Christopher Zanoli
Student Number: 23-942-394

1. Let $S = \{v_1, \dots, v_p\}$ be a set of orthonormal non-zero vectors with $v_i \in \mathbb{R}^n\ \forall i$. Prove that S is an orthonormal basis of span(S).

2 pts ☐

We want to prove that $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ are linearly independent, they $\in$ *span*(S) and *dim*(*span*$(S)) = p$.
Since S is orthonormal:

$$\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 1 \quad \forall i \quad (1)$$
$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0 \quad \forall i \neq j \quad (2)$$

Let

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_p \mathbf{v}_p = 0,$$

then $\forall i$:

$$\begin{aligned} 0 &= \langle \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_p \mathbf{v}_p, \mathbf{v}_i \rangle && \text{by definition} \\ &= \alpha_1 \langle \mathbf{v}_1, \mathbf{v}_i \rangle + \alpha_2 \langle \mathbf{v}_2, \mathbf{v}_i \rangle + \cdots + \alpha_p \langle \mathbf{v}_p, \mathbf{v}_i \rangle && \text{by linearity of inner product} \\ &= \alpha_i \langle \mathbf{v}_i, \mathbf{v}_i \rangle && \text{due to (2)} \\ &= \alpha_i && \text{due to (1)} \end{aligned}$$

$\Rightarrow \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ are linearly independent. Furthermore, since $\{\mathbf{v}_1, \dots, \mathbf{v}_p\} \in S$, then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\} \in$ *span*(S) as well, since, by definition of *span*, *span*(S) is the set of all possible linear combination of vectors in S. Hence, S spans *span*(S) because any vector in *span*(S) can be written as a linear combination of vectors in S. Consequently, this means that *dim*(*span*$(S)) = p$, where p is the cardinality of S.

QED

2. Let $A \in \mathbb{R}^{m \times n}$ be a matrix of rank $r \leq n$ and $n \leq m$.

   Further, let $D = \{v_1, \dots v_n\}$ be an orthonormal basis of $\mathbb{R}^n$, where $\{v_1, \dots v_n\}$ are eigenvectors of $A^T A$ arranged such that their associated eigenvalues are decreasing $\lambda_1 \geq \ldots \geq \lambda_n$. Given A is of rank r, $A^T A$ is also of rank r and we have $\forall i \leq r, \lambda_i \neq 0$ and $\forall i > r, \lambda_i = 0$.

   Show the following

   (a) For all $i \in \{1, \dots n\}$ $\lambda_i \geq 0$. *Hint: what is* $\|Av_i\|^2$*?*

1 pts ☐

We need to show that $\lambda_i \geq 0 \ \forall i \in \{1, \dots, n\}$

$$\begin{aligned}
\|A\mathbf{v}_i\|^2 &\equiv \langle A\mathbf{v}_i, A\mathbf{v}_i\rangle && \text{by definition} \\
&= (A\mathbf{v}_i)^\top (A\mathbf{v}_i) && \text{assuming inner product as dot product} \\
&= \mathbf{v}_i^\top A^\top A\mathbf{v}_i && \text{property of transpose} \\
&= \langle A^\top A\mathbf{v}_i, \mathbf{v}_i\rangle && \text{by definition, inner product as dot product}
\end{aligned}$$

By definition, we know that $A^\top A\mathbf{v}_i = \lambda_i \mathbf{v}_i$.

$$\begin{aligned}
\Rightarrow \langle A^\top A\mathbf{v}_i, \mathbf{v}_i\rangle &= \langle \lambda_i \mathbf{v}_i, \mathbf{v}_i\rangle \\
&= \lambda_i \langle \mathbf{v}_i, \mathbf{v}_i\rangle \\
&= \lambda_i && \text{because } D \text{ is orthonormal}
\end{aligned}$$

Since by definition $\|A\mathbf{v}_i\|^2 \geq 0 \ \forall i \in \{1, \dots, n\}$ (being squared), and since we proved $\|A\mathbf{v}_i\|^2 = \lambda_i \ \forall i \in \{1, \dots, n\}$, it follows that $\lambda_i \geq 0 \ \forall i \in \{1, \dots, n\}$.

QED

(b) $\{Av_1, \dots, Av_r\}$, $1 \leq k \leq r$ is an orthogonal basis for Col(A). *Reminder: if $a_1, \cdots, a_n$ are the columns of A, Col(A)=span($\{a_1, \cdots, a_n\}$).*

3 pts

We need to show that:

i. $\{A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_r\}$ are orthogonal, and, as a consequence, linearly independent

ii. The vectors $\{A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_r\}$ must span $Col(A)$. This means that any vector in Col(A) can be expressed as a linear combination of $\{A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_r\}$

iii. $dim(Col(A)) = r$ where r is the cardinality of $\{A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_r\}$

For the first point, we want to show that $\{A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_r\}$ is a set of non-zero vectors and that $\langle A\mathbf{v}_i, A\mathbf{v}_j\rangle = 0 \ \forall i \neq j,\ i \leq r,\ j \leq r$.

From exercise 2(a) we have shown that $\forall i \in \{1, \dots, n\}$ $\|A\mathbf{v}_i\|^2 = \lambda_i$, and, from the problem statement, $\forall i \leq r \ \lambda_i \neq 0$. Hence, $\{A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_r\}$ is a set of non-zero vectors.

It suffices to prove orthogonality for $i, j \leq r$, as these correspond to the non-zero eigenvalues of $A^T A$ and thus the non-zero vectors $A\mathbf{v}_i$, which are relevant for the basis of $Col(A)$.

To establish this, we know that $\langle A\mathbf{v}_i, A\mathbf{v}_j\rangle = (A\mathbf{v}_j)^\top (A\mathbf{v}_i) = \mathbf{v}_j^\top A^\top A\mathbf{v}_i = \langle A^\top A\mathbf{v}_i, \mathbf{v}_j\rangle$.

But we know that $A^\top A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ implies $\langle A^\top A\mathbf{v}_i, \mathbf{v}_j\rangle = \langle \lambda_i \mathbf{v}_i, \mathbf{v}_j\rangle = \lambda_i \langle \mathbf{v}_i, \mathbf{v}_j\rangle$.

But since $\mathbf{v}_i$ and $\mathbf{v}_j$ are orthogonal (they are part of an orthonormal basis), we have $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = 0 \ \forall i \neq j$, which implies $\langle A\mathbf{v}_i, A\mathbf{v}_j\rangle = 0 \ \forall i \neq j$.

Since $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}$ are non-zero and orthogonal, they form a basis for a subspace of $Col(A)$. In order to show that they span $Col(A)$, we need to show that any vector $\mathbf{y}$ in $Col(A)$ can be expressed as a linear combination of $A\mathbf{v}_i$ $\forall i \in \{1, ..., r\}$.

Given that $\mathbf{y} \in Col(A)$ and $Col(A) = Im(A)$, $\forall \mathbf{y} \in Col(A)$ there exists an $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{y}$. Since $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthonormal basis (from problem statement), we can write $\mathbf{x}$ as the following linear combination of the orthonormal basis vectors $\mathbf{v}_i$:

$$\mathbf{x} = \sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i \qquad \Rightarrow \qquad A\mathbf{x} = A\left(\sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i\right) = \sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{v}_i \rangle A\mathbf{v}_i$$

However, for $i > r$, $A\mathbf{v}_i = \mathbf{0}$ (from exercise 2(a)), so we only need to consider $i \leq r$, which shows that $\mathbf{y}$ can be written as a linear combination of $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}$.

As a consequence, also the third point is proved since A is of rank r (there are r linearly independent columns of A) $\Rightarrow dim(Col(A)) = r$, and the cardinality of $\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_r\}$ is r (by definition).

QED

Next, we will prove the Singular Value Decomposition Theorem (SVD). We define the matrix $V = [v_1, \ldots, v_n]$ using D defined in Q2. We set $U = [u_1, \ldots, u_m]$ an orthonormal basis of $\mathbb{R}^m$ such that $\forall i, u_i = \frac{Av_i}{\|Av_i\|}$.

We define $\Sigma \in \mathbb{R}^{m \times n}, \Sigma = \begin{bmatrix} \Delta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ where $\Delta = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}$ with $\forall i \leq r, \sigma_i > 0$.

3. Prove that

   (a) Using $\sqrt{\lambda_i} = \sigma_i$, $AV = \left[\sigma_1 u_1, \ldots, \sigma_r u_r, 0, \ldots, 0\right]$

2 pts ☐

$$AV = A[\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n] = [A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_n] = [u_1\|A\mathbf{v}_1\|, \ldots, u_r\|A\mathbf{v}_r\|, 0, \ldots, 0]$$

Since:

- $\forall i \leq r$, $\mathbf{u}_i = \frac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|}$
- From exercise $2(a)$, $\lambda_i = 0$ $\forall i > r$. Since we proved $\|A\mathbf{v}_i\|^2 = \lambda_i$, then $\lambda_i = 0 \iff \|A\mathbf{v}_i\| = 0$ $\forall i > r$

Now, $\|A\mathbf{v}_i\|^2 = \lambda_i$ implies $\|A\mathbf{v}_i\| = \sqrt{\lambda_i}$ and $\sqrt{\lambda_i} = \sigma_i$ (by definition),

$$\Rightarrow [u_1\|A\mathbf{v}_1\|, \ldots, u_r\|A\mathbf{v}_r\|, 0, \ldots, 0] = [\sigma_1 u_1, \ldots, \sigma_r u_r, 0, \ldots, 0]$$

QED

(b) $U\Sigma = AV$, and

1 pts $\square$

We need to prove that

$$U\Sigma = AV = [\sigma_1 u_1, \dots, \sigma_r u_r, 0, \dots, 0]$$

where $U \in \mathbb{R}^{m\times m}$ and $\Sigma \in \mathbb{R}^{m\times n}$

$$U\Sigma = [u_1, u_2, \dots, u_m] \begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$

$$= \begin{bmatrix} u_{1,1} & u_{1,2} & \dots & u_{1,m} \\ u_{2,1} & u_{2,2} & \dots & u_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m,1} & u_{m,2} & \dots & u_{m,m} \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$

$$= [u_1\sigma_1, u_2\sigma_2, \dots, u_r\sigma_r, 0, \dots, 0] = AV$$

QED

(c) $A = U\Sigma V^T$.

1 pts $\square$

To prove that $A = U\Sigma V^\top$ we can start by observing that V has been defined as an orthogonal matrix since, from $2(a)$, $\{v_1, v_2, \dots, v_n\}$ are orthonormal vectors $\Rightarrow V^{-1} = V^\top$.
Hence, since in the previous point we showed that $U\Sigma = AV$:

$$\begin{aligned} U\Sigma V^{-1} &= AVV^{-1} && \text{by multiplying each right hand side by } V^{-1} \\ U\Sigma V^\top &= AI && \text{where } I \text{ is the identity matrix} \\ U\Sigma V^\top &= A \end{aligned}$$

QED

We have shown that any matrix $A \in \mathbb{R}^{m\times n}$ of rank r can be decomposed into $U\Sigma V^T$ where U, V are orthogonal matrices and $\Sigma = \begin{bmatrix} \Delta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ is a diagonal matrix containing the singular values of A.