MATH 173A: Homework #2

Due on Oct 22, 2024 at 23:59pm

Professor Cloninger

Ray Tsai

A16848188

Problem 1

Using the conditions of optimality, find the extreme points of the following functions and determine whether they are maxima or minima. 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 $f(x_1, x_2) = x_1^4 + 2x_2^4 - 4x_1x_2$

Proof. Note that

$$\nabla f(x) = (4x_1^3 - 4x_2, 8x_2^3 - 4x_1) = 0$$
$$\nabla^2 f(x) = \begin{bmatrix} 12x_1^2 & -4 \\ -4 & 24x_2^2 \end{bmatrix}.$$

Thus, the critical points are $x^* = (0, 0)$ or $(\pm 2^{-1/8}, \pm 2^{-3/8})$. We can then check

$$\nabla^2 f(0,0) = \begin{bmatrix} 0 & -4 \\ -4 & 0 \end{bmatrix}$$
$$\nabla^2 f(\pm 2^{-1/8}, \pm 2^{-3/8}) = \begin{bmatrix} 12 \cdot 2^{-1/4} & -4 \\ -4 & 24 \cdot 2^{-3/4} \end{bmatrix}.$$

Since the eigenvalues of $\nabla^2 f(0, 0)$ are ± 4, it is a saddle point. Since $\det \nabla^2 f(\pm 2^{-1/8}, \pm 2^{-3/8}) = 128 > 0$ amd $\frac{\partial^2 f}{\partial x_1^2} > 0$, the critical points $(\pm 2^{-1/8}, \pm 2^{-3/8})$ are local minima. □

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

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

Proof.

$$\nabla f(\vec{x}) = 2A\vec{x} + b,$$
$$\nabla^2 f(\vec{x}) = 2A.$$

Setting $\nabla f(\vec{x}) = 0$ yields

$$\vec{x}^* = -\frac{1}{2}A^{-1}b = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{2} \\ \frac{4}{3} \end{bmatrix}.$$

Since the eigenvalues of A are $-\frac{1}{2}, -1, -\frac{3}{2}$, we know $A \prec 0$ and the critical point is a local maximum. □

Problem 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

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

Proof. We already know f is convex and the gradient of f is

$$\nabla f(x) = 2A^T(Ax - b).$$

Thus, the gradient descent algorithm is

Pick $x^{(0)} \in \mathbb{R}^n$
for $t = 0, 1, 2, \dots$ **do**
    Set $x^{(t+1)} = x^{(t)} - \alpha\nabla f(x^{(t)}) = x^{(t)} - 2\alpha A^T(Ax^{(t)} - b)$
end for
return the last $x^{(t)}$.

□

Problem 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

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

  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

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

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

Proof. Note that

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

$$\nabla^2 F(w) = \frac{1}{N} \sum_{i=1}^N \frac{x_i^2 y_i^2 e^{-\langle w, x_i \rangle y_i}}{(1 + e^{-\langle w, x_i \rangle y_i})^2}.$$

Since the Hessian is positive semidefinite, $F(w)$ is convex. $\square$

(b) Find a gradient descent algorithm for minimizing F.