Optimization
Homework 1

Erika Rivadeneira
erika.rivadeneira@cimat.mx

February 2, 2021

1. *Let* $f_1(x_1, x_2) = x_1^2 - x_2^2, f_2(x_1, x_2) = 2x_1x_2$. *Represent the level sets associated with* $f_1(x_1, x_2) = 12$ *and* $f_2(x_1, x_2) = 16$ *on the same figure using Python. Indicate on the figure, the points* $\boldsymbol{x} = [x_1, x_2]^T$ *for which* $f(\boldsymbol{x}) = [f_1(x_1, x_2), f_2(x_1, x_2)]^T = [12, 16]^T$
Solution
Let's solve the equation system in order to find the intersection points. We have that

$$x_1^2 - x_2^2 = 12 \tag{1}$$
$$2x_1x_2 = 16 \tag{2}$$

From equation (2) we get that

$$x_2 = \frac{8}{x_1}$$

substituting $x_2 = \frac{8}{x_1}$ in equation (1)

$$x_1^2 - \frac{64}{x_1^2} = 12$$
$$x_1^2(x_1^2 - 12) = 64$$

Then, the solution of the system is $(-4, -2)$ and $(4, 2)$
Finally, the corresponding plot of the level sets is

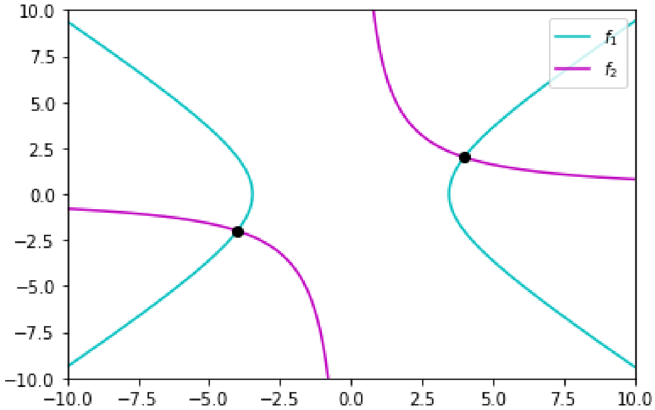

Figura 1: Level sets associated with $f_1(x_1, x_2) = 12$ and $f_2(x_1, x_2) = 16$

2. *Consider the function* $f(\boldsymbol{x}) = \left(\boldsymbol{a}^T\boldsymbol{x}\right)\left(\boldsymbol{b}^T\boldsymbol{x}\right)$, *where* $\boldsymbol{a}, \boldsymbol{b}$, *and* $\boldsymbol{x}$ *are* n− *dimensional vectors. Compute the gradient* $\nabla f(\boldsymbol{x})$ *and the Hessian* $\nabla^2 f(\boldsymbol{x})$.
Solution
Let's recall that the gradient of $f(x)$ is defined as

$$\nabla f(\boldsymbol{x}) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\right]^T$$
$$Df(\boldsymbol{x}) = \nabla f(\boldsymbol{x})^T$$

Then, we have that the gradien of $f(x)$ is

$$\begin{aligned}\frac{\partial f}{\partial x_k} &= a_k \sum_{i=1}^{n} b_i x_i + b_k \sum_{i=1}^{n} a_i x_i \\ &= \sum_{j=1}^{n} (a_k b_j + b_k a_j)\, x_j\end{aligned}$$

Now, the Hessian $\nabla^2 f(\boldsymbol{x})$ is given by

$$\nabla^2 f(\boldsymbol{x}) = \frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{bmatrix} 2a_1b_1 & a_1b_2 + a_2b_1 & \dots & a_1b_n + a_nb_1 \\ a_2b_1 + a_1b_2 & 2a_2b_2 & \dots & a_2b_n + a_nb_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_nb_1 + a_1b_n & a_nb_2 + a_2b_n & \dots & 2a_nb_n \end{bmatrix}$$

3. *Let* $f(x) = \frac{1}{1+e^{-x}}$ *and* $g(\boldsymbol{z}) = f\left(\boldsymbol{a}^T\boldsymbol{z} + b\right)$ *with* $\|\boldsymbol{a}\|_2 = 1$*. Show that*

$$D_a g(\boldsymbol{z}) = g(\boldsymbol{z})(1 - g(\boldsymbol{z})) \tag{3}$$

Solution
By the directional derivative and since $||a||_2 = 1$ we have that

$$\begin{aligned}D_a g(z) &= \lim_{h\to 0} \frac{f(a^T(z+ha)+b) - f(a^Tz+b)}{h} \\ &= \lim_{h\to 0} \frac{f(a^Tz+b+h) - f(a^Tz+b)}{h} \\ &= f'(a^Tz+b)\end{aligned}$$

Let's note that

$$f'(x) = \frac{e^{-x}}{1+e^{-x}}$$

Then,

$$D_a g(z) = f'(a^Tz+b) = \frac{e^{-(a^Tz+b)}}{1+e^{-(a^Tz+b)}}$$

Now, from (3) we get

$$\begin{aligned}g(\boldsymbol{z})(1-g(\boldsymbol{z})) &= \frac{1}{1+e^{-(a^Tz+b)}}\left(1 - \frac{1}{1+e^{-(a^Tz+b)}}\right) \\ &= \frac{1}{1+e^{-(a^Tz+b)}} - \frac{1}{(1+e^{-(a^Tz+b)})^2} \\ &= \frac{e^{-(a^Tz+b)}}{(1+e^{-(a^Tz+b)})^2}\end{aligned}$$

Therefore,

$$D_a g(z) = f'(a^Tz+b) = \frac{e^{-(a^Tz+b)}}{(1+e^{-(a^Tz+b)})^2} = g(\boldsymbol{z})(1-g(\boldsymbol{z}))$$

4. *Compute the gradient of*

$$f(\theta) \overset{def}{=} \frac{1}{2}\sum_{i=1}^{n} \left[g\left(\boldsymbol{x}_i\right) - g\left(\mathbf{A}\boldsymbol{x}_i + \boldsymbol{b}\right)\right]^2$$

with respect to θ*, where* $\theta = [a_{11}, a_{12}, a_{21}, a_{22}, b_1, b_2]^T$, $x_i \in \mathbb{R}^2$, $\mathrm{A} \in \mathbb{R}^{2\times 2}, \boldsymbol{b} \in \mathbb{R}^2$ *are defined as follows*

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

and $g : \mathbb{R}^2 \to \mathbb{R} \in \mathcal{C}^1$.
Solution
Let's note that

$$Ax_i + b = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + b_1 \\ a_{21}x_1 + a_{22}x_2 + b_2 \end{bmatrix}$$

Then,

$$D_\theta(Ax_i + b) = \begin{bmatrix} x_1 & x_2 & 0 & 0 & 1 & 0 \\ 0 & 0 & x_1 & x_2 & 0 & 1 \end{bmatrix}$$

Therefore, we get that

$$\begin{aligned} \nabla f(\theta) &= -\sum_{i=1}^{n} \left(g(x_i) - g(Ax_i + b)\right) D_\theta g(Ax_i + b) D_\theta(Ax_i + b) \\ &= -\sum_{i=1}^{n} \left(g(x_i) - g(Ax_i + b)\right) \left(\frac{dg}{dv_1}, \frac{dg}{dv_2}\right) \begin{bmatrix} x_1 & x_2 & 0 & 0 & 1 & 0 \\ 0 & 0 & x_1 & x_2 & 0 & 1 \end{bmatrix} \end{aligned}$$

5. *Show that* $\kappa(\mathbf{A}) \geq 1$ *where* $\|\mathbf{A}\| = \text{máx}_x \frac{\|\mathbf{A}x\|}{\|x\|}$. *(Hint: show that* $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$*)*
Solution
Let's prove that $||AB|| \leq ||A||||B||$ first. Let's note that

$$||A|| = \text{máx}_x \frac{||Ax||}{||x||} \geq \frac{||Ay||}{||y||} \Rightarrow ||Ay|| \leq ||A||||y|| \tag{4}$$

Now, considering (4) we get

$$\begin{aligned} ||AB|| &= \text{máx}_x \frac{||ABx||}{||x||} \\ &\leq \text{máx}_x \frac{||A||||Bx||}{||x||} \\ &\leq \text{máx}_x \frac{||A||||B||||x||}{||x||} \\ &= ||AB|| \end{aligned}$$

Therefore, we have proven that

$$||AB|| \leq ||A||||B|| \tag{5}$$

Now, by using (5) and the definition of the condition number of a matrix we get

$$\begin{aligned} \kappa(A) &= ||A||||A^{-1}|| \\ &\geq ||A \cdot A^{-1}|| \\ &= ||I|| = 1 \end{aligned}$$

6. *Show that* $x - \sin x = o\left(x^2\right)$, *as* $x \to 0$
Solution
We know by definition that

$$\text{lím}_{x \to a} \frac{f(x)}{g(x)} = 0$$

Then, using L'Hopital we get

$$\begin{aligned} \text{lím}_{x \to 0} \frac{x - sin(x)}{x^2} &= \text{lím}_{x \to 0} \frac{1 - cos(x)}{2x} \\ &= \text{lím}_{x \to 0} \frac{sin(x)}{2} = 0 \end{aligned}$$

Therefore, we have proven that $x - sin(x) = o(x^2)$ as $x \to 0$

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

   Solution

   Let's assume that $f(x) = o(g(x))$. Then by definition of $o(\cdot)$ we have that

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

   Then, by definition of limit for all $\epsilon > 0$ there exists $\delta > 0$ such that if $||x - 0|| < \delta$ then

$$\left|\frac{f(x)}{g(x)} - 0\right| < \epsilon$$
$$\Rightarrow |f(x)| < \epsilon|g(x)|$$

   which is the definition of $O(\cdot)$

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

   Solution

   We have that

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

   Then, by definition of $o(\cdot)$ we get

$$\begin{aligned}\lim_{x\to 0} \frac{f(x) + g(x)}{g(x)} &= \lim_{x\to 0} \frac{f(x)}{g(x)} + 1 = 0 \\ &\Rightarrow \lim_{x\to 0} \frac{f(x)}{g(x)} = -1\end{aligned}$$

   Therefore, $f(x) < 0$ since $g(x) > 0$ by hypothesis.