Names: ____________________

NetIDs: ____________________

Discussion section: **D** ________ **Group number:** ________

MATH 257 - WORKSHEET 14

Given a real $m \times n$ matrix A, recall we have a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ defined by $L(\mathbf{x}) = A\mathbf{x}$. When A is also square and symmetric, the Spectral Theorem assures us an orthonormal eigenbasis $\mathcal{B}$ for A. This implies that $L_{\mathcal{B},\mathcal{B}}$ is a diagonal matrix with diagonal entries the eigenvalues of A. Note that the basis $\mathcal{B}$ is doubly nice: (1) it is orthonormal, and (2) it is an eigenbasis for A. For general rectangular matrices, can we find nice ordered bases $\mathcal{F}, \mathcal{G}$ for $\mathbb{R}^n, \mathbb{R}^m$ respectively so that $L_{\mathcal{G},\mathcal{F}}$ is diagonal, i.e. having non-zero entries only along the main-diagonal?
In this worksheet we will investigate how to do this by finding certain orthonormal bases (**ONB**) for each of the four fundamenatal subspaces so that they together "play well" with the matrix A to yield a coordinate matrix that is diagonal. This will lead us to the powerful **Singular Value Decomposition** (**SVD**).
Note: Occasionally we will use brackets to denote matrices but this has no additional mathematical meaning. It is for ease of typesetting.

(1) First, an example: Consider the matrix $A = \begin{pmatrix} 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 1 & 1 & -1 & -1 \end{pmatrix}$ which has the following four fundamental subspaces:

$$Col(A^T) = span\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \} \qquad Nul(A) = span\{ \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \end{pmatrix} \}$$

$$Col(A) = span\{ \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix} \} \qquad Nul(A^T) = span\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \}$$

where the given tuples are basis vectors for their respective subspaces. Let

$$\mathcal{F} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4) = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \end{pmatrix} \right)$$

$$\mathcal{G} = (\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3) = \left(\begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right)$$

(a) Assuming $\mathcal{G}, \mathcal{F}$ are bases for $\mathbb{R}^m, \mathbb{R}^n$ respectively, find $L_{\mathcal{G},\mathcal{F}}$.

We compute that

$$\begin{aligned} A\mathbf{v}_1 &= 2\mathbf{w}_1 + 2\mathbf{w}_2 \\ A\mathbf{v}_2 &= 2\mathbf{w}_2 \end{aligned}$$

so that

$$L_{\mathcal{G},\mathcal{F}} = \left[\begin{array}{cc|cc} 2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right]$$

which is (lower) "triangular".

(b) (**Optional**, slight messy) What if we apply the Gram-Schmidt process to get orthonormal bases $\hat{\mathcal{F}}, \hat{\mathcal{G}}$? What is $L_{\hat{\mathcal{G}},\hat{\mathcal{F}}}$.

We compute that

$$\hat{\mathcal{F}} = (\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \hat{\mathbf{v}}_3, \hat{\mathbf{v}}_4) = \left(\frac{1}{2}\begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \frac{1}{2}\begin{pmatrix}-1\\-1\\1\\1\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\\0\\0\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}0\\0\\-1\\1\end{pmatrix}\right)$$

$$\hat{\mathcal{G}} = (\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, \hat{\mathbf{w}}_3) = \left(\frac{1}{3}\begin{pmatrix}2\\2\\1\end{pmatrix}, \frac{1}{3\sqrt{2}}\begin{pmatrix}1\\1\\-4\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\\0\end{pmatrix}\right)$$

Let V, U be matrices with columns the elements of $\hat{\mathcal{F}}, \hat{\mathcal{G}}$ respectively so that $V = I_{\mathcal{E}_4,\hat{\mathcal{F}}}$ and $U = I_{\mathcal{E}_3,\hat{\mathcal{G}}}$ are orthogonal matrices where $\mathcal{E}_k$ is the standard basis in $\mathbb{R}^k$. Then,

$$L_{\mathcal{G},\mathcal{F}} = U^T A V = \left[\begin{array}{c|c} B & 0 \\ \hline 0 & 0 \end{array}\right]_{4\times 3} \qquad B = \frac{1}{3}\begin{pmatrix} 16 & -2 \\ 4\sqrt{2} & 4\sqrt{2} \end{pmatrix}$$

In particular,

$$\begin{aligned} A\hat{\mathbf{v}}_1 &= \frac{16}{3}\hat{\mathbf{w}}_1 + \frac{4\sqrt{2}}{3}\hat{\mathbf{w}}_2 \\ A\hat{\mathbf{v}}_2 &= -\frac{2}{3}\hat{\mathbf{w}}_1 + \frac{4\sqrt{2}}{3}\hat{\mathbf{w}}_2 \end{aligned}$$

(2) We didn't quite achieve our goal, but the preceding example(s) have been informative. Let's consider what the previous example(s) indicate.

(a) First, let's clear up the assumption we used earlier: Suppose $rk(A) = r$ with $\mathcal{A} = (\mathbf{v}_1, \dots, \mathbf{v}_r)$ an ordered ONB for $Col(A^T)$ and $\mathcal{B} = (\mathbf{v}_{r+1}, \dots, \mathbf{v}_n)$ an ordered ONB for $Nul(A)$. Explain why $\mathcal{F} = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ is an ordered ONB for $\mathbb{R}^n$. Similarly, if $\mathcal{C} = (\mathbf{u}_1, \dots, \mathbf{u}_r)$ is an ordered ONB for $Col(A)$ and $\mathcal{D} = (\mathbf{u}_{r+1}, \dots, \mathbf{u}_m)$ an ordered ONB for $Nul(A^T)$, then $\mathcal{G} = (\mathbf{u}_1, \dots, \mathbf{u}_m)$ is an ordered ONB for $\mathbb{R}^m$.

$Col(A^T), \mathrm{Nul}(A)$ are orthogonal complements of each other in $\mathbb{R}^n$ so that $\mathcal{F}$ is a spanning set for $\mathbb{R}^n$. Since it contains $n = \dim(\mathbb{R}^n)$ vectors, it is a basis for $\mathbb{R}^n$. The claim for $\mathcal{G}$ follows analogously.

(b) Next, notice that we could choose *any* ONB for each of $Nul(A), Nul(A^T)$ without changing the fact that $L_{\mathcal{G},\mathcal{F}} = \left[\begin{array}{c|c} B & 0 \\ \hline 0 & 0 \end{array}\right]_{m\times n}$ for some matrix $B_{r\times r}$. Why is this true?

For any vector $\mathbf{v} \in Nul(A)$ we have $A\mathbf{v} = \mathbf{0}$ so that columns $r+1, \ldots, n$ of $L_{\mathcal{G},\mathcal{F}}$ are always zero. For any $\mathbf{v} \in V$ we have $A\mathbf{v} \in Col(A)$ so that written as a linear combination of basis vectors, we need only utilize those for $Col(A)$. Thus, the lower-left block of $L_{\mathcal{G},\mathcal{F}}$ is entirely zero.

(c) If we want B diagonal, we need ONB $(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ for $Col(A^T)$ and ONB $(\mathbf{u}_1, \ldots, \mathbf{u}_r)$ for $Col(A)$ compatible in the sense that for each $i = 1, \ldots, r$ the tuple $A\mathbf{v}_i$ is a multiple of $\mathbf{u}_i$. Show that this implies the collection $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ is *orthogonal.*

The collection $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is orthogonal, and $A\mathbf{v}_i$ is parallel to $\mathbf{u}_i$ for each $i = 1, \ldots, r$. Thus, $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ is orthogonal.

(3) The preceding result implies for $i \neq j$ we have $0 = (A\mathbf{v}_i) \cdot (A\mathbf{v}_j) = \mathbf{v}_i^T(A^TA)\mathbf{v}_j$ so that we want to choose ONB $(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ for $Col(A^T)$ with the property that $0 = \mathbf{v}_i \cdot (A^TA)\mathbf{v}_j$ whenever $i \neq j$.

(a) Show that this condition implies that the $\mathbf{v}_i$ are eigenvectors of A^TA.

The condition implies that $A^TA\mathbf{v}_i$ is orthogonal to all $\mathbf{v}_j$ for $i \neq j$. Since $A^TA\mathbf{v}_i \in Col(A^T)$ and $\mathbf{v}_1, \ldots, \mathbf{v}_r$ is an ONB for $Col(A^T)$, we have $A^TA\mathbf{v}_i \in span(\mathbf{v}_i)$ so that $\mathbf{v}_i$ is an eigenvector of A^TA.

(b) This directs our attention to the matrix A^TA. Recall from the lecture modules that $\mathrm{Nul}(A^TA) = \mathrm{Nul}(A)$. Explain why an ONB for the 0-eigenspace of A^TA is an ONB for $Nul(A)$ and why $rk(A^TA) = r$

The 0-eigenspace of a square matrix is the same as the nullspace of the matrix. Since $\mathrm{Nul}(A^TA) = \mathrm{Nul}(A)$, the first claim follows. Then, $\dim Nul(A^TA) = n - r$ so that $rk(A^TA) = r$ by the Rank-Nullity Theorem.

(c) Since A^TA is symmetric, the Spectral Theorem yields an orthonormal eigenbasis $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$. Suppose $\mathcal{B} = (\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n)$ is an ONB for $Nul(A^TA)$. Why is $\mathcal{A} = (\mathbf{v}_1, \ldots, \mathbf{v}_r)$ an ONB for $Col(A^T)$?

$Col(A^T)$ and $Nul(A)$ are orthogonal complements in $\mathbb{R}^n$ and so are $Col((A^TA)^T)$ and $Nul(A^TA)$. Since $Nul(A^TA) = Nul(A)$, we have $\mathcal{A}$ an ONB for $Col((A^TA)^T) = Col(A^T)$.

(4) Let $\mathcal{F} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ orthonormal eigenbasis for A^TA, which we now recognize as merging ONB for each of $Col(A^T)$ and $\mathrm{Nul}(A)$.

(a) By our earlier work, we know that $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ is an orthogonal collection in $Col(A)$. Show that the collection is actually an orthogonal basis for $Col(A)$.

It suffices to show that the collection is either linearly independent or a spanning set for $Col(A)$. Since $\dim(Col(A)) = r$ this will show that the collection is a basis.
Say we want to show the former: Since $\mathbf{v}_1, \ldots, \mathbf{v}_r \in Col(A^T) = \mathrm{Nul}(A)^\perp$ the orthogonal collection $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ does not contain zero and is hence linearly independent.
Alternatively, $\mathbf{0} = \sum_{i=1}^r a_i A\mathbf{v}_i = A(\sum_{i=1}^r a_i \mathbf{v}_i)$ implies $\sum_{i=1}^r a_i \mathbf{v}_i \in Nul(A)$. Since $\mathbf{v}_i \in Col(A^T)$ so is their linear combination. Thus, $\sum_{i=1}^r a_i \mathbf{v}_i = 0$ and the linear independence of $\mathbf{v}_1, \ldots, \mathbf{v}_r$ implies $a_i = 0$ for all $i = 1, \ldots, r$.
Say we want to show the former: Any vector $\mathbf{w} \in Col(A)$ has the form $\mathbf{w} = A\mathbf{v}$ for some $\mathbf{v} \in V$. Writing $\mathbf{v} = \sum_{i=1}^n a_i \mathbf{v}_i$ as a linear combination of the basis vectors $\{\mathbf{v}_i\}$ we have $\mathbf{w} = A\mathbf{v} = \sum_{i=1}^n a_i A\mathbf{v}_i = \sum_{i=1}^r a_i A\mathbf{v}_i$. Thus, the collection $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ is a spanning set for (A).

(b) Find the lengths $\|A\mathbf{v}_i\|$ and so obtain an ONB $\mathcal{C} = (\mathbf{w}_1, \ldots, \mathbf{w}_r)$ for $Col(A)$.

For $i = 1, \ldots, r$ we have

$$\|A\mathbf{v}_i\|^2 = \mathbf{v}_i^T A^T A \mathbf{v}_i = \lambda_i \mathbf{v}_i^T \mathbf{v}_i = \lambda_i$$

since $(A^T A)\mathbf{v}_i = \lambda_i \mathbf{v}_i$ and the $\mathbf{v}_i$ are unit length. This computation also shows that the non-zero eigenvalues of $A^T A$ are strictly positive. Thus, if we define $\mathbf{u}_i = \frac{1}{\sqrt{\lambda_i}} A\mathbf{v}_i$ then the collection $\mathcal{C} = (\mathbf{u}_1, \ldots, \mathbf{u}_r)$ is an ONB for $Col(A)$.

(c) Pick any ONB $\mathcal{D} = (\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n)$ for $Nul(A^T)$. Let $\mathcal{G} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$. Find $L_{\mathcal{G},\mathcal{F}}$.

We compute

$$\begin{aligned} A\mathbf{v}_i &= \sqrt{\lambda_i}\frac{1}{\sqrt{\lambda_i}} A\mathbf{v}_i = \sqrt{\lambda_i}\mathbf{u}_i \qquad && i = 1, \ldots, r \\ A\mathbf{v}_i &= \mathbf{0} \qquad i = r+1, \ldots, n && \end{aligned}$$

so that

$$L_{\mathcal{G},\mathcal{F}} = \left[\begin{array}{c|c} D & 0 \\ \hline 0 & 0 \end{array}\right]_{m\times n} \qquad D = \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_r} \end{pmatrix}_{r\times r}$$

with D diagonal as desired.

(d) Let $\mathcal{E}_n, \mathcal{E}_m$ denote the standard bases in $\mathbb{R}^n, \mathbb{R}^m$ respectively. Since $L_{\mathcal{E}_m,\mathcal{E}_n} = A$, use the change of coordinates formula $L_{\mathcal{E}_m,\mathcal{E}_n} = I_{\mathcal{E}_m,\mathcal{G}} L_{\mathcal{G},\mathcal{F}} I_{\mathcal{F},\mathcal{E}_n}$ to write A as a product of three matrices.

Let $V_{n\times n} = \begin{pmatrix}\mathbf{v}_1 & \ldots & \mathbf{v}_n\end{pmatrix}$ be a matrix with columns $\mathbf{v}_i$ and define $U_{m\times m} = \begin{pmatrix}\mathbf{u}_1 & \ldots & \mathbf{u}_m\end{pmatrix}$ similarly. Then, $I_{\mathcal{E}_n,\mathcal{F}} = V$ and $I_{\mathcal{E}_n,\mathcal{G}} = U$. Since $\mathcal{F}$ is an orthonormal basis, we have $I_{\mathcal{F},\mathcal{E}_n} = I_{\mathcal{E}_n,\mathcal{F}}^T = V^T$ so that

$$A = L_{\mathcal{E}_m,\mathcal{E}_n} = I_{\mathcal{E}_m,\mathcal{G}} L_{\mathcal{G},\mathcal{F}} I_{\mathcal{F},\mathcal{E}_n} = U \left[\begin{array}{c|c} D & 0 \\ \hline 0 & 0 \end{array}\right]_{m\times n} V^T \qquad D = \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_r} \end{pmatrix}_{r\times r}$$

(5) **Congratulations!** You've just rediscovered Singular Value Decomposition (SVD). The positive values $\sigma_i = \sqrt{\lambda_i}$ are called **singular values** and we usually order the basis for $Col(A^T)$ so that the singular values will be arranged in non-increasing order $\sigma_1 \geq \ldots \geq \sigma_r$ within the diagonal block D. We often write $\Sigma = L_{\mathcal{G},\mathcal{F}}$ so that SVD is written $A = U\Sigma V^T$.

Is SVD unique, i.e. given A are there unique U, V such that $A = U\Sigma V^T$ with the diagonal entries of Σ in non-increasing order?

No. Recall that the choices of ONB for $Nul(A), Nul(A^T)$ are immaterial to the desired final diagonal form for $L_{\mathcal{G},\mathcal{F}}$.

Epilogue: We see that Singular Value Decomposition of real matrices is not magic but rather the nigh inexorable conclusion of a specific goal and specific approach:

- Goal: Find bases $\mathcal{F}, \mathcal{G}$ of $\mathbb{R}^n, \mathbb{R}^m$ respectively such that $L_{\mathcal{G},\mathcal{F}}$ is diagonal.
- Approach: choose appropriate ONB for each of the four fundamental subspaces to obtain the goal.

These observations support SVD's claim as a beautiful and significant result of linear algebra.