Ex 1

$\text{rank}(A) = \dim(\mathcal{R}(A)) \Rightarrow$ the dimension of $\mathcal{R}(A)$ is less than n. Unpacking more,

$$\mathcal{R}(A) = \{ y = Ax : x \in \mathbb{R}^n \}$$
$$= \text{span}(\{a_1, \dots, a_n\})$$

So the range of A is what can be built from the columns of A. Hence, $\text{rank}(A) < n$ means the columns of A are linearly dependent. It also means that there are vectors in $\mathbb{R}^m$ that cannot be built from the columns of A, even when $m = n$.

Ex 2

Following pg. 3.24, we let $\underset{p \times n}{B^T = \begin{bmatrix} | & & | \\ b_1 & \cdots & b_n \\ | & & | \end{bmatrix}}$ and $\underset{n \times m}{A^T = \begin{bmatrix} - a_1^T - \\ \vdots \\ - a_n^T - \end{bmatrix}}$

so that

$$B^T A^T = \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix} \begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} = \sum_{i=1}^{n} b_i a_i^T$$

Examining the j^{th} column, we get

$$(B^T A^T)_{:j} = \sum_{i=1}^{n} \underbrace{b_i}_{\text{vector}} \underbrace{a_{ij}}_{\text{scalar}},$$ which is a linear combination of the columns of $B^T \Leftrightarrow$ of the rows of B.

Ex 3

From pg. 3.24, we know that for $K=2$, $\text{rank}(A_1 A_2) \le \min(\text{rank}(A_1), \text{rank}(A_2))$. This can serve as the base case for an inductive proof. Now assume

$$\text{rank}\left(\prod_{k=1}^{K-1} A_k\right) \le \min \{\text{rank}(A_k)\}_{k=1}^{K-1}. \quad \text{(inductive hypothesis)}$$

Let $B = \prod_{k=1}^{K-1} A_k$. Then

$$\text{rank}(BA_K) \le \min(\text{rank}(B), \text{rank}(A_K)) \quad \text{(by base case)}$$

$$\le \min\left(\{\text{rank}(A_k)\}_{k=1}^{K-1}, \text{rank}(A_K)\right)$$

$$= \min \{\text{rank}(A_k)\}_{k=1}^{K}.$$

Ex 4

Since x, y are vectors, $\text{rank}(x) = 1$ and $\text{rank}(y) = 1 = \text{rank}(y^T)$. Therefore

$$\text{rank}(xy^T) \le \text{rank}(x)\,\text{rank}(y^T) = 1.$$

When $x^T y = 0$, we still have $\text{rank}(xy^T) = 1$ even though all eigenvalues are 0, as we showed in HW1.

Ex 5

Unitary matrices are "like" rotation matrices, so they can be interpreted as a change of basis for the vectors/subspace. Changing the axes that define a vector does not change what vectors can be built.

Ex 6

We'll show a stronger statement, which is that $\mathcal{R}(AQ) = \mathcal{R}(A)$. Since these are two sets, to show equality we show that $\mathcal{R}(AQ) \subset \mathcal{R}(A)$ and $\mathcal{R}(A) \subset \mathcal{R}(AQ)$. First take $y \in \mathcal{R}(AQ)$. We want to show $y \in \mathcal{R}(A)$ to prove $\mathcal{R}(AQ) \subset \mathcal{R}(A)$.

$$y \in \mathcal{R}(AQ) \Rightarrow y = AQx \quad \text{for some } x$$
$$= Av \quad \text{for } v = Qx$$

but $Av \in \mathcal{R}(A)$ by definition. Next show $\mathcal{R}(A) \subset \mathcal{R}(AQ)$ by taking an arbitrary $y \in \mathcal{R}(A)$. Then

$$y = Ax$$
$$= AQQ^Tx \qquad \text{since } QQ^T = I$$
$$= AQw \qquad \text{for } w = Q^Tx$$

but $AQw \in \mathcal{R}(AQ)$ by definition, which completes the proof.

Ex 7

$\mathcal{R}^{\perp}(A) = \{ y \in \mathbb{R}^m : y \neq Ax \text{ for any } x \in \mathbb{R}^n \}$ so y has size m.

$\mathcal{N}(A) = \{ x \in \mathbb{R}^n : Ax = 0 \}$ so x has size n.

We can only compute $x^T y$ if $m = n$ (i.e., if A is square), but even if this is the case, we cannot say anything about $x^T y$ in general.

Ex 8

Recall $S^{\perp} = \{ v \in V : \langle s, v \rangle = 0 \ \forall s \in S \}$. Thus

$$\mathcal{N}^{\perp}(A) = \{ v \in \mathbb{R}^n : v^T x = 0 \ \forall x \in \mathcal{N}(A) \}$$

which implies $x_0^T x_1 = 0$. Similarly, $y_0^T y_1 = 0$.

Ex 9

$\underline{\mathcal{R}^{\perp}(A) \subset \mathcal{N}(A^T)}$: Take $x \in \mathcal{R}^{\perp}(A)$. Then $x^T A y = 0 \ \forall y$, and $(A^T x)^T y = 0 \ \forall y$, which can only hold if $A^T x = 0$, so $x \in \mathcal{N}(A^T)$.

$\underline{\mathcal{N}(A^T) \subset \mathcal{R}^{\perp}(A)}$: Take $x \in \mathcal{N}(A^T)$, so that $A^T x = 0$. Then $y^T (A^T x) = 0$ for all y, which implies

$$(y^T A^T) x = 0 \iff (Ay)^T x = 0 \quad \forall y$$

So x is orthogonal to anything in $\mathcal{R}(A)$, meaning $x \in \mathcal{R}^{\perp}(A)$.

Ex 10

Recall the four fundamental subspaces are $\mathcal{R}(A)$, $\mathcal{R}^\perp(A)$, $\mathcal{N}(A)$, and $\mathcal{N}^\perp(A)$.

Let $\operatorname{rank}(A) = r$. Then the SVD is broken down as

$$A = U\Sigma V^T = \begin{bmatrix} U_r & U_0 \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^T \\ V_0^T \end{bmatrix}$$

where

U_r = basis for $\mathcal{R}(A)$ $\quad$ V_r = basis for $\mathcal{N}^\perp(A)$

U_0 = basis for $\mathcal{R}^\perp(A)$ $\quad$ V_0 = basis for $\mathcal{N}(A)$

Ex 11

For $\operatorname{range}(A) = \mathbb{R}^m$, A must have m linearly independent columns. This implies $n \geq m$. If $n > m$, some columns of A are linearly dependent.

Ex 12

For the full SVD, the matrices U and V are square and orthogonal, while Σ is diagonal. For the thin SVD, U_r and V_r are tall with orthonormal columns (but <u>not</u> orthonormal rows in general), and Σ_r is square. Both are always valid ways to write a matrix.

Ex 13

Let A be size $m \times n$. Similar to the above, we get

$$UU^T = U^T U = I_m$$

$$U_r^T U_r = I_r$$

$$U_r U_r^T = ??$$ (projection matrix onto $\mathcal{R}(A)$ - discussed next week)

Ex 14

$$\|U_r U_r^T x\|_2^2 = (U_r U_r^T x)^T (U_r U_r^T x)$$

$$= x^T U_r \underbrace{U_r^T U_r}_{I_r} U_r^T x$$

$$= x^T U_r U_r^T x$$

$$= \|U_r^T x\|_2^2$$

So $\|U_r^T x\| = \|U_r U_r^T x\|$. Unfortunately, $\|U_r^T x\| \neq \|x\|$ in general.

Ex 15

Following 3.42, $z^T z$ is the only nonzero eigenvalue of A and has corresponding eigenvector z. Therefore, normalizing z, we get

$$A = \lambda_1 q_1 q_1^T = (z^T z) \frac{z}{\|z\|} \frac{z^T}{\|z\|}$$

$$= \|z\|^2 \frac{z z^T}{\|z\|^2} = z z^T$$ ← this line is a sanity check

Ex 16

Check: All eigenvectors and singular vectors have unit norm, singular values are nonnegative, and $A = Q \Lambda Q^T$ or $A = U \Sigma V^T$. Do on your own :)

Ex 17

$$y = Ax = U \Sigma V^T x = U_r \Sigma_r V_r^T x = U_r \left(\Sigma_r V_r^T x \right)$$

We can either think of the coordinates as elements of the vector $\Sigma_r V_r^T x$ or as elements of $V_r^T x$ with gains defined by Σ_r. The latter is a more intuitive description. See pg. 3.45.