Homework 2 for Math 173A - Fall 2024

1. Using the conditions of optimality, find the local maximizers and local minimizers, of the following functions and determine whether they are local maximizers or local minimimizers. You may use a computer to find the eigenvalues, but these questions should have easily accessible eigenvalues by hand.

   (a) $f : \mathbb{R}^2 \to \mathbb{R}$ for $f(x_1, x_2) = x_1^4 + 2x_2^4 - 4x_1 x_2$

   (b) $f : \mathbb{R}^3 \to \mathbb{R}$ for $f(\vec{x}) = \vec{x}^T A \vec{x} + b^T \vec{x}$, where

   $$A = \begin{bmatrix} -1 & 0 & 1/2 \\ 0 & -1 & 0 \\ 1/2 & 0 & -1 \end{bmatrix}, \qquad b = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$

   Bonues question: Are the local maximizers/minimizers you find also global maximizers/minimizers? If so, why?

2. Consider the problem $f : \mathbb{R}^n \to \mathbb{R}$ for $f(x) = \|Ax - b\|_2^2$ for $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. (Note: this was on the last homework). Write down the gradient descent algorithm to solve the optimization

   $$\min_{x \in \mathbb{R}^n} f(x).$$

   This doesn't have to be a computer program, just something of the form

   $$\begin{aligned} x^{(0)} &= ... \\ x^{(t+1)} &= ... \text{ (where the right hand side is in terms of } x^{(t)}). \end{aligned}$$

   You need to specify the formula of the gradient, but you do not need to specify the exact choice of the stepsize.

3. **Implementing Classification Model:** First some background for classification:

   - You are given labeled data $\{(x_i, y_i)\}_{i=1}^N$ for $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$.
   - Logistic regression involves choosing a label according to

     $$y = \operatorname{sign}(\langle w, x \rangle).$$

     Note we ignore the y-intercept term here, so we only need the optimal $w \in \mathbb{R}^d$.

- It turns out the correct function to minimize to find the weights is

$$F(w) = \frac{1}{N}\sum_{i=1}^{N} \log\left(1 + e^{-\langle w, x_i\rangle y_i}\right).$$

Questions:

(a) Is $F(w)$ a convex function?

(b) Write down the gradient descent algorithm for minimizing F. You need to specify the starting point, the formula of the gradient, but you do not need to specify the exact choice of the stepsize.

4. **Coding Question:** Recall that the equation for an ellipse in $\mathbb{R}^2$ is

$$a_1 x^2 + a_2 y^2 = 1.$$

Given data $\{(x_i, y_i)\}_{i=1}^{N} \subset \mathbb{R}^2$ that lie on (or near) the ellipse, you can find the best fit ellipse by solving

$$\min_{a\in\mathbb{R}^2} f(a)$$

where

$$f(a) = \sum_{i=1}^{N} (a_1 x_i^2 + a_2 y_i^2 - 1)^2$$

(a) Find an $A \in \mathbb{R}^{N\times 2}$ and $b \in \mathbb{R}^N$ such that $f(a) = \|Aa - b\|_2^2$. What is A in terms of (x_i, y_i)?

(b) Download the data provided on the HW page (called HW2_ellipse.csv), and create a scatter plot of the points (submit your code and the plot).

(c) Using Problem 2, create computer code to compute the gradient descent algorithm on this $f(a)$. The code must include a stopping condition. Use a step-size of $\eta = \frac{1}{2\|A^T A\|}$. Note, you cannot use a built-in gradient descent algorithm it must be written with a while or for loop. Also note, the norm of a matrix $\|X\| = \lambda_{\max}(X)$ is the largest eigenvalue (in magnitude) of X, and can be computed using "norm(X,2)" in matlab or "np.linalg.norm(X,2)" in Python. (submit the code)

(d) Using the data provided and your gradient descent code, estimate the solution a. Report a and $f(a)$. Given $f(a)$ and N, do you think you fit the data well or poorly? Given the convexity of f, do you think this is the optimal a?