

# COMPSCI590.01 Deep Learning Theory Assignment 1

Feb. 14, 2025

## **Policy**

1. Homework must be typed.
2. You can collaborate with other students (up to groups of size 5), but each student must write down their individual solution. You should acknowledge all the students you discussed with.
3. You can use a LLM if you want, as long as you make sure that you understand what it is outputting and try to verify that the output is correct. You need to describe how you used the LLM (simply copying the prompt and answer works as long as the answer is not too long).

**Problem 1** (Monkey Saddle). Monkey saddle (see [https://en.wikipedia.org/wiki/Monkey\\_saddle](https://en.wikipedia.org/wiki/Monkey_saddle)) is a 3rd order saddle point. For function  $f(x, y) = x^3 - 3xy^2 = x(x - \sqrt{3}y)(x + \sqrt{3}y)$  at point  $(x, y)$ , it is easy to check that both gradient and Hessian are equal to 0. Yet the point is not a local minimum or local maximum. In fact it has 3 descending directions.

(a) (10 points) Run gradient descent starting from  $(x, y) = (-0.1, 0.1)$  with step size  $\eta = 0.5$  for 10 iterations. Give the  $(x, y)$  values for each iteration. (Obviously, please write a program to do this. Feel free to do that using a LLM. You don't need to submit the code or acknowledge LLM use for the code.)

(b) (10 points) Prove the following 3rd order sufficient condition for saddle points. Suppose function  $f(x) : \mathbb{R}^d \rightarrow \mathbb{R}$  has bounded 4th order derivative. That is, for any  $x$  and  $v$  where  $\|v\| = 1$ , let  $g(\delta) = f(x + \delta v)$  and  $g^{(4)}(\delta)$  be its 4th order derivative, then  $|g^{(4)}(0)| \leq Z$ .

For any  $x$  with  $\nabla f(x) = 0$ , let  $S$  be the nullspace of  $\nabla^2 f(x)$  (that is, for any  $v \in S$ ,  $[\nabla^2 f(x)]v = 0$ ), if there exists a direction  $v \in S$  with  $\|v\| = 1$  such that  $g(\delta) = f(x + \delta v)$  has 3rd order derivative  $g^{(3)}(0) \neq 0$ , then  $x$  is a saddle point of  $f(x)$ .

(Not homework) Does perturbed gradient descent allow you to find an approximate 3rd order stationary point (see definitions in <https://arxiv.org/abs/1602.05908>)? I believe so but I don't know how to prove that. As far as I know this is still an open problem.

**Problem 2** (Phase Retrieval). Phase retrieval is a problem where there is an unknown signal  $u$  and we observe measures of the form  $|\langle u, a \rangle|$ . In this problem for simplicity we assume  $u \in \mathbb{R}^d$ , and the measurement  $a \sim N(0, I)$  is a standard Gaussian. Suppose we are given infinitely many observations  $a$  on unit vector  $\|u\|$ , one way to find  $u$  by optimization is to solve the following optimization problem:

$$\min_x f(x) := \mathbb{E}[|\langle u, a \rangle|^2 - \langle x, a \rangle|^2]^2.$$

(a) (10 points) Compute the expectation over  $a$  in the definition of  $f(x)$ , write  $f(x)$  as a linear combination of  $(\|x\|^2 - 1)^2$  and  $\|x\|^2 - \langle x, u \rangle^2$ . Hint: both  $\langle u, a \rangle$  and  $\langle x, a \rangle$  are single variate Gaussians. For  $z \sim N(0, \sigma^2)$ ,  $\mathbb{E}[z^4] = 3\sigma^4$ . The cross term can be handled if you write  $x = \lambda u + v$  where  $v \perp u$ .

(b) (10 points) What is the set of first order stationary points for  $f(x)$ ?

(c) (10 points) Show that  $x = \pm u$  are the only two second order stationary points for  $f(x)$ .

Recall the second order stationary point definition we used in class: a point  $x$  is a second order stationary point if  $\nabla f(x) = 0$  and  $\nabla^2 f(x) \succeq 0$ .