Let $\underline{x}$ be the ~~unit~~ (normalized) eigenvector ~~associ~~ of $A^H A$ associated with the largest eigenvalue. Let

$$\underline{A}\underline{x} = \sigma \underline{y}$$

where $\|\underline{y}\|_2 = 1$, and $\sigma \geq 0$. It can be verified that $\|A\|_2 = \sigma$.

We can construct two unitary matrices $\underline{U}$ & $\underline{V}$ so that $\underline{x}$ and y form their first columns: (recall one of the lemmas in the proof of Schur Δ theorem?)

$$U = [\, y \,,\ U_1 \,]$$
$$V = [\, x \,,\ V_1 \,]$$

Let

$$A_1 = \underline{U}^H \underline{A}\, \underline{V}$$

The matrix A_1 can be decomposed as

$$\begin{bmatrix} y^H \\ \underline{U}_1^H \end{bmatrix} A\, [\, x \ \ \underline{V}_1 \,] = \begin{bmatrix} y^H \\ \underline{U}_1^H \end{bmatrix} [\, \sigma y \ \ \underline{A}\underline{V}_1 \,]$$

$$= \begin{bmatrix} \sigma y^H y & y^H \underline{A}\, \underline{V}_1 \\ \sigma \underline{U}_1^H y & \underline{U}_1 \underline{A}\, \underline{V}_1 \end{bmatrix}$$

$$\overset{\#}{=} \begin{bmatrix} \sigma & y^H \underline{A}\, \underline{V}_1 \\ \underline{0} & U_1 \underline{A}\, \underline{V}_1 \end{bmatrix}$$

$$\triangleq \begin{bmatrix} \sigma & w^H \\ \underline{0} & \underline{B} \end{bmatrix}$$

We now show that $\underline{w} = \underline{0}$, thereby $\underline{A}_1 = \begin{bmatrix} \sigma & \underline{0} \\ \underline{0} & \underline{B} \end{bmatrix}$.

We note that

$$\sigma^2 = \|\underline{A}\|_2^2 = \|\underline{U}^H \underline{A}\, \underline{V}\|_2^2$$
$$= \|\underline{A}_1\|_2^2 .$$

By the definition of the 2-norm,

$$\|\underline{A}_1\|_2^2 \geq \frac{\|\underline{A}_1 \underline{z}\|_2^2}{\|\underline{z}\|_2^2}, \quad \text{for any } \underline{z} \in \mathbb{C}^n$$

Putting $\underline{z} = \begin{bmatrix} \sigma \\ \underline{w} \end{bmatrix}$ into the above eqn.,

$$\sigma^2 = \|\underline{A}_1\|_2^2 \geq \left\| \begin{bmatrix} \sigma & \underline{w}^H \\ \underline{0} & \underline{B} \end{bmatrix} \begin{bmatrix} \sigma \\ \underline{w} \end{bmatrix} \right\|_2^2 \Big/ (\sigma^2 + \underline{w}^H \underline{w})$$

$$= \left\| \begin{bmatrix} \sigma^2 + \underline{w}^H \underline{w} \\ \underline{B}\underline{w} \end{bmatrix} \right\|_2^2 \Big/ (\sigma^2 + \underline{w}^H \underline{w})$$

$$= \left\{ (\sigma^2 + \underline{w}^H \underline{w})^2 + \|\underline{B}\underline{w}\|_2^2 \right\} \Big/ (\sigma^2 + \underline{w}^H \underline{w})$$

$$\geq (\sigma^2 + \underline{w}^H \underline{w})^2 \Big/ (\sigma^2 + \underline{w}^H \underline{w}) = \sigma^2 + \underline{w}^H \underline{w}$$

This implies that

$$\|\underline{w}\|_2^2 = \underline{w}^H \underline{w} \leq 0$$

which is possible only when $\underline{w} = \underline{0}$.

We now have

$$\underline{A}_1 = \begin{bmatrix} \sigma & \underline{0} \\ \underline{0} & \underline{B} \end{bmatrix}$$

The whole process repeats using only $\underline{B}$, until $\underline{A}_n$ becomes diagonal.

Property 4.3: Recall $\mathcal{R}(A) = \{ y \in \mathbb{C}^m \mid y = Ax,\ x \in \mathbb{C}^n \}$.

$$y = U_1 \tilde{\Sigma} V_1^H x$$
$$= U_1 \tilde{\Sigma} z \quad , \quad z\ \cancel{x} = V_1^H x$$

Hence,

$$\mathcal{R}(A) = \{ y \in \mathbb{C}^m \mid y = U_1 \tilde{\Sigma} z,\ z \in \mathcal{R}(V_1^H) \}$$

Since $\dim \mathcal{R}(V_1^H) = \operatorname{rank}(V_1^H) = r$ (bear in mind that the columns of V_1 form an orthonormal set of vectors),

$$\mathcal{R}(A) = \{ y \in \mathbb{C}^m \mid y = U_1 \tilde{\Sigma} z,\ z \in \cancel{\mathcal{R}(}\ \mathbb{C}^r \}$$

Since $\sigma_i \neq 0$ for $i = 1, \dots, r$,

$$\mathcal{R}(A) = \mathcal{R}(U_1).$$

Property 4.1: $\operatorname{rank}(A) = \dim \mathcal{R}(A) = \dim \mathcal{R}(U_1) = r$.

Property 4.2: Recall $N(A) = \{ x \in \mathbb{C}^n \mid Ax = 0 \}$

Let

$$x = V c$$
$$= [\, V_1 \mid V_2 \,] \begin{array}{c} \left[\begin{array}{c} c_1 \\ \hline c_2 \end{array} \right] \end{array} \begin{array}{l} \updownarrow r \\ \updownarrow n-r \end{array}$$
$$= V_1 c_1 + V_2 c_2.$$

Since $\mathcal{R}(V) = \mathbb{C}^r$,

$$N(A) = \{ x \in \mathbb{C}^r \mid x = V_1 c_1 + V_2 c_2,\ Ax = 0,\ c_1 \in \mathbb{C}^r,\ c_2 \in \mathbb{C}^{n-r} \}$$

Now,

$$Ax = 0 \iff U_1 \tilde{\Sigma} V_1^H (V_1 c_1 + V_2 c_2) = 0$$
$$\iff U_1 \tilde{\Sigma} c_1 = 0 \iff c_1 = 0.$$

Therefore,

$$\mathcal{N}(\underline{A}) = \{ \underline{x} \in \mathbb{C}^n \mid \underline{x} = \underline{V}_2 \underline{c}_2 ,\ \underline{c}_2 \in \mathbb{C}^{n-r} \}$$
$$= \mathcal{R}(\underline{V}_2).$$

Property 4.4: $\underline{A}^H = \underline{V} \underline{\Sigma}^T \underline{U}^H = \underline{V}_1 \tilde{\underline{\Sigma}} \underline{U}_1^H$. Following the proof for Property 4.3, we can show that $\mathcal{R}(\underline{A}^H) = \mathcal{R}(\underline{V}_1)$.

Property 4.5: Recall $\mathcal{R}_\perp(\underline{A}) = \{ \underline{y} \in \mathbb{C}^m \mid \underline{y}^H \underline{z} = \underline{0} ,\ \forall\, \underline{z} \in \mathcal{R}(\underline{A}) \}$.
From Property 4.3, $\mathcal{R}(\underline{A}) = \mathcal{R}(\underline{U}_1)$. In other words we can represent any $\underline{z} \in \mathcal{R}(\underline{U}_1)$ by

$$\underline{z} = \underline{U}_1 \underline{d}_1 ,\quad \underline{d}_1 \in \mathbb{C}^r \qquad (*)$$

Moreover, any $\underline{y} \in \mathbb{C}^m$ can be represented by

$$\underline{y} = \underline{U}_1 \underline{c}_1 + \underline{U}_2 \underline{c}_2 ,\quad \underline{c}_1 \in \mathbb{C}^r,\quad \underline{c}_2 \in \mathbb{C}^{m-r} \qquad (**)$$

From (*) and (**), the condition $\underline{y}^H \underline{z} = \underline{0}$ is equiv. to

$$\underline{0} = \underline{y}^H \underline{z} = (\underline{U}_1 \underline{c}_1 + \underline{U}_2 \underline{c}_2)^H \underline{U}_1 \underline{d}_1$$
$$= \underline{c}_1^H \underline{d}_1$$

To have $\underline{y}^H \underline{z} = \underline{0}$ $\forall$ $\underline{z} \in \mathcal{R}(\underline{U}_1)$, we need $\underline{c}_1 = \underline{0}$. Moreover $\underline{c}_2$ does not affect the condition $\underline{y}^H \underline{z} = \underline{0}$. Hence,

$$\mathcal{R}_\perp(\underline{A}) = \{ \underline{y} \in \mathbb{C}^m \mid \underline{y} = \underline{U}_2 \underline{c}_2 ,\ \underline{c}_2 \in \mathbb{C}^{m-r} \}$$
$$= \mathcal{R}(\underline{U}_2).$$

Proof of Theorem 4.2

We prove this theorem by showing that $\|\underline{A} - \underline{B}\|_2 \geq \|\underline{A} - \underline{A}_k\|_2$ for all $\underline{B}$ with $\operatorname{rank}(\underline{B}) = k$. From Property 4.2, we ~~obtain~~ have that $\dim N(\underline{B}) = n-k$. Hence, fixing an $\underline{B}$, $N(\underline{B})$ can be constructed by $R(\underline{X})$ for some full rank $\underline{X} \in \mathbb{C}^{n\times(n-k)}$. Let $\{\underline{v}_1, \cdots, \underline{v}_r\}$ be the right singular vectors corresponding to the nonzero singular values, denoted by $\{\sigma_1, \cdots, \sigma_r\}$. A dimension argument suggest that (proof omitted here)

$$R(\underline{X}) \cap R([\underline{v}_1, \cdots, \underline{v}_{k+1}]) \neq \{\underline{0}\}$$

Let $\underline{z}$ be a unit 2-norm vector lying in $R(\underline{X}) \cap R([\underline{v}_1, \cdots, \underline{v}_{k+1}])$. Then,

$$
\begin{aligned}
\|\underline{A} - \underline{B}\|_2^2 &\geq \|(\underline{A} - \underline{B})\underline{z}\|_2^2 \\
&= \|\underline{A}\underline{z}\|_2^2 \\
&= \underline{z}^H \underline{A}^H \underline{A}\, \underline{z} \\
&\cancel{= \underline{z}^H \underline{V}\, \underline{\Lambda}\, \underline{V}^H \underline{z}} \\
&= \underline{z}^H \underline{V}_1 \tilde{\underline{\Sigma}}^2 \underline{V}_1^H \underline{z} \qquad \left(\begin{aligned} \underline{V}_1 &= [\underline{v}_1, \cdots, \underline{v}_r] \\ \tilde{\Sigma} &= \operatorname{Diag}(\sigma_1, \cdots, \sigma_r) \end{aligned} \right)
\end{aligned}
$$

Since $\underline{V}_1^H \underline{z} = [\,(\underline{v}_1^H \underline{z})\ (\underline{v}_2^H \underline{z}) \cdots (\underline{v}_{k+1}^H \underline{z})\ 0, \cdots, 0\,]^T$,

$$\|\underline{A} - \underline{B}\|_2^2 \geq \sum_{i=1}^{k+1} \sigma_i^2 (\underline{v}_i^H \underline{z})^2$$

$$\geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} (\underline{v}_i^H \underline{z})^2$$

$$= \sigma_{k+1}^2$$

(why $\sum_{i=1}^{k+1} (\underline{v}_i^H \underline{z})^2 = 1$? This is left as an exercise.)

Since $\|\underline{A} - \underline{A}_k\|_2 = \left\| U \begin{bmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \sigma_{k+1} & & \\ & & & & \ddots & \\ & & & & & \sigma_p \end{bmatrix} V^H \right\|_2$

$= \sigma_{k+1}$,

we have

$$\|\underline{A} - \underline{B}\|_2 \geq \|\underline{A} - \underline{A}_k\|_2$$

for any $\underline{B}$, $\text{rank}(\underline{B}) = k$.

The proof of Theorem 4.3 is omitted, since it involves some matrix concepts that we haven't gone through at this point.

Theorem 4.4

Suppose that there is full column rank $\mathbf{Y} \in \mathbb{C}^{\cancel{m\times m}\, n\times m}$, $\mathbf{Y} \neq \mathbf{X}$, such that $\cancel{\mathcal{R}(\mathbf{X})} = \mathcal{R}(\mathbf{Y})$. Let

$$\mathbf{P}_1 = \mathbf{Y}(\mathbf{Y}^H\mathbf{Y})^{-1}\mathbf{Y}^H.$$

Our goal is to show that $\mathbf{P}_1 = \mathbf{P}$. Since $\mathcal{R}(\mathbf{X}) = \mathcal{R}(\mathbf{Y})$, there exists a matrix $\mathbf{C} \in \mathbb{C}^{m\times m}$ such that

$$\mathbf{X} = \mathbf{Y}\mathbf{C}$$

The matrix $\mathbf{C}$ must be non-singular — if $\mathbf{C}$ were singular with rank $r<m$ then $\text{rank}(\mathbf{X}) \leq \cancel{\text{rank}(\mathbf{Y})\,\text{rank}(\mathbf{C})}\ \min(\text{rank}(\mathbf{B}), \text{rank}(\mathbf{C})) = r$ which contradicts the fact that $\text{rank}(\mathbf{X}) = m$.

$$\begin{aligned}
\mathbf{P}_1 &= \mathbf{X}\mathbf{C}^{-1}(\mathbf{C}^{-H}\mathbf{X}^H\mathbf{X}\mathbf{C}^{-1})^{-1}\mathbf{C}^{-H}\mathbf{X}^H \\
&= \mathbf{X}\mathbf{C}^{-1}\mathbf{C}(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{C}^H\mathbf{C}^{-H}\mathbf{X}^H \\
&= \mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H = \mathbf{P}.
\end{aligned}$$

Property 4.6 follows directly from Properties 4.1 – 4.5.

Property 4.7. ~~By notice~~ From Property 4.6,

$$\begin{aligned}
\mathbf{P} &= \mathbf{U}_1\mathbf{U}_1^H \\
&= [\mathbf{U}_1\ \mathbf{U}_2]\begin{bmatrix} \mathbf{I}_r & \\ & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{U}_1^H \\ \mathbf{U}_2^H \end{bmatrix}.
\end{aligned}$$

which is an eigendecomposition.

Theorem 4.5

$$\operatorname{dist}(S_1,S_2) = \| \underline{W}_1\underline{W}_1^H - \underline{Z}_1\underline{Z}_1^H \|_2$$

$$= \| \underline{W}^H (\underline{W}_1\underline{W}_1^H - \underline{Z}_1\underline{Z}_1^H) \underline{Z} \|_2$$

$$= \left\| \begin{bmatrix} \underline{W}_1^H (\underline{W}_1\underline{W}_1^H - \underline{Z}_1\underline{Z}_1^H) \underline{Z}_1 & \underline{W}_1^H (\underline{W}_1\underline{W}_1^H - \underline{Z}_1\underline{Z}_1^H) \underline{Z}_2 \\ \underline{W}_2^H (\underline{W}_1\underline{W}_1^H - \underline{Z}_1\underline{Z}_1^H) \underline{Z}_1 & \underline{W}_2^H (\underline{W}_1\underline{W}_1^H - \underline{Z}_1\underline{Z}_1^H) \underline{Z}_2 \end{bmatrix} \right\|_2$$

$$= \left\| \begin{bmatrix} \underline{W}_1^H\underline{Z}_1 - \underline{W}_1^H\underline{Z}_1 & \underline{W}_1^H\underline{Z}_2 \\ -\underline{W}_2^H\underline{Z}_1 & \underline{0} \end{bmatrix} \right\|_2$$

$$= \left\| \begin{bmatrix} \underline{0} & \underline{W}_1^H\underline{Z}_2 \\ -\underline{W}_2^H\underline{Z}_1 & \underline{0} \end{bmatrix} \right\|_2$$

Let

$$\underline{Q} = \underline{W}^H\underline{Z}$$

$$= \begin{bmatrix} \underline{W}_1^H\underline{Z}_1 & \underline{W}_1^H\underline{Z}_2 \\ \underline{W}_2^H\underline{Z}_1 & \underline{W}_2^H\underline{Z}_2 \end{bmatrix} = \begin{bmatrix} \underline{Q}_{11} & \underline{Q}_{12} \\ \underline{Q}_{21} & \underline{Q}_{22} \end{bmatrix}.$$

Let $\underline{x}$ be a unit 2-norm vector.

$$1 = \left\| \underline{Q} \begin{bmatrix} \underline{x} \\ \underline{0} \end{bmatrix} \right\|_2^2 = \left\| \begin{bmatrix} \underline{Q}_{11}\underline{x} \\ \underline{Q}_{21}\underline{x} \end{bmatrix} \right\|_2^2$$

$$= \| \underline{Q}_{11}\underline{x} \|_2^2 + \| \underline{Q}_{21}\underline{x} \|_2^2$$

Therefore,

$$\| \underline{Q}_{21} \|_2^2 = \max_{\|\underline{x}\|_2=1} \| \underline{Q}_{21}\underline{x} \|_2^2$$

$$= 1 - \min_{\|\underline{x}\|_2=1} \| \underline{Q}_{11}\underline{x} \|_2^2$$

$$= 1 - \sigma_{\min}^2(\underline{Q}_{11}).$$

where $\sigma_{\min}(\underline{A})$ stands for the min. singular value of $\underline{A}$.

~~Likewise~~. By working with $\underline{Q}^H$ in a similar fashion, we obtain

$$\|\underline{Q}_{12}^H\|_2^2 = 1 - \sigma_{min}^2(\underline{Q}_{11}^H).$$

$$\Rightarrow \quad \|\underline{Q}_{12}\|_2^2 = 1 - \sigma_{min}^2(\underline{Q}_{11}).$$

Hence, we have $\|\underline{Q}_{12}\|_2^2 = \|\underline{Q}_{21}\|_2^2$. As a consequence,

$$\text{dist}(S_1, S_2)^2 = \left\| \begin{bmatrix} \underline{0} & \underline{Q}_{12} \\ -\underline{Q}_{21} & \underline{0} \end{bmatrix} \right\|_2^2$$

$$= \max_{\|\underline{x}\|_2^2 + \|\underline{y}\|_2^2 = 1} \left\| \begin{bmatrix} \underline{0} & \underline{Q}_{12} \\ -\underline{Q}_{21} & \underline{0} \end{bmatrix} \begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix} \right\|_2^2$$

$$= \max_{\|\underline{x}\|_2^2 + \|\underline{y}\|_2^2 = 1} \|\underline{Q}_{12}\underline{x}\|_2^2 + \|\underline{Q}_{21}\underline{y}\|_2^2$$

$$\leq \max_{\|\underline{x}\|_2^2 + \|\underline{y}\|_2^2 = 1} \|\underline{Q}_{12}\|_2^2 \left(\|\underline{x}\|_2^2 + \|\underline{y}\|_2^2\right) = \|\underline{Q}_{12}\|_2^2.$$

It can be easily verified that the inequality is achievable by some $\underline{x}, \underline{y}$ (this is left as an exercise).

Properties 4.8 – 4.10 are left as your exercise.