CMSE 820 Homework 01. Due Sep 15, 2020.

I. Let $A \in \mathbb{R}^{m \times n}$ $(m < n)$, and let $A = U\Sigma V^T$ be the singular value decomposition, where U and V are $m \times m$ and $n \times n$ orthogonal matrices respectively, and

$$\Sigma = \begin{pmatrix} \sigma_1 & & & & & 0 & \dots & 0 \\ & \ddots & & & & 0 & & 0 \\ & & \sigma_r & & & 0 & & 0 \\ & & & 0 & & 0 & & 0 \\ & & & & \ddots & 0 & & 0 \\ & & & & 0 & 0 & \dots & 0 \end{pmatrix}_{m \times n}$$

where r is the rank of A and $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$ are the singular values of A in decreasing order. Write the columns of U as $u_1, \dots, u_m$ and the columns of V as $v_1, \dots, v_n$. Prove

(a) $\mathrm{Ker}(A) = \mathrm{span}\{v_{r+1}, \dots, v_n\}$.

Solution. Suppose $x \in \mathrm{span}\{v_{r+1}, \dots, v_n\} \subseteq \mathbb{R}^n$ is a vector of the following form,

$$x^T = \begin{pmatrix} 0 & 0 & \dots & 0 & x_{r+1} & \dots & x_n \end{pmatrix}$$

where r is the rank of A. Let $w = Vx$. Consider $Aw = U\Sigma V^T w$. Since V is orthogonal, it follows that $V^T = V^{-1}$, thus $Aw = U\Sigma x$. The singular values of A are arranged such that the first r diagonal elements of Σ are nonzero, and the remaining $m-r$ elements are 0. However, since the first r components of x are zero and the remaining $n-r$ elements are nonzero, the multiplication of $\Sigma x = \mathbf{0}$ for any x of the prescribed form. Therefore, $Aw = \mathbf{0}$. We can say that $\mathrm{span}\{v_{r+1}, \dots, v_n\} \subseteq \mathrm{Ker}(A)$.

Now suppose $x \in \mathrm{Ker}(A) \Rightarrow Ax = 0$. We rewrite this as $U\Sigma V^T x = 0 \Rightarrow \Sigma V^T x = 0$ since U is orthogonal, which implies $U^T = U^{-1}$. We can write this as a system of equations

$$\begin{aligned} \sigma_1 v_1^T x &= 0 \\ \sigma_2 v_2^T x &= 0 \\ &\vdots \\ \sigma_r v_r^T x &= 0 \\ 0 &= 0 \\ &\vdots \\ 0 &= 0. \end{aligned}$$

Since $\sigma_1 \geq \dots \geq \sigma_r > 0$, it follows that

$$\begin{aligned} v_1^T x &= 0 \\ v_2^T x &= 0 \\ &\vdots \\ v_r^T x &= 0. \end{aligned}$$

V is an orthogonal matrix, which means $v_i^T v_j = \delta_{ij}$. Additionally, the columns of V, v_i, automatically form a basis for $\mathbb{R}^n$ because mutual orthogonality implies linear independence. We can thus write x as a linear combination of the basis vectors, $x = \sum_{j=1}^{n} c_i v_i$ where $c_i \in \mathbb{R}$. Since $v_i^T x = 0$ for $i \leq r$, it follows that $c_i = 0$ for $i \leq r$. We can conclude that $x \in \text{span}\{v_{r+1}, \dots, v_n\} \Rightarrow \text{Ker}(A) \subseteq \text{span}\{v_{r+1}, \dots, v_n\}$. ■

(b) $\text{Ran}(A) = \text{span}\{u_1, \dots, u_r\}$.

Solution. The singular value decomposition of A can be rewritten as $A = \sum_{j=1}^{m} u_j \sigma_j v_j^T$, where u_j and v_j are the columns of U and V respectively. Suppose $y \in \text{Ran}(A) \Rightarrow \exists x$ s.t. $Ax = y$. $Ax = \left(\sum_{j=1}^{m} u_j \sigma_j v_j^T\right) x = \sum_{j=1}^{m} u_j \sigma_j v_j^T x = y$. The values σ_j and $v_j^T x$ are both scalars, therefore, we can let $\gamma_j = \sigma_j v_j^T x$. Thus, $Ax = \sum_{j=1}^{m} \gamma_j u_j = y$. However, for $j > r$, $\sigma_j = 0 \Rightarrow \gamma_j = 0$, so we can rewrite the sum as $Ax = \sum_{j=1}^{r} \gamma_j u_j + \sum_{j=r+1}^{m} 0 = \sum_{j=1}^{r} \gamma_j u_j = y$. Therefore $y \in \text{span}\{u_1, \dots, u_r\} \Rightarrow \text{Ran}(A) \subseteq \text{span}\{u_1, \dots, u_r\}$.
Suppose now $y \in \text{span}\{u_1, \dots, u_r\}$. We can write y as a linear combination $y = x_1 u_1 + \cdots + x_r u_r$ where $x_j \in \mathbb{R}$. From the SVD of A, we can write $A = U\Sigma V^T \Rightarrow Av_j = \sigma_j u_j$. The first r singular values are nonzero, so for $j \leq r$, $u_j = \frac{Av_j}{\sigma_j}$. Therefore,

$$y = x_1 \frac{Av_1}{\sigma_1} + \cdots + x_r \frac{Av_r}{\sigma_r}.$$

Let $x' = \frac{x_1}{\sigma_1} v_1 + \cdots + \frac{x_r}{\sigma_r} v_r$ such that $y = Ax'$. Therefore, $y \in \text{Ran}(A)$ and $\text{span}\{u_1, \dots, u_r\} \subseteq \text{Ran}(A)$. ■

II. Recall that a square matrix P is called an orthogonal projection matrix if $P^2 = P^T = P$. Let $A \in \mathbb{R}^{m\times n}$ and suppose $(A^T A)^{-1}$ exists. Define

$$H := A(A^T A)^{-1} A^T.$$

(a) Prove that H is an orthogonal projection matrix.

Solution.

$$\begin{aligned} H^2 &= A(A^T A)^{-1} A^T (A(A^T A)^{-1} A^T) \\ &= A(A^T A)^{-1} (A^T A)(A^T A)^{-1} A^T \\ &= A(A^T A)^{-1} A^T = H. \end{aligned}$$

$$\begin{aligned} H^T &= (A(A^T A)^{-1} A^T)^T \\ &= (A^T)^T (A^T A)^{-T} A^T \\ &= A(A^T (A^T)^T)^{-1} A^T \\ &= A(A^T A)^{-1} A^T = H. \end{aligned}$$

Since $H^2 = H^T = H$, it follows that H is an orthogonal projection matrix. ■

(b) Let $x \in \mathbb{R}^m$ be a vector. If $x \in \text{Ran}(H)$, prove that $Hx = x$. If $x \in \text{Ran}(H)^\perp$, prove that $Hx = 0$.

Solution. Suppose $x \in \text{Ran}(H)$. H is of size $m \times m$, so $\exists y \in \mathbb{R}^m$ such that $Hy = x$. If we multiply on the left by H, we get $HHy = Hx$. Since H is an orthogonal projection, $H^2 = H$, thus $Hy = Hx$. Therefore, $y = x$, and $\forall x \in \text{Ran}(H)$, $Hx = x$.

Suppose instead that $x \in \text{Ran}(H)^\perp$. From this, $\forall y \in \text{Ran}(H), y^T x = 0$. The vector x is therefore in the kernel of H, and thus $Hx = 0$. ■

(c) Let $a_1, \dots, a_n$ be the columns of A. Prove that

$$\text{Ran}(H) = \text{span}\{a_1, \dots, a_n\}.$$

Solution. Suppose $x \in \text{Ran}(H)$. From 2(b), we know that this implies $Hx = x$. We can rewrite H as $A(A^TA)^{-1}A^Tx = x$. Let $y = (A^TA)^{-1}A^Tx \in \mathbb{R}^n$. Thus, $Ay = x$. Since $\exists y$ s.t. $Ay = x$, it follows that $x \in \text{Ran}(A) = \text{span}\{a_1, \dots, a_n\}$, and therefore $\text{Ran}(H) \subseteq \text{span}\{a_1, \dots, a_n\}$.

Suppose now $x \in \text{span}\{a_1, \dots, a_n\}$. It follows that $\exists y$ such that $Ay = x$. Left multiplying by A^T we get $A^TAy = A^Tx$. Since $(A^TA)^{-1}$ exists, we can equate this to $y = (A^TA)^{-1}A^Tx$. Left multiplying by A on both sides, we see $Ay = A(A^TA)^{-1}A^Tx = Hx \Rightarrow Ay = x = Hx$. Therefore, we can conclude that $x \in \text{Ran}(H)$ and $\text{span}\{a_1, \dots, a_n\} \subseteq \text{Ran}(H)$. ■

III. Let f be a function with domain $D(f)$.

(a) If f is a convex function, prove that for any $M \in \mathbb{R}$, $\{x \in D(f) : f(x) \leq M\}$ is a convex set.

Solution. Suppose $x_1, x_2 \in \{x \in D(f) : f(x) \leq M\}$ for a given value of M and f is a convex function. It follows by the convexity of f that for any $\lambda \in [0,1], f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) \leq \lambda M + (1-\lambda)M = M$. Therefore, $\lambda x_1 + (1-\lambda)x_2 \in \{x \in D(f) : f(x) \leq M\}$. The set $\{x \in D(f) : f(x) \leq M\}$ is thus a convex set. ■

(b) Give a counter-example to show the converse of (a) is not true.

Solution. Consider $f(x) = \log(x)$. Let $x, y \in \mathbb{R}, \lambda \in [0,1]$. $f(\lambda x + (1-\lambda)y) = \log(\lambda x + (1-\lambda)y)$, and $\lambda f(x) + (1-\lambda)f(y) = \lambda \log(x) + (1-\lambda)\log(y)$. Suppose $\lambda = 0.3, x = 1, y = 2$ as an example. $\log(\lambda x + (1-\lambda)y) \approx 0.53$ and $\lambda \log(x) + (1-\lambda)\log(y) \approx 0.49$. Thus, $f(\lambda x + (1-\lambda)y) \not\leq \lambda f(x) + (1-\lambda)f(y)$, and therefore f is not convex.

Let $x_1, x_2 \in \{x \in D(f) : f(x) \leq M\}$ for some $M \in \mathbb{R}$, and let $\lambda \in [0,1]$. Without loss of generality, assume that $x_1 \leq x_2$. Consider $f(\lambda x_1 + (1-\lambda)x_2)$. We know that $f(x_1) = \log(x_1) \leq M$ and $f(x_2) = \log(x_2) \leq M$. Additionally, $x_1 \leq \lambda x_1 + (1-\lambda)x_2 \leq x_2$. Since $f(x)$ is a monotonically increasing function, it follows that $f(x_1) \leq f(\lambda x_1 + (1-\lambda)x_2) \leq f(x_2) \Rightarrow f(\lambda x_1 + (1-\lambda)x_2) \leq M$. Thus, $\lambda x_1 + (1-\lambda)x_2 \in \{x \in D(f) : f(x) \leq M\}$.

$f(x)$ has convex level sets but is not convex, therefore the reverse statement of (a) is not true. ■

IV. Prove that a function f is a convex function if and only if the following set

$$epi(f) := \{(x, t) \in D(f) \times \mathbb{R} : f(x) \leq t\}$$

is a convex set.

Solution. Suppose $(x_1, t_1), (x_2, t_2) \in epi(f)$ and f is a convex function. From this, we know $f(x_1) \leq t_1$ and $f(x_2) \leq t_2$. By the convexity of f, we know for any $\lambda \in [0, 1]$, $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) \leq \lambda t_1 + (1-\lambda)t_2$, which implies that $\lambda t_1 + (1-\lambda)t_2 \in epi(f)$. As such $epi(f)$ is a convex set.

Suppose $epi(f)$ is a convex set. Let $x_1, x_2 \in \mathbb{R}$. We know $(x_1, f(x_1))$ and $(x_2, f(x_2))$ lie in the epigraph of f, and by the convexity of $epi(f)$, for any $\lambda \in [0, 1], (\lambda x_1 + (1-\lambda)x_2, \lambda f(x_1) + (1-\lambda)f(x_2)) \in epi(f) \Rightarrow f(\lambda x_1 + (1-\lambda)x_2) \leq f(x_1) + (1-\lambda)f(x_2)$. f therefore must be a convex function. ■