Homework 3 for Math 173A - Fall 2024

1. Determine whether each function is Lipschitz, and if so find the smallest possible Lipschitz constant for the function. For all problems, $\|\cdot\|$ represents the Euclidean norm (2-norm).
   (a) $f : \mathbb{R}^n \to \mathbb{R}$ for $f(x) = \|x\|$
   (b) $f : \mathbb{R}^n \to \mathbb{R}$ for $f(x) = \|x\|^2$
   (c) $\rho : \mathbb{R} \to \mathbb{R}$ for $\rho(x) = \frac{1}{1+e^{-x}}$.
   (d) $f : \mathbb{R}^n \to \mathbb{R}$ for $f(x) = \rho(w^T x + b)$ for some weight vector $w \in \mathbb{R}^n$, $b \in \mathbb{R}$, and ρ from part (c).

2. Let f be a convex and differentiable. Let x^* be the global minimum and suppose $x^{(0)}$ is the initialization such that $\|x^* - x^{(0)}\| \leq 5$.
   (a) Let f be $L-$Lipschitz function where $L = 3$. Determine the step size [1] μ and number of steps needed to satisfy
   $$\left\| f\left(\frac{1}{t}\sum_{s=0}^{s-1} x^{(s)}\right) - f\left(x^*\right) \right\| \leq 10^{-4}.$$
   (b) Let f be $L-$smooth where $L = 3$. Determine the step size μ and number of steps needed to satisfy
   $$\left\| f\left(x^{(t)}\right) - f\left(x^*\right) \right\| \leq 10^{-4}.$$

3. Consider the function $f(x_1, x_2) = (2x_1 - 1)^4 + (x_1 + x_2 - 1)^2$.
   (a) Find the global minimum of f, and justify your answer.
   (b) Starting at $x^{(0)} = (0, 0)$, perform gradient descent with backtracking line-search.
      i. Starting at $x^{(0)} = (0, 0)$ with stepsize, which is also called learning rate in the machine learning community, $\mu^{(0)}$, write down the gradient descent equation for $x^{(1)}$.
      ii. Suppose we want to set $\mu^{(0)}$ using backtracking line search with $\gamma = 0.2$ and Armijo's condition $f(x^{(1)}) \leq f(x^{(0)}) - \mu^{(0)}\gamma\|\nabla f(x^{(0)})\|_2^2$. Find a value of $\mu^{(0)}$ that satisfies this.
      iii. Suppose instead you started with $\mu^{(0)} = 1$ and an update of $\mu^{(0)} \leftarrow \frac{1}{2}\mu^{(0)}$ (i.e. $\beta = \frac{1}{2}$). In the worst case, how many steps of back-tracking would you have to take before accepting $x^{(1)}$?

[1] Note that we mostly use η in the lectures to denote the stepsize.