Discussion 7

1. Gradient Descent with A Wide Matrix (Fall 2022 Midterm)

Consider a matrix $X \in \mathbb{R}^{n \times d}$ with $n < d$ and a vector $\vec{y} \in \mathbb{R}^n$, both of which are known and given to you. Suppose X has full row rank.

(a) Consider the following problem:

$$X\vec{w} = \vec{y} \tag{1}$$

where $\vec{w} \in \mathbb{R}^d$ is unknown. How many solutions does 1 have? *Justify your answer.*

(b) Consider the minimum-norm problem

$$\vec{w}_\star = \underset{\substack{\vec{w} \in \mathbb{R}^d \\ X\vec{w} = \vec{y}}}{\operatorname{argmin}} \|\vec{w}\|_2^2 . \tag{2}$$

We know that the optimal solution to this problem is $\vec{w}_\star = X^\top (XX^\top)^{-1}\vec{y}$. Now let $X = U\Sigma V^\top = U \begin{bmatrix} \Sigma_1 & 0 \end{bmatrix} V^\top$ be the SVD of X, where $\Sigma_1 \in \mathbb{R}^{n \times n}$. Recall that this is possible because $n < d$ and X is full row rank. Prove that $\vec{w}_\star$ is given by

$$\vec{w}_\star = V \begin{bmatrix} \Sigma_1^{-1} \\ 0 \end{bmatrix} U^\top \vec{y}. \tag{3}$$

(c) Let $\eta > 0$, and I be the identity matrix of appropriate dimension. Using the SVD $X = U \begin{bmatrix} \Sigma_1 & 0 \end{bmatrix} V^\top$, prove the following identity for all positive integers $i > 0$:

$$(I - \eta X^\top X)^i = V \left(I - \eta \begin{bmatrix} \Sigma_1^2 & 0 \\ 0 & 0 \end{bmatrix} \right)^i V^\top. \tag{4}$$

(d) Recall that $X \in \mathbb{R}^{n \times d}$, and that we can write the SVD of X as $X = U \begin{bmatrix} \Sigma_1 & 0 \end{bmatrix} V^\top$. We will use gradient descent to solve the minimization problem

$$\min_{\vec{w} \in \mathbb{R}^d} \frac{1}{2} \left\| X\vec{w} - \vec{y} \right\|_2^2 \tag{5}$$

with step-size $\eta > 0$. Let $\vec{w}_0 = \vec{0}$ be the initial state, and $\vec{w}_k$ be the k^{th} iterate of gradient descent. Use the identity:

$$(I - \eta X^\top X)^i = V \left(I - \eta \begin{bmatrix} \Sigma_1^2 & 0 \\ 0 & 0 \end{bmatrix} \right)^i V^\top. \tag{6}$$

to prove that after k steps, we have

$$\vec{w}_k = \eta \sum_{i=0}^{k-1} V \left(I - \eta \begin{bmatrix} \Sigma_1^2 & 0 \\ 0 & 0 \end{bmatrix} \right)^i \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} U^\top \vec{y}. \tag{7}$$

HINT: Remember to set $\vec{w}_0 = \vec{0}$.

(e) Now let $0 < \eta < \frac{1}{\sigma_1^2}$, where σ_1 denotes the maximum singular value of $X = U \begin{bmatrix} \Sigma_1 & 0 \end{bmatrix} V^\top$. Let $\vec{w}_k$ be given as

$$\vec{w}_k = \eta \sum_{i=0}^{k-1} V \left(I - \eta \begin{bmatrix} \Sigma_1^2 & 0 \\ 0 & 0 \end{bmatrix} \right)^i \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} U^\top \vec{y}. \tag{8}$$

and let $\vec{w}_\star$ be the minimum norm solution given as

$$\vec{w}_\star = V \begin{bmatrix} \Sigma_1^{-1} \\ 0 \end{bmatrix} U^\top \vec{y}. \tag{9}$$

Prove that $\lim_{k \to \infty} \vec{w}_k = \vec{w}_\star$.

HINT: You may use the following result without proof. When all eigenvalues of $A \in \mathbb{R}^{n \times n}$ have magnitude < 1, we have the identity $(I - A)^{-1} = I + A + A^2 + \dots$.

2. Convexity and Composition of Functions

Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$. Define the composition of f with g as $h = f \circ g : \mathbb{R}^n \to \mathbb{R}$ such that $h(\vec{x}) = f(g(\vec{x}))$.

(a) Show that if f is convex and non decreasing and g is convex, then h is convex.

(b) Show that there exists f non decreasing and g convex, such that $h = f \circ g$ is not convex.

(c) Show that there exists f convex and g convex such that $h = f \circ g$ is not convex.