ existe $\delta > 0$ para el cual si

$$0 < ||x|| < \delta \text{ entonces } \left| \frac{f(x)}{g(x)} \right| < \epsilon.$$

Además sabemos que como este límite existe, $g(x) \neq 0$ para $0 < ||x|| < \delta$. Entonces

$$|f(x)| < \epsilon |g(x)|.$$

Es decir, $f(x) = O(g(x))$ para $0 < ||x|| < \delta$. □

Problema 8. Demuestre que si las funciones $f : \mathbb{R}^n \to \mathbb{R}$ y $g : \mathbb{R}^n \to \mathbb{R}$ satisface $f(x) = -g(x) + o(g(x))$ y $g(x) > 0$ para todo $x \neq 0$, entonces para todo $x \neq 0$ suficientemente pequeño, se tiene que $f(x) < 0$.

Demostración. Notemos que

$$f(x) = -g(x) + o(g(x))$$

es equivalente a

$$f(x) + g(x) = o(g(x)).$$

Por el problema anterior, sabemos que para todo $\epsilon > 0$, existe $\delta > 0$ tal que si $0 < ||x|| < \delta$ entonces

$$|f(x) + g(x)| < \epsilon |g(x)|$$

Luego, como $g(x) > 0$ para todo $x \neq 0$, $|g(x)| = g(x)$, entonces

$$-\epsilon g(x) \quad < f(x) + g(x) < \quad \epsilon g(x) \tag{6}$$

de donde

$$-(\epsilon + 1)g(x) = -\epsilon g(x) - g(x) < f(x) < \epsilon g(x) - g(x) = (\epsilon - 1)g(x)$$

Así que se elige $0 < \hat{\epsilon} < 1$ entonces $\hat{\epsilon} - 1 < 0$ por lo que existe $\hat{\delta} > 0$ tal que si $||x|| < \hat{\delta}$ entonces

$$f(x) < (\hat{\epsilon} - 1)g(x) < 0.$$

Es decir, $f(x) < 0$ para un x lo suficientemente pequeño $\square$