OPTIMIZATION. HOMEWORK 1

OSCAR DALMAU

(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 $f(\boldsymbol{x}) = [f_1(x_1,x_2), f_2(x_1,x_2)]^T = [12,16]^T$.

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

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

(4) Compute the gradient of

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

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

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

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

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

(6) Show 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 $0 < \|\boldsymbol{x}\| < \delta$.

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