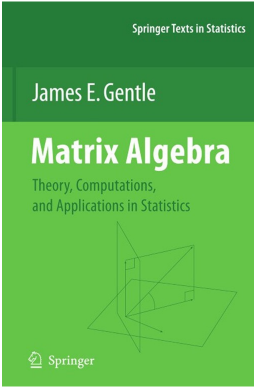
# Theory of Matrices

**Chapter 3. Basic Properties of Matrices**

3.8. Eigenvalues; Canonical Factorizations—Proofs of Theorems



# Table of contents

# Theorem 3.8.1

**Theorem 3.8.1.** If $v$ is an eigenvector of $A$ and $w$ is a left eigenvector of $A$ with a different associated eigenvalue, then $v \perp w$.

**Proof.** Let $Av = c_1 v$ and $w^T A = c_2 w^T$ where $c_1 \neq c_2$. Then $(w^T A)v = c_2 w^T v$ and $w^T(Av) = w^T(c_1 v) = c_1 w^T v$ so $c_1 w^T v = c_2 w^T v$, but since $c_1 \neq c_2$ it must be that $w^T v = \langle w, v \rangle = 0$ and $v \perp w$. □

# Theorem 3.8.1

**Theorem 3.8.1.** If $v$ is an eigenvector of $A$ and $w$ is a left eigenvector of $A$ with a different associated eigenvalue, then $v \perp w$.

**Proof.** Let $Av = c_1 v$ and $w^T A = c_2 w^T$ where $c_1 \neq c_2$. Then $(w^T A)v = c_2 w^T v$ and $w^T(Av) = w^T(c_1 v) = c_1 w^T v$ so $c_1 w^T v = c_2 w^T v$, but since $c_1 \neq c_2$ it must be that $w^T v = \langle w, v \rangle = 0$ and $v \perp w$. □

# Corollary 3.8.3

**Corollary 3.8.3.** The set of eigenvectors of a $n \times n$ matrix $A$ associated with given eigenvalue $c$, along with the 0 vector, form a subspace of $\mathbb{C}^n$ (or of $\mathbb{R}^n$ if we restrict ourselves to real numbers). The subspace is the *eigenspace* of $A$ associated with eigenvalue $c$.

**Proof.** By the definition of vector space of $n$ vectors from $\mathbb{R}^n$ (which also holds for $\mathbb{C}^n$; in fact it holds for $\mathbb{F}^n$ where $\mathbb{F}$ is any field) in Section 2.1, we need only show that for any scalars $a$ and $b$ and any eigenvectors $v_1$ and $v_2$, we have $av_1 + bv_2$ is either an eigenvector of $A$ with associated eigenvalue $c$ or is the 0 vector.

# Corollary 3.8.3

**Corollary 3.8.3.** The set of eigenvectors of a $n \times n$ matrix $A$ associated with given eigenvalue $c$, along with the 0 vector, form a subspace of $\mathbb{C}^n$ (or of $\mathbb{R}^n$ if we restrict ourselves to real numbers). The subspace is the *eigenspace* of $A$ associated with eigenvalue $c$.

**Proof.** By the definition of vector space of $n$ vectors from $\mathbb{R}^n$ (which also holds for $\mathbb{C}^n$; in fact it holds for $\mathbb{F}^n$ where $\mathbb{F}$ is any field) in Section 2.1, we need only show that for any scalars $a$ and $b$ and any eigenvectors $v_1$ and $v_2$, we have $av_1 + bv_2$ is either an eigenvector of $A$ with associated eigenvalue $c$ or is the 0 vector. We have

$$\begin{aligned} A(av_1 + bv_2) &= A(av_1) + A(bv_2) \\ &= aA(v_1) + bA(v_2) \\ &= a(cv_1) + b(cv_2) \text{ since } v_1 \text{ and } v_2 \text{ are} \\ &\quad\ \text{eigenvectors with eigenvalue } c \\ &= c(av_1 + bv_2). \end{aligned}$$

# Corollary 3.8.3

**Corollary 3.8.3.** The set of eigenvectors of a $n \times n$ matrix $A$ associated with given eigenvalue $c$, along with the 0 vector, form a subspace of $\mathbb{C}^n$ (or of $\mathbb{R}^n$ if we restrict ourselves to real numbers). The subspace is the *eigenspace* of $A$ associated with eigenvalue $c$.

**Proof.** By the definition of vector space of $n$ vectors from $\mathbb{R}^n$ (which also holds for $\mathbb{C}^n$; in fact it holds for $\mathbb{F}^n$ where $\mathbb{F}$ is any field) in Section 2.1, we need only show that for any scalars $a$ and $b$ and any eigenvectors $v_1$ and $v_2$, we have $av_1 + bv_2$ is either an eigenvector of $A$ with associated eigenvalue $c$ or is the 0 vector. We have

$$\begin{aligned} A(av_1 + bv_2) &= A(av_1) + A(bv_2) \\ &= aA(v_1) + bA(v_2) \\ &= a(cv_1) + b(cv_2) \text{ since } v_1 \text{ and } v_2 \text{ are} \\ &\quad \text{ eigenvectors with eigenvalue } c \\ &= c(av_1 + bv_2). \end{aligned}$$

# Corollary 3.8.3 (continued)

**Corollary 3.8.3.** The set of eigenvectors of a $n \times n$ matrix $A$ associated with given eigenvalue $c$, along with the 0 vector, form a subspace of $\mathbb{C}^n$ (or of $\mathbb{R}^n$ if we restrict ourselves to real numbers). The subspace is the *eigenspace* of $A$ associated with eigenvalue $c$.

**Proof (continued).** So $av_1 + bv_2$ is either the 0 vector in $\mathbb{C}^n$ or an eigenvector of $A$ with associated eigenvalue $c$. That is, the eigenvector associated with eigenvalue $c$ along with the 0 vector is a subspace of $\mathbb{C}^n$. □

# Theorem 3.8.4

**Theorem 3.8.4. The Cayley-Hamilton Theorem.**
For $n \times n$ matrix $A$ with characteristic polynomial $p_A$ we have $p_A(A) = 0$.

**Proof.** By Theorem 3.1.3, $(A - c\mathcal{I}_n)\text{adj}(A - c\mathcal{I}_n) = p_A(c)\mathcal{I}_n$. Since $p_A(c)$ is a polynomial of degree $n$, then $p_A(c) = s_0 + s_1 c + s_2 c^2 + \cdots + s_n c^n$ for some $s_0, s_1, \ldots, s_n$. Then $p_A(c)\mathcal{I}_n = p_A(c\mathcal{I}_n) = (A - c\mathcal{I}_n)\text{adj}(A - c\mathcal{I}_n)$, and so $\text{adj}(A - c\mathcal{I}_n)$ must be some $n-1$ degree polynomial with $n \times n$ matrix coefficients, say $B_0, B_1, \ldots, B_{n-1}$:

$$\text{adj}(A - c\mathcal{I}_n) = B_0 + B_1 c + B_2 c^2 + \cdots + B_{n-1} c^{n-1}.$$

So

$$(A - c\mathcal{I}_n)(B_0 + B_1 c + B_2 c^2 + \cdots + B_{n-1} c^{n-1}) = (s_0 + s_1 c + s_2 c^2 + \cdots + s_n c^n)\mathcal{I}_n$$

or . . .

# Theorem 3.8.4

**Theorem 3.8.4. The Cayley-Hamilton Theorem.**
For $n \times n$ matrix $A$ with characteristic polynomial $p_A$ we have $p_A(A) = 0$.

**Proof.** By Theorem 3.1.3, $(A - c\mathcal{I}_n)\text{adj}(A - c\mathcal{I}_n) = p_A(c)\mathcal{I}_n$. Since $p_A(c)$ is a polynomial of degree $n$, then $p_A(c) = s_0 + s_1 c + s_2 c^2 + \cdots + s_n c^n$ for some $s_0, s_1, \ldots, s_n$. Then $p_A(c)\mathcal{I}_n = p_A(c\mathcal{I}_n) = (A - c\mathcal{I}_n)\text{adj}(A - c\mathcal{I}_n)$, and so $\text{adj}(A - c\mathcal{I}_n)$ must be some $n - 1$ degree polynomial with $n \times n$ matrix coefficients, say $B_0, B_1, \ldots, B_{n-1}$:

$$\text{adj}(A - c\mathcal{I}_n) = B_0 + B_1 c + B_2 c^2 + \cdots + B_{n-1} c^{n-1}.$$

So

$$(A - c\mathcal{I}_n)(B_0 + B_1 c + B_2 c^2 + \cdots + B_{n-1} c^{n-1}) = (s_0 + s_1 c + s_2 c^2 + \cdots + s_n c^n)\mathcal{I}_n$$

or ...

# Theorem 3.8.4 (continued)

**Proof (continued).**

$$AB_0 + (AB_1 - B_0)c + (AB_2 - B_1)c^2 + \cdots + (AB_{n-1} - B_{n-2})c^{n-1} + (-B_{n-1}c^n)$$
$$= (s_0 + s_1 c + s_2 c^2 + \cdots + s_n c^n)\mathcal{I}_n.$$

Equating the coefficients of $c$:

$$\begin{array}{rcl} AB_0 & = & s_0\mathcal{I}_n \\ AB_1 - B_0 & = & s_1\mathcal{I}_n \\ AB_2 - B_1 & = & s_2\mathcal{I}_n \\ & \vdots & \\ AB_{n-1} - B_{n-2} & = & s_{n-1}\mathcal{I}_n \\ -B_{n-1} & = & s_n\mathcal{I}_n \end{array} \quad \text{and} \quad \begin{array}{rcl} AB_0 & = & s_0\mathcal{I}_n \\ A^2B_1 - AB_0 & = & s_1 A \\ A^3B_2 - A^2B_1 & = & s_2 A^2 \\ & \vdots & \\ A^nB_{n-1} - A^{n-1}B_{n-2} & = & s_{n-1}A^{n-1} \\ -A^nB_{n-1} & = & s_n A^n. \end{array}$$

Summing these $n+1$ equations gives $0 = p_A(A)$, as claimed. □

# Theorem 3.8.4 (continued)

**Proof (continued).**

$$AB_0+(AB_1-B_0)c+(AB_2-B_1)c^2+\cdots+(AB_{n-1}-B_{n-2})c^{n-1}+(-B_{n-1}c^n)$$
$$=(s_0+s_1c+s_2c^2+\cdots+s_nc^n)\mathcal{I}_n.$$

Equating the coefficients of $c$:

$$\begin{array}{rcl} AB_0 & = & s_0\mathcal{I}_n \\ AB_1-B_0 & = & s_1\mathcal{I}_n \\ AB_2-B_1 & = & s_2\mathcal{I}_n \\ & \vdots & \\ AB_{n-1}-B_{n-2} & = & s_{n-1}\mathcal{I}_n \\ -B_{n-1} & = & s_n\mathcal{I}_n \end{array} \quad \text{and} \quad \begin{array}{rcl} AB_0 & = & s_0\mathcal{I}_n \\ A^2B_1-AB_0 & = & s_1A \\ A^3B_2-A^2B_1 & = & s_2A^2 \\ & \vdots & \\ A^nB_{n-1}-A^{n-1}B_{n-2} & = & s_{n-1}A^{n-1} \\ -A^nB_{n-1} & = & s_nA^n. \end{array}$$

Summing these $n+1$ equations gives $0=p_A(A)$, as claimed. □

# Theorem 3.8.5

**Theorem 3.8.5.** Let $q(c) = s_0 + s_1 c + s_2 c^2 + \cdots + s_{n-1} c^{n-1} + c^n$ be a monic polynomial. Then $q(c) = \det(c\mathcal{I} - A)$ for some $n \times n$ matrix $A$. In particular, $q(c) = \det(c\mathcal{I} - A)$ for

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -s_0 & -s_1 & -s_2 & \cdots & -s_{n-1} \end{bmatrix}.$$

Matrix $A$ is called a *companion matrix* for polynomial $q$.

**Proof.** We prove $\det(c\mathcal{I} - A) = q(c)$ by induction on $n$. If $n = 1$ then $A = [-s_0]$ and $\det(c\mathcal{I} - A) = s_0 + c$.

# Theorem 3.8.5

**Theorem 3.8.5.** Let $q(c) = s_0 + s_1 c + s_2 c^2 + \cdots + s_{n-1} c^{n-1} + c^n$ be a monic polynomial. Then $q(c) = \det(c\mathcal{I} - A)$ for some $n \times n$ matrix $A$. In particular, $q(c) = \det(c\mathcal{I} - A)$ for

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -s_0 & -s_1 & -s_2 & \cdots & -s_{n-1} \end{bmatrix}.$$

Matrix $A$ is called a *companion matrix* for polynomial $q$.

**Proof.** We prove $\det(c\mathcal{I} - A) = q(c)$ by induction on $n$. If $n = 1$ then $A = [-s_0]$ and $\det(c\mathcal{I} - A) = s_0 + c$. For clarity, we also observe that for $n = 2$, $A = \begin{bmatrix} 0 & 1 \\ -s_0 & -s_1 \end{bmatrix}$, $c\mathcal{I} - A = \begin{bmatrix} c & -1 \\ s_0 & c + s_1 \end{bmatrix}$, and

$$\det(c\mathcal{I} - A) = (c)(c + s_1) - (s_0)(-1) = s_0 + s_1 c + c^2.$$

# Theorem 3.8.5

**Theorem 3.8.5.** Let $q(c) = s_0 + s_1 c + s_2 c^2 + \cdots + s_{n-1} c^{n-1} + c^n$ be a monic polynomial. Then $q(c) = \det(c\mathcal{I} - A)$ for some $n \times n$ matrix $A$. In particular, $q(c) = \det(c\mathcal{I} - A)$ for

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -s_0 & -s_1 & -s_2 & \cdots & -s_{n-1} \end{bmatrix}.$$

Matrix $A$ is called a *companion matrix* for polynomial $q$.

**Proof.** We prove $\det(c\mathcal{I} - A) = q(c)$ by induction on $n$. If $n = 1$ then $A = [-s_0]$ and $\det(c\mathcal{I} - A) = s_0 + c$. For clarity, we also observe that for $n = 2$, $A = \begin{bmatrix} 0 & 1 \\ -s_0 & -s_1 \end{bmatrix}$, $c\mathcal{I} - A = \begin{bmatrix} c & -1 \\ s_0 & c + s_1 \end{bmatrix}$, and

$$\det(c\mathcal{I} - A) = (c)(c + s_1) - (s_0)(-1) = s_0 + s_1 c + c^2.$$

# Theorem 3.8.5 (continued 1)

**Proof (continued).** Suppose the result holds for $k = n$ and consider the case $k = n + 1$. We have

$$c\mathcal{I} - A = \begin{bmatrix} c & -1 & 0 & \cdots & 0 & 0 \\ 0 & c & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c & -1 \\ s_0 & s_1 & s_2 & \cdots & s_{k-1} & c + s_k \end{bmatrix}.$$

Then $\det(c\mathcal{I} - A)$ can be computed using cofactors and column 1 by Theorem 3.1.F to give

$$\det(c\mathcal{I} - A) = c\det\left(\begin{bmatrix} c & -1 & 0 & \cdots & 0 & 0 \\ 0 & c & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c & -1 \\ s_1 & s_2 & s_3 & \cdots & s_{k-1} & c + s_k \end{bmatrix}\right) \cdots$$

# Theorem 3.8.5 (continued 1)

**Proof (continued).** Suppose the result holds for $k = n$ and consider the case $k = n + 1$. We have

$$c\mathcal{I} - A = \begin{bmatrix} c & -1 & 0 & \cdots & 0 & 0 \\ 0 & c & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c & -1 \\ s_0 & s_1 & s_2 & \cdots & s_{k-1} & c + s_k \end{bmatrix}.$$

Then $\det(c\mathcal{I} - A)$ can be computed using cofactors and column 1 by Theorem 3.1.F to give

$$\det(c\mathcal{I} - A) = c\det\left(\begin{bmatrix} c & -1 & 0 & \cdots & 0 & 0 \\ 0 & c & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c & -1 \\ s_1 & s_2 & s_3 & \cdots & s_{k-1} & c + s_k \end{bmatrix}\right) \cdots$$

# Theorem 3.8.5 (continued 2)

**Proof (continued).** . . .

$$+(-1)^k s_0 \det\left(\begin{bmatrix} -1 & 0 & \cdots & 0 & 0 \\ c & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 0 \\ 0 & 0 & \cdots & c & -1 \end{bmatrix}\right).$$

By the induction hypothesis, the first determinant is $s_1 + s_2 c + s_3 c^2 + \cdots + s_{k-1} c^{k-2} + c^{k-1}$. Since the second determinant involves a lower triangular matrix by Theorem 3.1.H (with $A = -\mathcal{I}$ and

$$T = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -c & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & -c & 1 \end{bmatrix}$$

) we have that this determinant is $\det(-\mathcal{I}) = (-1)^k$; $\det(-\mathcal{I})$ follows from Note 3.1.B.

# Theorem 3.8.5 (continued 2)

**Proof (continued).** . . .

$$+(-1)^k s_0 \det\left(\begin{bmatrix} -1 & 0 & \cdots & 0 & 0 \\ c & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 0 \\ 0 & 0 & \cdots & c & -1 \end{bmatrix}\right).$$

By the induction hypothesis, the first determinant is $s_1 + s_2 c + s_3 c^2 + \cdots + s_{k-1} c^{k-2} + c^{k-1}$. Since the second determinant involves a lower triangular matrix by Theorem 3.1.H (with $A = -\mathcal{I}$ and

$T = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -c & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & -c & 1 \end{bmatrix}$) we have that this determinant is $\det(-\mathcal{I}) = (-1)^k$; $\det(-\mathcal{I})$ follows from Note 3.1.B.

# Theorem 3.8.5 (continued 3)

**Theorem 3.8.5.** Let $q(c) = s_0 + s_1 c + s_2 c^2 + \cdots + s_{n-1} c^{n-1} + c^n$ be a monic polynomial. Then $q(c) = \det(c\mathcal{I} - A)$ for some $n \times n$ matrix $A$. In particular, $q(c) = \det(c\mathcal{I} - A)$ for

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -s_0 & -s_1 & -s_2 & \cdots & -s_{n-1} \end{bmatrix}.$$

**Proof (continued).** Hence

$$\det(c\mathcal{I} - A) = c(s_1 + s_2 c + s_3 c^2 + \cdots + s_{k-1} c^{k-2} + c^{k-1}) + (-1)^k (-1)^k s_0$$

$$= s_0 + s_1 c + s_2 c^2 + \cdots + s_{k-1} c^{k-1} + c^k = s_0 + s_1 c + s_2 c^2 + \cdots + s_n c^n + c^{n+1}$$

and the result holds for $k = n + 1$. Therefore, by Mathematical Induction, it holds for all $n \in \mathbb{N}$. □

# Theorem 3.8.6

**Theorem 3.8.6.** Let $A$ be an $n \times n$ matrix with eigenvalues $c_1, c_2, \ldots, c_n$. Then $\det(A) = \prod_{i=1}^{n} c_i$ and $\operatorname{tr}(A) = \sum_{i=1}^{n} c_i$.

**Proof.** Since the eigenvalues of $A$ are the roots of the characteristic polynomial $p_A(c)$, then $p_A(c) = (-1)^n(c - c_1)(c - c_2)\cdots(c - c_n)$ (the coefficient of $c^n$ is $(-1)^n$ as explained in Note 3.8.A). So

$$\det(A - c\mathcal{I}) = (-1)^n(c^n + (-c_1 - c_2 - \cdots - c_n)c^{n-1} + \cdots + (-1)^n c_1 c_2 \cdots c_n) \quad (*)$$

and by setting variable $c = 0$ we see that $\det(A) = c_1 c_2 \cdots c_n$.

# Theorem 3.8.6

**Theorem 3.8.6.** Let $A$ be an $n \times n$ matrix with eigenvalues $c_1, c_2, \ldots, c_n$. Then $\det(A) = \prod_{i=1}^n c_i$ and $\mathrm{tr}(A) = \sum_{i=1}^n c_i$.

**Proof.** Since the eigenvalues of $A$ are the roots of the characteristic polynomial $p_A(c)$, then $p_A(c) = (-1)^n(c - c_1)(c - c_2)\cdots(c - c_n)$ (the coefficient of $c^n$ is $(-1)^n$ as explained in Note 3.8.A). So

$$\det(A - c\mathcal{I}) = (-1)^n(c^n + (-c_1 - c_2 - \cdots - c_n)c^{n-1} + \cdots + (-1)^n c_1 c_2 \cdots c_n) \quad (*)$$

and by setting variable $c = 0$ we see that $\det(A) = c_1 c_2 \cdots c_n$.

We also have $\det(A - c\mathcal{I}) =$

$$\det\left(\begin{bmatrix} a_{11} - c & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - c & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - c \end{bmatrix}\right) = \sum_{\pi \in S_n} \sigma(\pi) \prod_{i=1}^n b_{i\,\pi(i)}$$

where $b_{i\,\pi(i)}$ is the $(i, \pi(i))$ entry of $A - c\mathcal{I}$.

# Theorem 3.8.6

**Theorem 3.8.6.** Let $A$ be an $n \times n$ matrix with eigenvalues $c_1, c_2, \ldots, c_n$. Then $\det(A) = \prod_{i=1}^n c_i$ and $\operatorname{tr}(A) = \sum_{i=1}^n c_i$.

**Proof.** Since the eigenvalues of $A$ are the roots of the characteristic polynomial $p_A(c)$, then $p_A(c) = (-1)^n(c - c_1)(c - c_2)\cdots(c - c_n)$ (the coefficient of $c^n$ is $(-1)^n$ as explained in Note 3.8.A). So

$$\det(A - c\mathcal{I}) = (-1)^n(c^n + (-c_1 - c_2 - \cdots - c_n)c^{n-1} + \cdots + (-1)^n c_1 c_2 \cdots c_n) \quad (*)$$

and by setting variable $c = 0$ we see that $\det(A) = c_1 c_2 \cdots c_n$.

We also have $\det(A - c\mathcal{I}) =$

$$\det\left(\begin{bmatrix} a_{11} - c & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - c & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - c \end{bmatrix}\right) = \sum_{\pi \in S_n} \sigma(\pi) \prod_{i=1}^n b_{i\,\pi(i)}$$

where $b_{i\,\pi(i)}$ is the $(i, \pi(i))$ entry of $A - c\mathcal{I}$.

# Theorem 3.8.6 (continued)

**Theorem 3.8.6.** Let $A$ be an $n \times n$ matrix with eigenvalues $c_1, c_2, \ldots, c_n$. Then $\det(A) = \prod_{i=1}^n c_i$ and $\text{tr}(A) = \sum_{i=1}^n c_i$.

**Proof (continued).** As described in Note 3.8.A, the only $\sigma(\pi) \prod_{i=1}^n b_{i\,\pi(i)}$ which contains powers of $c^n$ or $c^{n-1}$ is the case when $\pi$ is the identity. In this case,

$$\sigma(\pi) \prod_{i=1}^n b_{i\,\pi(i)} = \prod_{i=1}^n (a_{ii} - c)$$

$$= (-1)^n c^n + (-1)^{n-1}(a_{11} + a_{22} + \cdots + a_{nn})c^{n-1} + \cdots + a_{11}a_{22} \cdots a_{nn}.$$

Equating this with $(*)$ we see that

$$(-1)^{n-1}\text{tr}(A) = (-1)^{n-1}(a_{11} + a_{22} + \cdots + a_{nn}) = (-1)^n(-c_1 - c_2 - \cdots - c_n)$$

or $\text{tr}(A) = c_1 + c_2 + \cdots + c_n$. □

# Theorem 3.8.6 (continued)

**Theorem 3.8.6.** Let $A$ be an $n \times n$ matrix with eigenvalues $c_1, c_2, \ldots, c_n$. Then $\det(A) = \prod_{i=1}^n c_i$ and $\operatorname{tr}(A) = \sum_{i=1}^n c_i$.

**Proof (continued).** As described in Note 3.8.A, the only $\sigma(\pi)\prod_{i=1}^n b_{i\,\pi(i)}$ which contains powers of $c^n$ or $c^{n-1}$ is the case when $\pi$ is the identity. In this case,

$$\sigma(\pi)\prod_{i=1}^n b_{i\,\pi(i)} = \prod_{i=1}^n (a_{ii} - c)$$

$$= (-1)^n c^n + (-1)^{n-1}(a_{11} + a_{22} + \cdots + a_{nn})c^{n-1} + \cdots + a_{11}a_{22}\cdots a_{nn}.$$

Equating this with $(*)$ we see that

$$(-1)^{n-1}\operatorname{tr}(A) = (-1)^{n-1}(a_{11}+a_{22}+\cdots+a_{nn}) = (-1)^n(-c_1-c_2-\cdots-c_n)$$

or $\operatorname{tr}(A) = c_1 + c_2 + \cdots + c_n$. □

# Theorem 3.8.8

**Theorem 3.8.8.** Let $A$ be an $n \times n$ matrix with distinct eigenvalues $\{c_1, c_2, \ldots, c_k\}$ and corresponding eigenvectors $\{x_1, x_2, \ldots, x_k\}$ where $(c_i, x_i)$ is an eigenpair for $A$. Then $\{x_1, x_2, \ldots, x_k\}$ is a set of linearly independent vectors. That is, eigenvectors associated with distinct eigenvalues are linearly independent.

**Proof.** Suppose not. ASSUME that $\{x_1, x_2, \ldots, x_k\}$ is not linearly independent. Then there is some maximal subset $\{y_1, y_2, \ldots, y_j\} \subset \{x_1, x_2, \ldots, x_k\}$ which is linearly independent and $j < k$. Let the corresponding eigenvalues for the $y_i$ be $\{\mu_1, \mu_2, \ldots, \mu_j\} \subset \{c_1, c_2, \ldots, c_k\}$.

# Theorem 3.8.8

**Theorem 3.8.8.** Let $A$ be an $n \times n$ matrix with distinct eigenvalues $\{c_1, c_2, \ldots, c_k\}$ and corresponding eigenvectors $\{x_1, x_2, \ldots, x_k\}$ where $(c_i, x_i)$ is an eigenpair for $A$. Then $\{x_1, x_2, \ldots, x_k\}$ is a set of linearly independent vectors. That is, eigenvectors associated with distinct eigenvalues are linearly independent.

**Proof.** Suppose not. ASSUME that $\{x_1, x_2, \ldots, x_k\}$ is not linearly independent. Then there is some maximal subset $\{y_1, y_2, \ldots, y_j\} \subset \{x_1, x_2, \ldots, x_k\}$ which is linearly independent and $j < k$. Let the corresponding eigenvalues for the $y_i$ be $\{\mu_1, \mu_2, \ldots, \mu_j\} \subset \{c_1, c_2, \ldots, c_k\}$. Then for some element in $\{x_1, x_2, \ldots, x_k\} \setminus \{y_1, y_2, \ldots y_j\}$, say $y_{j+1}$, we have $y_{j+1} = \sum_{i=1}^{j} t_i y_i$ for some $t_i \in \mathbb{C}$ (not all $t_i = 0$) since $\{y_1, y_2, \ldots, y_{j+1}\}$ is a linearly dependent set. Since $y_{j+1}$ is an eigenvector of $A$, then there is an eigenvalue $\mu_{j+1}$ for $y_{j+1}$ in $\{c_1, c_2, \ldots, c_k\}$ and by construction, $\mu_{j+1}$ is distinct from $\mu_1, \mu_2, \ldots, \mu_j$.

# Theorem 3.8.8

**Theorem 3.8.8.** Let $A$ be an $n \times n$ matrix with distinct eigenvalues $\{c_1, c_2, \ldots, c_k\}$ and corresponding eigenvectors $\{x_1, x_2, \ldots, x_k\}$ where $(c_i, x_i)$ is an eigenpair for $A$. Then $\{x_1, x_2, \ldots, x_k\}$ is a set of linearly independent vectors. That is, eigenvectors associated with distinct eigenvalues are linearly independent.

**Proof.** Suppose not. ASSUME that $\{x_1, x_2, \ldots, x_k\}$ is not linearly independent. Then there is some maximal subset $\{y_1, y_2, \ldots, y_j\} \subset \{x_1, x_2, \ldots, x_k\}$ which is linearly independent and $j < k$. Let the corresponding eigenvalues for the $y_i$ be $\{\mu_1, \mu_2, \ldots, \mu_j\} \subset \{c_1, c_2, \ldots, c_k\}$. Then for some element in $\{x_1, x_2, \ldots, x_k\} \setminus \{y_1, y_2, \ldots y_j\}$, say $y_{j+1}$, we have $y_{j+1} = \sum_{i=1}^{j} t_i y_i$ for some $t_i \in \mathbb{C}$ (not all $t_i = 0$) since $\{y_1, y_2, \ldots, y_{j+1}\}$ is a linearly dependent set. Since $y_{j+1}$ is an eigenvector of $A$, then there is an eigenvalue $\mu_{j+1}$ for $y_{j+1}$ in $\{c_1, c_2, \ldots, c_k\}$ and by construction, $\mu_{j+1}$ is distinct from $\mu_1, \mu_2, \ldots, \mu_j$.

# Theorem 3.8.8 (continued)

**Theorem 3.8.8.** Let $A$ be an $n \times n$ matrix with distinct eigenvalues $\{c_1, c_2, \ldots, c_k\}$ and corresponding eigenvectors $\{x_1, x_2, \ldots, x_k\}$ where $(c_i, x_i)$ is an eigenpair for $A$. Then $\{x_1, x_2, \ldots, x_k\}$ is a set of linearly independent vectors. That is, eigenvectors associated with distinct eigenvalues are linearly independent.

**Proof (continued).** Now $Ay_{j+1} = \mu_{j+1} y_{j+1}$, so we have $0 = Ay_{j+1} - \mu_{j+1} y_{j+1} = (A - \mu_{j+1}\mathcal{I}) y_{j+1} = (A - \mu_{j+1}\mathcal{I}) \sum_{i=1}^{j} t_i y_i$ or

$$0 = \sum_{i=1}^{j} t_i (Ay_i - \mu_{j+1}\mathcal{I} y_i) = \sum_{i=1}^{j} t_i (\mu_i y_i - \mu_{j+1} y_i) = \sum_{i=1}^{j} t_i (\mu_i - \mu_{j+1}) y_i.$$

But then the coefficients $t_i(\mu_i - \mu_{j+1})$ for $1 \leq i \leq j$ are not all 0 and so this gives a dependence relation on $\{y_1, y_2, \ldots, y_j\}$, a CONTRADICTION to the fact that this is a linearly independent set. So the assumption that $\{x_1, x_2, \ldots, x_k\}$ is not linearly independent is false and so the set is linearly independent, as claimed. □

# Theorem 3.8.8 (continued)

**Theorem 3.8.8.** Let $A$ be an $n \times n$ matrix with distinct eigenvalues $\{c_1, c_2, \ldots, c_k\}$ and corresponding eigenvectors $\{x_1, x_2, \ldots, x_k\}$ where $(c_i, x_i)$ is an eigenpair for $A$. Then $\{x_1, x_2, \ldots, x_k\}$ is a set of linearly independent vectors. That is, eigenvectors associated with distinct eigenvalues are linearly independent.

**Proof (continued).** Now $Ay_{j+1} = \mu_{j+1}y_{j+1}$, so we have $0 = Ay_{j+1} - \mu_{j+1}y_{j+1} = (A - \mu_{j+1}\mathcal{I})y_{j+1} = (A - \mu_{j+1}\mathcal{I})\sum_{i=1}^{j} t_i y_i$ or

$$0 = \sum_{i=1}^{j} t_i(Ay_i - \mu_{j+1}\mathcal{I}y_i) = \sum_{i=1}^{j} t_i(\mu_i y_i - \mu_{j+1}y_i) = \sum_{i=1}^{j} t_i(\mu_i - \mu_{j+1})y_i.$$

But then the coefficients $t_i(\mu_i - \mu_{j+1})$ for $1 \leq i \leq j$ are not all 0 and so this gives a dependence relation on $\{y_1, y_2, \ldots, y_j\}$, a CONTRADICTION to the fact that this is a linearly independent set. So the assumption that $\{x_1, x_2, \ldots, x_k\}$ is not linearly independent is false and so the set is linearly independent, as claimed. □

# Theorem 3.8.9

**Theorem 3.8.9.** For any square matrix $A$, a Schur factorization exists.

**Proof.** If $A$ is $1 \times 1$, the result is trivial; take $Q = [1]$ and $B = A$. If $A$ is the zero matrix, then we let $Q$ be an identity matrix of the appropriate size and left $B$ be a zero matrix (which is, in fact, upper triangular).

# Theorem 3.8.9

**Theorem 3.8.9.** For any square matrix $A$, a Schur factorization exists.

**Proof.** If $A$ is $1 \times 1$, the result is trivial; take $Q = [1]$ and $B = A$. If $A$ is the zero matrix, then we let $Q$ be an identity matrix of the appropriate size and left $B$ be a zero matrix (which is, in fact, upper triangular).

For $n > 1$, let $(c, v)$ be an eigenpair of $A$ with eigenvector $v$ normalized. Form an orthogonal matrix $U$ with $v$ as the first column (this can be done by taking $v$ followed by the standard basis vectors for $\mathbb{R}^n$ and the applying the Gram-Schmidt process; this produces an orthonormal basis of $\mathbb{R}^n$ which includes vector $v$ [and one of the vectors will be a linear combination of the others and will become the zero vector leaving $n$ nonzero vectors]). Let matrix $U_2$ consist of the remaining columns of the basis so that $U = [v \mid U_2]$.

# Theorem 3.8.9

**Theorem 3.8.9.** For any square matrix $A$, a Schur factorization exists.

**Proof.** If $A$ is $1 \times 1$, the result is trivial; take $Q = [1]$ and $B = A$. If $A$ is the zero matrix, then we let $Q$ be an identity matrix of the appropriate size and left $B$ be a zero matrix (which is, in fact, upper triangular).

For $n > 1$, let $(c, v)$ be an eigenpair of $A$ with eigenvector $v$ normalized. Form an orthogonal matrix $U$ with $v$ as the first column (this can be done by taking $v$ followed by the standard basis vectors for $\mathbb{R}^n$ and the applying the Gram-Schmidt process; this produces an orthonormal basis of $\mathbb{R}^n$ which includes vector $v$ [and one of the vectors will be a linear combination of the others and will become the zero vector leaving $n$ nonzero vectors]). Let matrix $U_2$ consist of the remaining columns of the basis so that $U = [v \mid U_2]$.

# Theorem 3.8.9 (continued 1)

**Proof (continued).** So

$$\begin{aligned} U^TAU &= \begin{bmatrix} v^T \\ U_2^T \end{bmatrix} A[v \mid U_2] = \begin{bmatrix} v^TA \\ U_2^TA \end{bmatrix} [v \mid U_2] = \begin{bmatrix} v^TAv & v^TAU_2 \\ U_2^TAv & U_2^TAU_2 \end{bmatrix} \\ &= \begin{bmatrix} v^Tcv & v^TAU_2 \\ U_2^Tcv & U_2^TAU_2 \end{bmatrix} \text{ since } Av = cv \\ &= \begin{bmatrix} cv^Tv & v^TAU_2 \\ cU_2^Tv & U_2^TAU_2 \end{bmatrix} = \begin{bmatrix} c & v^TAU_2 \\ 0 & U_2^TAU_2 \end{bmatrix} = B \qquad (*) \end{aligned}$$

since $v^Tv = \|v\|^2 = 1$ and $v$ is orthogonal to each column of $U_2$ (so the inner product of each column of $U_2$ with $v$ is 0 and $U_2v^T$ is a $(n-1) \times 1$ zero matrix)

where $U_2^TAU_2$ is an $(n-1) \times (n-1)$ matrix.

# Theorem 3.8.9 (continued 2)

**Proof (continued).** Since $U$ is orthogonal, by Theorem 3.7.1, $U^T = U^{-1}$. By Theorem 3.8.2(8), the eigenvalues of $U^TAU = U^{-1}AU$ are the same as the eigenvalues of $A$. If $n = 2$, then $U_2^TAU_2$ is a scalar (well a $1 \times 1$ matrix) and the two eigenvalues of $A$ must be $c$ and this scalar (notice that $U^TAU = B$ in this case is upper triangular and so the eigenvalues are the diagonal entries by Theorem 3.8.2(5)). So the result holds for $k \times k$ where $k = 2$.

We now show the result holds by induction. Suppose a Schur factorization exists for all $k \times k$ matrices where $k = n - 1$. Let $A$ be an $n \times n$ matrix with eigenpair $(c, v)$ where $v$ is normalized.

# Theorem 3.8.9 (continued 2)

**Proof (continued).** Since $U$ is orthogonal, by Theorem 3.7.1, $U^T = U^{-1}$. By Theorem 3.8.2(8), the eigenvalues of $U^T AU = U^{-1}AU$ are the same as the eigenvalues of $A$. If $n = 2$, then $U_2^T AU_2$ is a scalar (well a $1 \times 1$ matrix) and the two eigenvalues of $A$ must be $c$ and this scalar (notice that $U^T AU = B$ in this case is upper triangular and so the eigenvalues are the diagonal entries by Theorem 3.8.2(5)). So the result holds for $k \times k$ where $k = 2$.

We now show the result holds by induction. Suppose a Schur factorization exists for all $k \times k$ matrices where $k = n - 1$. Let $A$ be an $n \times n$ matrix with eigenpair $(c, v)$ where $v$ is normalized. As discussed above in $(*)$, $U^T AU = \begin{bmatrix} c & v^T AU_2 \\ 0 & U_2^T AU_2 \end{bmatrix}$ where $U_2^T AU_2$ is an $(n-1) \times (n-1)$ matrix. So by the induction hypothesis there exists $(n-1) \times (n-1)$ orthogonal matrix $V$ such that

$$V^T(U_2^T AU_2)V = T \text{ where } T \text{ is upper triangluar.} \qquad (**)$$

# Theorem 3.8.9 (continued 2)

**Proof (continued).** Since $U$ is orthogonal, by Theorem 3.7.1, $U^T = U^{-1}$. By Theorem 3.8.2(8), the eigenvalues of $U^T AU = U^{-1}AU$ are the same as the eigenvalues of $A$. If $n = 2$, then $U_2^T AU_2$ is a scalar (well a $1 \times 1$ matrix) and the two eigenvalues of $A$ must be $c$ and this scalar (notice that $U^T AU = B$ in this case is upper triangular and so the eigenvalues are the diagonal entries by Theorem 3.8.2(5)). So the result holds for $k \times k$ where $k = 2$.

We now show the result holds by induction. Suppose a Schur factorization exists for all $k \times k$ matrices where $k = n - 1$. Let $A$ be an $n \times n$ matrix with eigenpair $(c, v)$ where $v$ is normalized. As discussed above in $(*)$, $U^T AU = \begin{bmatrix} c & v^T AU_2 \\ 0 & U_2^T AU_2 \end{bmatrix}$ where $U_2^T AU_2$ is an $(n-1) \times (n-1)$ matrix. So by the induction hypothesis there exists $(n-1) \times (n-1)$ orthogonal matrix $V$ such that

$$V^T(U_2^T AU_2)V = T \text{ where } T \text{ is upper triangluar.} \qquad (**)$$

# Theorem 3.8.9 (continued 3)

**Proof (continued).** Let $Q = U \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix}$. Then

$$Q^T Q = \begin{bmatrix} 1 & 0 \\ 0 & V^T \end{bmatrix} U^T U \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & VV^T \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{I} \end{bmatrix}$$

by Theorem 3.7.1 (which implies $U^T U = \mathcal{I}$ and $VV^T = \mathcal{I}$) and so $Q$ is orthogonal (by Theorem 3.7.1, again). Next, let

$$\begin{aligned} Q^T AQ &= \begin{bmatrix} 1 & 0 \\ 0 & V^T \end{bmatrix} U^T AU \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & V^T \end{bmatrix} \begin{bmatrix} c & v^T AU_2 \\ 0 & U_2^T AU_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} \text{ from } (*) \\ &= \begin{bmatrix} c & v^T AU_2 \\ 0 & V^T U_2^T AU_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} = \begin{bmatrix} c & v^T AU_2 V \\ 0 & V^T U_2^T AU_2 V \end{bmatrix} \\ &= \begin{bmatrix} c & v^T AU_2 V \\ 0 & T \end{bmatrix} = B \text{ by } (**). \end{aligned}$$

# Theorem 3.8.9 (continued 3)

**Proof (continued).** Let $Q = U \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix}$. Then

$$Q^T Q = \begin{bmatrix} 1 & 0 \\ 0 & V^T \end{bmatrix} U^T U \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & VV^T \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{I} \end{bmatrix}$$

by Theorem 3.7.1 (which implies $U^T U = \mathcal{I}$ and $VV^T = \mathcal{I}$) and so $Q$ is orthogonal (by Theorem 3.7.1, again). Next, let

$$\begin{aligned} Q^T A Q &= \begin{bmatrix} 1 & 0 \\ 0 & V^T \end{bmatrix} U^T A U \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & V^T \end{bmatrix} \begin{bmatrix} c & v^T A U_2 \\ 0 & U_2^T A U_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} \text{ from } (*) \\ &= \begin{bmatrix} c & v^T A U_2 \\ 0 & V^T U_2^T A U_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} = \begin{bmatrix} c & v^T A U_2 V \\ 0 & V^T U_2^T A U_2 V \end{bmatrix} \\ &= \begin{bmatrix} c & v^T A U_2 V \\ 0 & T \end{bmatrix} = B \text{ by } (**). \end{aligned}$$

# Theorem 3.8.9 (continued 4)

**Theorem 3.8.9.** For any square matrix $A$, a Schur factorization exists.

**Proof (continued).** So

$$Q^T AQ = \begin{bmatrix} c & v^T AU_2 V \\ 0 & T \end{bmatrix} = B.$$

Since $c$ is a constant and $T$ is upper triangular, then $B$ is upper triangular and the result holds for $k = n$. Therefore, by Mathematical Induction, every $n \times n$ matrix has a Schur factorization. □

# Theorem 3.8.10

**Theorem 3.8.10.** Let $A$ be an $n \times n$ matrix, let $c_1, c_2, \ldots, c_n$ be (possibly complex) scalars, and let $v_1, v_2, \ldots, v_n$ be nonzero $n$-vectors. Let $V$ be an $n \times n$ matrix with $i$th column $v_i$ for $1 \leq i \leq n$ and let $C = \text{diag}(c_1, c_2, \ldots, c_n)$. Then $AV = VC$ if and only if $c_1, c_2, \ldots, c_n$ are eigenvalues of $A$ and $v_j$ is an eigenvector of $A$ corresponding to $c_j$ for $j = 1, 2, \ldots, n$.

**Proof.** The $j$th column of $VC = [v_1, v_2, \ldots, v_n]\text{diag}(c_1, c_2, \ldots, c_n)$ is $c_j v_j$. The $j$th column of $AV$ is $Av_j$. So $AV = CV$ if and only if $Av_j = c_j v_j$ for $1 \leq j \leq n$. That is, $AV = VC$ if and only if $v_j$ is an eigenvector of $A$ with corresponding eigenvalue $c_j$. □

# Theorem 3.8.10

**Theorem 3.8.10.** Let $A$ be an $n \times n$ matrix, let $c_1, c_2, \ldots, c_n$ be (possibly complex) scalars, and let $v_1, v_2, \ldots, v_n$ be nonzero $n$-vectors. Let $V$ be an $n \times n$ matrix with $i$th column $v_i$ for $1 \leq i \leq n$ and let $C = \text{diag}(c_1, c_2, \ldots, c_n)$. Then $AV = VC$ if and only if $c_1, c_2, \ldots, c_n$ are eigenvalues of $A$ and $v_j$ is an eigenvector of $A$ corresponding to $c_j$ for $j = 1, 2, \ldots, n$.

**Proof.** The $j$th column of $VC = [v_1, \, v_2, \, \ldots, \, v_n]\text{diag}(c_1, c_2, \ldots, c_n)$ is $c_j v_j$. The $j$th column of $AV$ is $Av_j$. So $AV = CV$ if and only if $Av_j = c_j v_j$ for $1 \leq j \leq n$. That is, $AV = VC$ if and only if $v_j$ is an eigenvector of $A$ with corresponding eigenvalue $c_j$. □

# Theorem 3.8.11

**Theorem 3.8.11. Diagonalizability Theorem.**
Let $A$ be an $n \times n$ matrix with distinct eigenvalues $c_1, c_2, \ldots, c_k$ with algebraic multiplicities $m_1, m_2, \ldots, m_k$, respectively. Then $A$ is diagonalizable if and only if $\text{rank}(A - c_i\mathcal{I}) = n - m_i$ for $i = 1, 2, \ldots, k$ (that is, each eigenvalue is semisimple).

**Proof.** For sufficiency, suppose that for each $i$, $1 \leq i \leq k$, we have $\text{rank}(A - c_i\mathcal{I}) = n - m_i$. By Theorem 3.5.4, $\dim(\mathcal{N}(A - c_i\mathcal{I})) = n - \text{rank}(A - c_i\mathcal{I}) = m_i$. Now $\mathcal{N}(A - c_i\mathcal{I})$ is the set of all $n$-vectors $x$ such that $(A - c_i\mathcal{I})x = 0$. Since $\dim(\mathcal{N}(A - c_i\mathcal{I})) = m_i$, there are $m_i$ linearly independent $x$ such that $Ax = c_i\mathcal{I}x = c_ix$ (these $m_i$ vectors are a basis for the eigenspace of $c_i$).

# Theorem 3.8.11

**Theorem 3.8.11. Diagonalizability Theorem.**
Let $A$ be an $n \times n$ matrix with distinct eigenvalues $c_1, c_2, \ldots, c_k$ with algebraic multiplicities $m_1, m_2, \ldots, m_k$, respectively. Then $A$ is diagonalizable if and only if $\text{rank}(A - c_i\mathcal{I}) = n - m_i$ for $i = 1, 2, \ldots, k$ (that is, each eigenvalue is semisimple).

**Proof.** For sufficiency, suppose that for each $i$, $1 \leq i \leq k$, we have $\text{rank}(A - c_i\mathcal{I}) = n - m_i$. By Theorem 3.5.4, $\dim(\mathcal{N}(A - c_i\mathcal{I})) = n - \text{rank}(A - c_i\mathcal{I}) = m_i$. Now $\mathcal{N}(A - c_i\mathcal{I})$ is the set of all $n$-vectors $x$ such that $(A - c_i\mathcal{I})x = 0$. Since $\dim(\mathcal{N}(A - c_i\mathcal{I})) = m_i$, there are $m_i$ linearly independent $x$ such that $Ax = c_i\mathcal{I}x = c_i x$ (these $m_i$ vectors are a basis for the eigenspace of $c_i$). By Theorem 3.8.8, eigenvectors associated with distinct eigenvalues are linearly independent. So there are $m_1 + m_2 + \cdots + m_k = n$ linearly independent eigenvectors for $A$. That is, for matrix $V$ with the linearly independent eigenvectors as its columns, we have $\text{rank}(V) = n$ and so $V^{-1}$ exists.

# Theorem 3.8.11

**Theorem 3.8.11. Diagonalizability Theorem.**
Let $A$ be an $n \times n$ matrix with distinct eigenvalues $c_1, c_2, \ldots, c_k$ with algebraic multiplicities $m_1, m_2, \ldots, m_k$, respectively. Then $A$ is diagonalizable if and only if $\text{rank}(A - c_i\mathcal{I}) = n - m_i$ for $i = 1, 2, \ldots, k$ (that is, each eigenvalue is semisimple).

**Proof.** For sufficiency, suppose that for each $i$, $1 \leq i \leq k$, we have $\text{rank}(A - c_i\mathcal{I}) = n - m_i$. By Theorem 3.5.4, $\dim(\mathcal{N}(A - c_i\mathcal{I})) = n - \text{rank}(A - c_i\mathcal{I}) = m_i$. Now $\mathcal{N}(A - c_i\mathcal{I})$ is the set of all $n$-vectors $x$ such that $(A - c_i\mathcal{I})x = 0$. Since $\dim(\mathcal{N}(A - c_i\mathcal{I})) = m_i$, there are $m_i$ linearly independent $x$ such that $Ax = c_i\mathcal{I}x = c_ix$ (these $m_i$ vectors are a basis for the eigenspace of $c_i$). By Theorem 3.8.8, eigenvectors associated with distinct eigenvalues are linearly independent. So there are $m_1 + m_2 + \cdots + m_k = n$ linearly independent eigenvectors for $A$. That is, for matrix $V$ with the linearly independent eigenvectors as its columns, we have $\text{rank}(V) = n$ and so $V^{-1}$ exists.

# Theorem 3.8.11 (continued)

**Proof (continued).** With $C$ a diagonal matrix with $c_{jj}$ as the eigenvalue associated with eigenvector $v_j$, we have $AV = VC$ by Theorem 3.8.10. Therefore, $A = VCV^{-1}$ and $A$ is diagonalizable.

To see the condition is necessary, suppose $A$ is diagonalizable. Then $A = VCV^{-1}$ for *some* invertible $V$ and diagonal $C$. Then $AV = VC$ and so by Theorem 3.8.10, $C = \text{diag}(c_1, c_2, \ldots, c_n)$ where $c_1, c_2, \ldots, c_n$ are eigenvalues of $A$ and $V$ has its $j$th column an eigenvector of $A$ corresponding to $c_j$. Since $V$ is invertible then $\text{rank}(V) = n$ and the eigenvectors in $V$ are linearly independent, with the eigenvalues repeated in $C$ according to multiplicity. So for $1 \leq i \leq k$, the diagonal matrix $C - c_i\mathcal{I}$ has exactly $m_i$ zeros on the diagonal and hence $\text{rank}(C - c_i\mathcal{I}) = n - m_i$.

# Theorem 3.8.11 (continued)

**Proof (continued).** With $C$ a diagonal matrix with $c_{jj}$ as the eigenvalue associated with eigenvector $v_j$, we have $AV = VC$ by Theorem 3.8.10. Therefore, $A = VCV^{-1}$ and $A$ is diagonalizable.

To see the condition is necessary, suppose $A$ is diagonalizable. Then $A = VCV^{-1}$ for *some* invertible $V$ and diagonal $C$. Then $AV = VC$ and so by Theorem 3.8.10, $C = \mathrm{diag}(c_1, c_2, \ldots, c_n)$ where $c_1, c_2, \ldots, c_n$ are eigenvalues of $A$ and $V$ has its $j$th column an eigenvector of $A$ corresponding to $c_j$. Since $V$ is invertible then $\mathrm{rank}(V) = n$ and the eigenvectors in $V$ are linearly independent, with the eigenvalues repeated in $C$ according to multiplicity. So for $1 \leq i \leq k$, the diagonal matrix $C - c_i\mathcal{I}$ has exactly $m_i$ zeros on the diagonal and hence $\mathrm{rank}(C - c_i\mathcal{I}) = n - m_i$. Since $V$ and $V^{-1}$ are invertible and so are of full rank, then by Theorem 3.3.12,

$$n - m_i = \mathrm{rank}(C - c_i\mathcal{I}) = \mathrm{rank}(V(C - c_i\mathcal{I})V^{-1})$$

$$= \mathrm{rank}(VCV^{-1} - c_i\mathcal{I}) = \mathrm{rank}(A - c_i\mathcal{I}), \text{ as claimed. } \square$$

# Theorem 3.8.11 (continued)

**Proof (continued).** With $C$ a diagonal matrix with $c_{jj}$ as the eigenvalue associated with eigenvector $v_j$, we have $AV = VC$ by Theorem 3.8.10. Therefore, $A = VCV^{-1}$ and $A$ is diagonalizable.

To see the condition is necessary, suppose $A$ is diagonalizable. Then $A = VCV^{-1}$ for *some* invertible $V$ and diagonal $C$. Then $AV = VC$ and so by Theorem 3.8.10, $C = \mathrm{diag}(c_1, c_2, \ldots, c_n)$ where $c_1, c_2, \ldots, c_n$ are eigenvalues of $A$ and $V$ has its $j$th column an eigenvector of $A$ corresponding to $c_j$. Since $V$ is invertible then $\mathrm{rank}(V) = n$ and the eigenvectors in $V$ are linearly independent, with the eigenvalues repeated in $C$ according to multiplicity. So for $1 \le i \le k$, the diagonal matrix $C - c_i\mathcal{I}$ has exactly $m_i$ zeros on the diagonal and hence $\mathrm{rank}(C - c_i\mathcal{I}) = n - m_i$. Since $V$ and $V^{-1}$ are invertible and so are of full rank, then by Theorem 3.3.12,

$$n - m_i = \mathrm{rank}(C - c_i\mathcal{I}) = \mathrm{rank}(V(C - c_i\mathcal{I})V^{-1})$$

$$= \mathrm{rank}(VCV^{-1} - c_i\mathcal{I}) = \mathrm{rank}(A - c_i\mathcal{I}), \text{ as claimed. } \square$$

# Theorem 3.8.A

**Theorem 3.8.A.** A (real) $n \times n$ matrix $A$ is orthogonally diagonalizable if and only if $A$ is symmetric.

**Proof.** First, suppose $A$ is orthogonally diagonalizable. Let $C$ be a diagonal matrix and let $Q$ be an orthogonal matrix such that $A = QCQ^T = QCQ^{-1}$ ($Q^T = Q^{-1}$ by Theorem 3.7.1). Then $A^T = (QCQ^T)^T = (Q^T)^T C^T Q^T = QCQ^T = A$ and so $A$ is symmetric.

# Theorem 3.8.A

**Theorem 3.8.A.** A (real) $n \times n$ matrix $A$ is orthogonally diagonalizable if and only if $A$ is symmetric.

**Proof.** First, suppose $A$ is orthogonally diagonalizable. Let $C$ be a diagonal matrix and let $Q$ be an orthogonal matrix such that $A = QCQ^T = QCQ^{-1}$ ($Q^T = Q^{-1}$ by Theorem 3.7.1). Then $A^T = (QCQ^T)^T = (Q^T)^T C^T Q^T = QCQ^T = A$ and so $A$ is symmetric. Now suppose $A$ is symmetric. We show that $A$ is orthogonally diagonalizable using induction. If $n = 1$ then we take $Q = [1]$ and we have $A = QAQ^T$ where $C = A$, so that $A$ is orthogonally diagonalizable. Now suppose the result holds for all $(n-1) \times (n-1)$ matrices.

# Theorem 3.8.A

**Theorem 3.8.A.** A (real) $n \times n$ matrix $A$ is orthogonally diagonalizable if and only if $A$ is symmetric.

**Proof.** First, suppose $A$ is orthogonally diagonalizable. Let $C$ be a diagonal matrix and let $Q$ be an orthogonal matrix such that $A = QCQ^T = QCQ^{-1}$ ($Q^T = Q^{-1}$ by Theorem 3.7.1). Then $A^T = (QCQ^T)^T = (Q^T)^T C^T Q^T = QCQ^T = A$ and so $A$ is symmetric. Now suppose $A$ is symmetric. We show that $A$ is orthogonally diagonalizable using induction. If $n = 1$ then we take $Q = [1]$ and we have $A = QAQ^T$ where $C = A$, so that $A$ is orthogonally diagonalizable. Now suppose the result holds for all $(n-1) \times (n-1)$ matrices. Since $A$ is real and symmetric then by Theorem 3.8.7, the eigenvalues of $A$ are real. Let $c$ be some eigenvalue of $A$. If $v$ is an eigenvector of $A$ associated with $c$ then $\det(A - c\mathcal{I}) = 0$ and since $A$ and $c$ are real then the system of equations $(A - c\mathcal{I})x = 0$ (or $Ax = cx$) has a nontrivial real solution $x = v$.

# Theorem 3.8.A

**Theorem 3.8.A.** A (real) $n \times n$ matrix $A$ is orthogonally diagonalizable if and only if $A$ is symmetric.

**Proof.** First, suppose $A$ is orthogonally diagonalizable. Let $C$ be a diagonal matrix and let $Q$ be an orthogonal matrix such that $A = QCQ^T = QCQ^{-1}$ ($Q^T = Q^{-1}$ by Theorem 3.7.1). Then $A^T = (QCQ^T)^T = (Q^T)^T C^T Q^T = QCQ^T = A$ and so $A$ is symmetric. Now suppose $A$ is symmetric. We show that $A$ is orthogonally diagonalizable using induction. If $n = 1$ then we take $Q = [1]$ and we have $A = QAQ^T$ where $C = A$, so that $A$ is orthogonally diagonalizable. Now suppose the result holds for all $(n-1) \times (n-1)$ matrices. Since $A$ is real and symmetric then by Theorem 3.8.7, the eigenvalues of $A$ are real. Let $c$ be some eigenvalue of $A$. If $v$ is an eigenvector of $A$ associated with $c$ then $\det(A - c\mathcal{I}) = 0$ and since $A$ and $c$ are real then the system of equations $(A - c\mathcal{I})x = 0$ (or $Ax = cx$) has a nontrivial real solution $x = v$.

# Theorem 3.8.A (continued 1)

**Proof (continued).** Define $v_1 = v/\|v\|$. Then $v_1$ is a real unit eigenvector of $A$. Expand the set $\{v_1\}$ to an orthonormal bases of $\mathbb{R}^n$, $\{v_1, v_2, \ldots, v_n\}$, which can be done as explained in the proof of Theorem 3.8.9. Form $n \times n$ matrix $P$ with $i$th column $v_i$ for $1 \le i \le n$. Then $P$ is orthogonal so $P^{-1} = P^T$ by Theorem 3.7.1.

# Theorem 3.8.A (continued 1)

**Proof (continued).** Define $v_1 = v/\|v\|$. Then $v_1$ is a real unit eigenvector of $A$. Expand the set $\{v_1\}$ to an orthonormal bases of $\mathbb{R}^n$, $\{v_1, v_2, \ldots, v_n\}$, which can be done as explained in the proof of Theorem 3.8.9. Form $n \times n$ matrix $P$ with $i$th column $v_i$ for $1 \leq i \leq n$. Then $P$ is orthogonal so $P^{-1} = P^T$ by Theorem 3.7.1. Consider $P^{-1}AP = P^TAP$. This matrix is symmetric because $(P^TAP)^T = P^TA^T(P^T)^T = P^TAP$ (since $A$ is symmetric). The first column of this matrix is

$$P^TAP \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = P^TAv_1 = P^Tcv_1 = cP^Tv_1 = c[v_1 \ v_2 \ \cdots \ v_n]^Tv_1 = \cdots$$

# Theorem 3.8.A (continued 1)

**Proof (continued).** Define $v_1 = v/\|v\|$. Then $v_1$ is a real unit eigenvector of $A$. Expand the set $\{v_1\}$ to an orthonormal bases of $\mathbb{R}^n$, $\{v_1, v_2, \ldots, v_n\}$, which can be done as explained in the proof of Theorem 3.8.9. Form $n \times n$ matrix $P$ with $i$th column $v_i$ for $1 \le i \le n$. Then $P$ is orthogonal so $P^{-1} = P^T$ by Theorem 3.7.1. Consider $P^{-1}AP = P^TAP$. This matrix is symmetric because $(P^TAP)^T = P^TA^T(P^T)^T = P^TAP$ (since $A$ is symmetric). The first column of this matrix is

$$P^TAP \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = P^TAv_1 = P^Tcv_1 = cP^Tv_1 = c[v_1\ v_2\ \cdots\ v_n]^Tv_1 = \cdots$$

# Theorem 3.8.A (continued 2)

**Proof (continued).**

$$\cdots = c \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix} v_1 = c \begin{bmatrix} v_1^T v_1 \\ v_2^T v_1 \\ \vdots \\ v_n^T v_1 \end{bmatrix} = c \begin{bmatrix} \langle v_1, v_1 \rangle \\ \langle v_2, v_1 \rangle \\ \vdots \\ \langle v_n, v_1 \rangle \end{bmatrix} = c \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} c \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Since $P^{-1}AP$ is symmetric, then its first row must be $[c\ 0\ 0\ \cdots\ 0]$. So we must have a partitioning of the form $P^{-1}AP = \begin{bmatrix} c & 0 \\ 0 & B \end{bmatrix}$ where $B$ is an $(n-1) \times (n-1)$ symmetric matrix. By the induction hypothesis, $B$ is orthogonally diagonalizable and so $B = UDU^{-1}$ or $D = U^{-1}BU = U^TBU$ for diagonal matrix $D$ and orthogonal matrix $U$ (matrices $B$, $U$, $D$, and $U^T$ are each $(n-1) \times (n-1)$).

# Theorem 3.8.A (continued 2)

**Proof (continued).**

$$\cdots = c \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix} v_1 = c \begin{bmatrix} v_1^T v_1 \\ v_2^T v_1 \\ \vdots \\ v_n^T v_1 \end{bmatrix} = c \begin{bmatrix} \langle v_1, v_1 \rangle \\ \langle v_2, v_1 \rangle \\ \vdots \\ \langle v_n, v_1 \rangle \end{bmatrix} = c \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} c \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Since $P^{-1}AP$ is symmetric, then its first row must be $[c\ 0\ 0\ \cdots\ 0]$. So we must have a partitioning of the form $P^{-1}AP = \begin{bmatrix} c & 0 \\ 0 & B \end{bmatrix}$ where $B$ is an $(n-1) \times (n-1)$ symmetric matrix. By the induction hypothesis, $B$ is orthogonally diagonalizable and so $B = UDU^{-1}$ or $D = U^{-1}BU = U^TBU$ for diagonal matrix $D$ and orthogonal matrix $U$ (matrices $B$, $U$, $D$, and $U^T$ are each $(n-1) \times (n-1)$).

# Theorem 3.8.A (continued 3)

**Proof (continued).** For this matrix $U$, define $R = \begin{bmatrix} 1 & 0 \\ 0 & U \end{bmatrix}$ and notice that $R^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & U^{-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & U^T \end{bmatrix} = R^T$. Let $Q = PR$; notice $QQ^T = (PR)(PR)^T = PRR^TP^T = \mathcal{I}$, since $P$ and $R$ are orthogonal, and so $Q$ is orthogonal. Then

$$Q^{-1}AQ = (R^{-1}P^{-1})A(PR) = R^{-1}(P^{-1}AP)R = R^{-1}\begin{bmatrix} c & 0 \\ 0 & B \end{bmatrix} R$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & U^{-1} \end{bmatrix}\begin{bmatrix} c & 0 \\ 0 & B \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & U \end{bmatrix} = \begin{bmatrix} c & 0 \\ 0 & U^{-1}B \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & U \end{bmatrix}$$

$$= \begin{bmatrix} c & 0 \\ 0 & U^{-1}BU \end{bmatrix} = \begin{bmatrix} c & 0 \\ 0 & D \end{bmatrix},$$

so

$$A = Q\begin{bmatrix} c & 0 \\ 0 & D \end{bmatrix}Q^{-1} = Q\begin{bmatrix} c & 0 \\ 0 & D \end{bmatrix}Q^T = QCQ^T, \ldots$$

# Theorem 3.8.A (continued 3)

**Proof (continued).** For this matrix $U$, define $R = \begin{bmatrix} 1 & 0 \\ 0 & U \end{bmatrix}$ and notice that $R^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & U^{-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & U^T \end{bmatrix} = R^T$. Let $Q = PR$; notice $QQ^T = (PR)(PR)^T = PRR^TP^T = \mathcal{I}$, since $P$ and $R$ are orthogonal, and so $Q$ is orthogonal. Then

$$Q^{-1}AQ = (R^{-1}P^{-1})A(PR) = R^{-1}(P^{-1}AP)R = R^{-1}\begin{bmatrix} c & 0 \\ 0 & B \end{bmatrix} R$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & U^{-1} \end{bmatrix}\begin{bmatrix} c & 0 \\ 0 & B \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & U \end{bmatrix} = \begin{bmatrix} c & 0 \\ 0 & U^{-1}B \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & U \end{bmatrix}$$

$$= \begin{bmatrix} c & 0 \\ 0 & U^{-1}BU \end{bmatrix} = \begin{bmatrix} c & 0 \\ 0 & D \end{bmatrix},$$

so

$$A = Q\begin{bmatrix} c & 0 \\ 0 & D \end{bmatrix}Q^{-1} = Q\begin{bmatrix} c & 0 \\ 0 & D \end{bmatrix}Q^T = QCQ^T, \ldots$$

# Theorem 3.8.A (continued 4)

**Theorem 3.8.A.** A (real) $n \times n$ matrix $A$ is orthogonally diagonalizable if and only if $A$ is symmetric.

**Proof (continued).** ...where $Q$ is orthogonal and $C$ is diagonal; that is, $A$ is orthogonally diagonalizable and the result holds for $n \times n$ matrices. Therefore, by Mathematical Induction, the result holds for all square matrices. □

# Theorem 3.8.12

**Theorem 3.8.12.** If $A$ is an $n \times n$ diagonalizable matrix where $A = VCV^{-1}$ for diagonal $C$, then

(1) there are $n$ linearly independent eigenvectors of $A$,

(2) the number of nonzero eigenvalues of $A$ is equal to $\text{rank}(A)$.

**Proof. (1)** Since $A = VCV^{-1}$ then $AV = VC$ and so by Theorem 3.8.10 the $i$th column of $V$ is $v_i$ where $v_i$ is an eigenvector of $c_i$ where $C = \text{diag}(c_1, c_2, \ldots, c_n)$. Since $V$ is invertible then $V$ is full rank $n$ (see the definition of inverse matrix in Section 3.3) and so the dimension of the column space of $V$ is $n$ and hence the $n$ columns of $V$ are linearly independent.

# Theorem 3.8.12

**Theorem 3.8.12.** If $A$ is an $n \times n$ diagonalizable matrix where $A = VCV^{-1}$ for diagonal $C$, then

(1) there are $n$ linearly independent eigenvectors of $A$,

(2) the number of nonzero eigenvalues of $A$ is equal to rank($A$).

**Proof. (1)** Since $A = VCV^{-1}$ then $AV = VC$ and so by Theorem 3.8.10 the $i$th column of $V$ is $v_i$ where $v_i$ is an eigenvector of $c_i$ where $C = \text{diag}(c_1, c_2, \ldots, c_n)$. Since $V$ is invertible then $V$ is full rank $n$ (see the definition of inverse matrix in Section 3.3) and so the dimension of the column space of $V$ is $n$ and hence the $n$ columns of $V$ are linearly independent.

**(2)** Since $V$ is invertible then it is full rank $n$ and so by Theorem 3.3.9, there are matrices $P$ and $Q$, products of elementary matrices, such that $PVQ = I$.

# Theorem 3.8.12

**Theorem 3.8.12.** If $A$ is an $n \times n$ diagonalizable matrix where $A = VCV^{-1}$ for diagonal $C$, then

(1) there are $n$ linearly independent eigenvectors of $A$,

(2) the number of nonzero eigenvalues of $A$ is equal to $\text{rank}(A)$.

**Proof. (1)** Since $A = VCV^{-1}$ then $AV = VC$ and so by Theorem 3.8.10 the $i$th column of $V$ is $v_i$ where $v_i$ is an eigenvector of $c_i$ where $C = \text{diag}(c_1, c_2, \ldots, c_n)$. Since $V$ is invertible then $V$ is full rank $n$ (see the definition of inverse matrix in Section 3.3) and so the dimension of the column space of $V$ is $n$ and hence the $n$ columns of $V$ are linearly independent.

**(2)** Since $V$ is invertible then it is full rank $n$ and so by Theorem 3.3.9, there are matrices $P$ and $Q$, products of elementary matrices, such that $PVQ = I$.

# Theorem 3.8.12 (continued)

**Theorem 3.8.12.** If $A$ is an $n \times n$ diagonalizable matrix where $A = VCV^{-1}$ for diagonal $C$, then

(1) there are $n$ linearly independent eigenvectors of $A$,

(2) the number of nonzero eigenvalues of $A$ is equal to $\text{rank}(A)$.

**Proof.** Now each elementary matrix is invertible (in the notation of Section 3.2, see the note after Theorem 3.2.3, $E_{pq}^{-1} = E_{pq}$, $E_{sp}^{-1} = E_{(1/s)p}$, and $E_{psq}^{-1} = E_{p(-s)q}$) so a product of elementary matrices is invertible and hence $Q^{-1}$ exists so that $PV = Q^{-1}$ and similarly $V = P^{-1}Q^{-1}$. [We have shown that an invertible matrix is a product of elementary matrices.] That is, both $V$ and $V^{-1}$ are products of elementary matrices, so by Theorem 3.3.3, $\text{rank}(A) = \text{rank}(VCV^{-1}) = \text{rank}(C)$. Since $C = \text{diag}(c_1, c_2, \ldots, c_n)$ then the $i$th column of $C$ is $c_i e_i$ where $e_i$ is the $i$th standard basis vector of $\mathbb{R}^n$. So $\text{rank}(C)$ is the number of nonzero $c_i e_i$, which is the number of nonzero eigenvalues of $C$. □

# Theorem 3.8.12 (continued)

**Theorem 3.8.12.** If $A$ is an $n \times n$ diagonalizable matrix where $A = VCV^{-1}$ for diagonal $C$, then

(1) there are $n$ linearly independent eigenvectors of $A$,

(2) the number of nonzero eigenvalues of $A$ is equal to $\text{rank}(A)$.

**Proof.** Now each elementary matrix is invertible (in the notation of Section 3.2, see the note after Theorem 3.2.3, $E_{pq}^{-1} = E_{pq}$, $E_{sp}^{-1} = E_{(1/s)p}$, and $E_{psq}^{-1} = E_{p(-s)q}$) so a product of elementary matrices is invertible and hence $Q^{-1}$ exists so that $PV = Q^{-1}$ and similarly $V = P^{-1}Q^{-1}$. [We have shown that an invertible matrix is a product of elementary matrices.] That is, both $V$ and $V^{-1}$ are products of elementary matrices, so by Theorem 3.3.3, $\text{rank}(A) = \text{rank}(VCV^{-1}) = \text{rank}(C)$. Since $C = \text{diag}(c_1, c_2, \ldots, c_n)$ then the $i$th column of $C$ is $c_i e_i$ where $e_i$ is the $i$th standard basis vector of $\mathbb{R}^n$. So $\text{rank}(C)$ is the number of nonzero $c_i e_i$, which is the number of nonzero eigenvalues of $C$. □

# Theorem 3.8.13

**Theorem 3.8.13.** If $A$ is a symmetric matrix where $(c, v)$ is an eigenpair for $A$ with $v^T v = \|v\|^2 = 1$, then for any $k \in \mathbb{N}$ we have $(A - cvv^T)^k = A^k - c^k vv^T$.

**Proof.** We prove the result by induction. Of course it holds for $k = 1$. Suppose the result holds for $k - 1$ so that $(A - cvv^T)^{k-1} = A^{k-1} - c^{k-1} vv^T$.

# Theorem 3.8.13

**Theorem 3.8.13.** If $A$ is a symmetric matrix where $(c, v)$ is an eigenpair for $A$ with $v^T v = \|v\|^2 = 1$, then for any $k \in \mathbb{N}$ we have $(A - cvv^T)^k = A^k - c^k vv^T$.

**Proof.** We prove the result by induction. Of course it holds for $k = 1$. Suppose the result holds for $k - 1$ so that $(A - cvv^T)^{k-1} = A^{k-1} - c^{k-1}vv^T$. Then

$$\begin{aligned}
(A - cvv^T)^k &= (A^{k-1} - c^{k-1}vv^T)(A - cvv^T) \\
&= A^k - c^{k-1}vv^T A - cA^{k-1}vv^T + c^k vv^T vv^T \\
&= A^k - c^{k-1}vv^T A - c(c^{k-1}v)v^T + c^k vv^T \text{ since} \\
&\quad A^{k-1}v = A^{k-2}(cv) = c(A^{k-2}v) = \cdots = c^{k-1}v \\
&= A^k - c^{k-1}vv^T A - c^k vv^T + c^k vv^T \\
&= A^k - c^{k-1}v(A^T v)^T = A^k - c^{k-1}v(Av)^T \text{ since } A^T = A \\
&= A^k - c^{k-1}v(cv^T) = A^k - c^k vv^T.
\end{aligned}$$

So the result holds for $k$ and the result follows by Math Induction. □

# Theorem 3.8.13

**Theorem 3.8.13.** If $A$ is a symmetric matrix where $(c, v)$ is an eigenpair for $A$ with $v^T v = \|v\|^2 = 1$, then for any $k \in \mathbb{N}$ we have $(A - cvv^T)^k = A^k - c^k vv^T$.

**Proof.** We prove the result by induction. Of course it holds for $k = 1$. Suppose the result holds for $k - 1$ so that $(A - cvv^T)^{k-1} = A^{k-1} - c^{k-1}vv^T$. Then

$$\begin{aligned}
(A - cvv^T)^k &= (A^{k-1} - c^{k-1}vv^T)(A - cvv^T) \\
&= A^k - c^{k-1}vv^T A - cA^{k-1}vv^T + c^k vv^T vv^T \\
&= A^k - c^{k-1}vv^T A - c(c^{k-1}v)v^T + c^k vv^T \text{ since} \\
& \quad A^{k-1}v = A^{k-2}(cv) = c(A^{k-2}v) = \cdots = c^{k-1}v \\
&= A^k - c^{k-1}vv^T A - c^k vv^T + c^k vv^T \\
&= A^k - c^{k-1}v(A^T v)^T = A^k - c^{k-1}v(Av)^T \text{ since } A^T = A \\
&= A^k - c^{k-1}v(cv^T) = A^k - c^k vv^T.
\end{aligned}$$

So the result holds for $k$ and the result follows by Math Induction. □

# Theorem 3.8.14

**Theorem 3.8.14.** Any real symmetric matrix is positive definite if and only if all of its eigenvalues are positive. Any real symmetric matrix is nonnegative definite if and only if all of its eigenvalues are nonnegative.

**Proof.** By Theorem 3.8.A, $A$ is orthogonally diagonalizable so $A = VCV^{-1} = VCV^T$ where $V$ is orthogonal (so $V^{-1} = V^T$). So for any $x \in \mathbb{R}^n$ (where $A$ is $n \times n$) we have

$$x^T Ax = x^T(VCV^T)x = (x^T V)C(V^T x) = y^T Cy$$

where $y = V^T x$.

# Theorem 3.8.14

**Theorem 3.8.14.** Any real symmetric matrix is positive definite if and only if all of its eigenvalues are positive. Any real symmetric matrix is nonnegative definite if and only if all of its eigenvalues are nonnegative.

**Proof.** By Theorem 3.8.A, $A$ is orthogonally diagonalizable so $A = VCV^{-1} = VCV^T$ where $V$ is orthogonal (so $V^{-1} = V^T$). So for any $x \in \mathbb{R}^n$ (where $A$ is $n \times n$) we have

$$x^T Ax = x^T(VCV^T)x = (x^T V)C(V^T x) = y^T Cy$$

where $y = V^T x$. So, for $x \neq 0$, $x^T Ax > 0$ (notice $x^T Ax$ is $1 \times 1$; it is a quadratic form), that is, $A$ is positive definite (by definition), if and only if $y^T Cy > 0$. Now with the entries of $y$ as $y_i$ for $i = 1, 2, \ldots, n$, we have $y^T Cy = \sum_{i=1}^n (y_i)^2 c_i$. So if each $c_i > 0$ then $y^T Cy > 0$ and hence $A$ is positive definite.

# Theorem 3.8.14

**Theorem 3.8.14.** Any real symmetric matrix is positive definite if and only if all of its eigenvalues are positive. Any real symmetric matrix is nonnegative definite if and only if all of its eigenvalues are nonnegative.

**Proof.** By Theorem 3.8.A, $A$ is orthogonally diagonalizable so $A = VCV^{-1} = VCV^T$ where $V$ is orthogonal (so $V^{-1} = V^T$). So for any $x \in \mathbb{R}^n$ (where $A$ is $n \times n$) we have

$$x^T Ax = x^T(VCV^T)x = (x^T V)C(V^T x) = y^T Cy$$

where $y = V^T x$. So, for $x \neq 0$, $x^T Ax > 0$ (notice $x^T Ax$ is $1 \times 1$; it is a quadratic form), that is, $A$ is positive definite (by definition), if and only if $y^T Cy > 0$. Now with the entries of $y$ as $y_i$ for $i = 1, 2, \ldots, n$, we have $y^T Cy = \sum_{i=1}^n (y_i)^2 c_i$. So if each $c_i > 0$ then $y^T Cy > 0$ and hence $A$ is positive definite.

# Theorem 3.8.14 (continued)

**Theorem 3.8.14.** Any real symmetric matrix is positive definite if and only if all of its eigenvalues are positive. Any real symmetric matrix is nonnegative definite if and only if all of its eigenvalues are nonnegative.

**Proof (continued).** By choosing $y$ as the $i$th standard basis vector for $\mathbb{R}^n$ (or equivalently by choosing $x = Vy$ where $y$ is the $i$th standard basis vector for $\mathbb{R}^n$), we have $y^T Cy = c_i$. So if $A$ is positive definite then $x^T Ax = y^T Cy = c_i > 0$ and so each eigenvalue $c_i > 0$. The proof for nonnegative definite is similar. □

# Theorem 3.8.14 (continued)

**Theorem 3.8.14.** Any real symmetric matrix is positive definite if and only if all of its eigenvalues are positive. Any real symmetric matrix is nonnegative definite if and only if all of its eigenvalues are nonnegative.

**Proof (continued).** By choosing $y$ as the $i$th standard basis vector for $\mathbb{R}^n$ (or equivalently by choosing $x = Vy$ where $y$ is the $i$th standard basis vector for $\mathbb{R}^n$), we have $y^T Cy = c_i$. So if $A$ is positive definite then $x^T Ax = y^T Cy = c_i > 0$ and so each eigenvalue $c_i > 0$. The proof for nonnegative definite is similar. □

# Theorem 3.8.15

**Theorem 3.8.15.**

(1) If symmetric matrix $A$ is positive definite then there is nonsingular $P$ such that $P^T AP = I$.

(2) If symmetric matrix $A$ is nonnegative definite and $A = VCV^T$ where $V$ is orthogonal (such $V$ exists by Theorem 3.8.A) and $C = \text{diag}(c_1, c_2, \ldots, c_n)$ where the eigenvalues of $A$ are $c_1, c_2, \ldots, c_n$. Then there is diagonal nonnegative definite matrix $S$ such that $(VSV^T)^2 = A$.

**Proof.** **(1)** By Theorem 3.8.A, $A = VCV^T$ for orthogonal $V$ where $C = \text{diag}(c_1, c_2, \ldots, c_n)$ and the eigenvalues of $A$ are $c_1, c_2, \ldots, c_n$. Since $A$ is positive definite, by Theorem 3.8.14 each $c_i > 0$. Define $S = \text{diag}(\sqrt{c_1}, \sqrt{c_2}, \ldots, \sqrt{c_n})$.

# Theorem 3.8.15

**Theorem 3.8.15.**

(1) If symmetric matrix $A$ is positive definite then there is nonsingular $P$ such that $P^T AP = I$.

(2) If symmetric matrix $A$ is nonnegative definite and $A = VCV^T$ where $V$ is orthogonal (such $V$ exists by Theorem 3.8.A) and $C = \text{diag}(c_1, c_2, \ldots, c_n)$ where the eigenvalues of $A$ are $c_1, c_2, \ldots, c_n$. Then there is diagonal nonnegative definite matrix $S$ such that $(VSV^T)^2 = A$.

**Proof. (1)** By Theorem 3.8.A, $A = VCV^T$ for orthogonal $V$ where $C = \text{diag}(c_1, c_2, \ldots, c_n)$ and the eigenvalues of $A$ are $c_1, c_2, \ldots, c_n$. Since $A$ is positive definite, by Theorem 3.8.14 each $c_i > 0$. Define $S = \text{diag}(\sqrt{c_1}, \sqrt{c_2}, \ldots, \sqrt{c_n})$. Then $S^2 = C$ and so $A = VS^2V^T = VSSV^T = VSS^TV^T = VS(VS)^T$. Now $V$ is orthogonal and the $i$th column of $VS$ is $\sqrt{c_i}$ times the $i$th column of $V$ (where $c_i > 0$), so the columns of $VS$ are linearly independent and $VS$ is full rank and hence (by definition) invertible.

# Theorem 3.8.15

**Theorem 3.8.15.**

(1) If symmetric matrix $A$ is positive definite then there is nonsingular $P$ such that $P^T AP = I$.

(2) If symmetric matrix $A$ is nonnegative definite and $A = VCV^T$ where $V$ is orthogonal (such $V$ exists by Theorem 3.8.A) and $C = \text{diag}(c_1, c_2, \ldots, c_n)$ where the eigenvalues of $A$ are $c_1, c_2, \ldots, c_n$. Then there is diagonal nonnegative definite matrix $S$ such that $(VSV^T)^2 = A$.

**Proof. (1)** By Theorem 3.8.A, $A = VCV^T$ for orthogonal $V$ where $C = \text{diag}(c_1, c_2, \ldots, c_n)$ and the eigenvalues of $A$ are $c_1, c_2, \ldots, c_n$. Since $A$ is positive definite, by Theorem 3.8.14 each $c_i > 0$. Define $S = \text{diag}(\sqrt{c_1}, \sqrt{c_2}, \ldots, \sqrt{c_n})$. Then $S^2 = C$ and so $A = VS^2V^T = VSSV^T = VSS^TV^T = VS(VS)^T$. Now $V$ is orthogonal and the $i$th column of $VS$ is $\sqrt{c_i}$ times the $i$th column of $V$ (where $c_i > 0$), so the columns of $VS$ are linearly independent and $VS$ is full rank and hence (by definition) invertible.

# Theorem 3.8.15 (continued)

**Theorem 3.8.15.**

(1) If symmetric matrix $A$ is positive definite then there is nonsingular $P$ such that $P^T AP = I$.

(2) If symmetric matrix $A$ is nonnegative definite and $A = VCV^T$ where $V$ is orthogonal (such $V$ exists by Theorem 3.8.A) and $C = \text{diag}(c_1, c_2, \ldots, c_n)$ where the eigenvalues of $A$ are $c_1, c_2, \ldots, c_n$. Then there is diagonal nonnegative definite matrix $S$ such that $(VSV^T)^2 = A$.

**Proof (continued). (1)** So $(VS)^{-1}A((VS)^T)^{-1} = (VS)^{-1}A((VS)^{-1})^T$ by Theorem 3.3.7. With $P = ((VS)^{-1})^T$, the claim follows.

# Theorem 3.8.15 (continued)

**Theorem 3.8.15.**

(1) If symmetric matrix $A$ is positive definite then there is nonsingular $P$ such that $P^T AP = I$.

(2) If symmetric matrix $A$ is nonnegative definite and $A = VCV^T$ where $V$ is orthogonal (such $V$ exists by Theorem 3.8.A) and $C = \text{diag}(c_1, c_2, \ldots, c_n)$ where the eigenvalues of $A$ are $c_1, c_2, \ldots, c_n$. Then there is diagonal nonnegative definite matrix $S$ such that $(VSV^T)^2 = A$.

**Proof (continued). (1)** So $(VS)^{-1}A((VS)^T)^{-1} = (VS)^{-1}A((VS)^{-1})^T$ by Theorem 3.3.7. With $P = ((VS)^{-1})^T$, the claim follows.

**(2)** Similar to the proof of part (1), we take $S = \text{diag}(\sqrt{c_1}, \sqrt{c_2}, \ldots, \sqrt{c_n})$. Then $S$ is diagonal with nonnegative eigenvalues $\sqrt{c_1}, \sqrt{c_2}, \ldots, \sqrt{c_n}$ and so $S$ is nonnegative definite by Theorem 3.8.14. Also $S^2 = C$ and so (since $V$ is orthogonal and $V^{-1} = V^T$):

$$A = VCV^T = VS^2V^T = VSISV^T = VSV^TVSV^T = (VSV^T)^2. \quad \square$$

# Theorem 3.8.15 (continued)

**Theorem 3.8.15.**

(1) If symmetric matrix $A$ is positive definite then there is nonsingular $P$ such that $P^T AP = I$.

(2) If symmetric matrix $A$ is nonnegative definite and $A = VCV^T$ where $V$ is orthogonal (such $V$ exists by Theorem 3.8.A) and $C = \text{diag}(c_1, c_2, \ldots, c_n)$ where the eigenvalues of $A$ are $c_1, c_2, \ldots, c_n$. Then there is diagonal nonnegative definite matrix $S$ such that $(VSV^T)^2 = A$.

**Proof (continued). (1)** So $(VS)^{-1}A((VS)^T)^{-1} = (VS)^{-1}A((VS)^{-1})^T$ by Theorem 3.3.7. With $P = ((VS)^{-1})^T$, the claim follows.

**(2)** Similar to the proof of part (1), we take $S = \text{diag}(\sqrt{c_1}, \sqrt{c_2}, \ldots, \sqrt{c_n})$. Then $S$ is diagonal with nonnegative eigenvalues $\sqrt{c_1}, \sqrt{c_2}, \ldots, \sqrt{c_n}$ and so $S$ is nonnegative definite by Theorem 3.8.14. Also $S^2 = C$ and so (since $V$ is orthogonal and $V^{-1} = V^T$):

$$A = VCV^T = VS^2V^T = VSISV^T = VSV^TVSV^T = (VSV^T)^2. \quad \square$$

# Theorem 3.8.16

**Theorem 3.8.16.** Let $A$ be an $n \times m$ matrix. Then there exists a singular value decomposition of $A$.

**Proof.** First, matrix $A^T A$ is a $m \times m$ symmetric matrix which is nonnegative definite by Theorem 3.3.14(2) and so by Theorem 3.8.14 the eigenvalues of $A^T A$ are nonnegative. By Theorem 3.8.A, $A^T A$ is orthogonally diagonalizable so there is $m \times m$ orthogonal $Q$ such that $A^T A = QCQ^T$ where $C = \text{diag}(c_1, c_2, \ldots, c_n)$ where $c_1 \geq c_2 \geq \cdots \geq c_n \geq 0$ are the eigenvalues of $A^T A$.

# Theorem 3.8.16

**Theorem 3.8.16.** Let $A$ be an $n \times m$ matrix. Then there exists a singular value decomposition of $A$.

**Proof.** First, matrix $A^T A$ is a $m \times m$ symmetric matrix which is nonnegative definite by Theorem 3.3.14(2) and so by Theorem 3.8.14 the eigenvalues of $A^T A$ are nonnegative. By Theorem 3.8.A, $A^T A$ is orthogonally diagonalizable so there is $m \times m$ orthogonal $Q$ such that $A^T A = QCQ^T$ where $C = \text{diag}(c_1, c_2, \ldots, c_n)$ where $c_1 \geq c_2 \geq \cdots \geq c_n \geq 0$ are the eigenvalues of $A^T A$. Let $r = \text{rank}(A)$. By Theorem 3.3.14(6), $\text{rank}(A^T A) = r$ and by Theorem 3.8.12, $r$ is the number of nonzero eigenvalues of $A^T A$. Define the $r \times r$ diagonal matrix of rank $r$, $D_1 = \text{diag}(\sqrt{c_1}, \sqrt{c_2}, \ldots, \sqrt{c_r})$. Since $D_1$ is full rank then $D_1^{-1}$ exists.

# Theorem 3.8.16

**Theorem 3.8.16.** Let $A$ be an $n \times m$ matrix. Then there exists a singular value decomposition of $A$.

**Proof.** First, matrix $A^T A$ is a $m \times m$ symmetric matrix which is nonnegative definite by Theorem 3.3.14(2) and so by Theorem 3.8.14 the eigenvalues of $A^T A$ are nonnegative. By Theorem 3.8.A, $A^T A$ is orthogonally diagonalizable so there is $m \times m$ orthogonal $Q$ such that $A^T A = QCQ^T$ where $C = \text{diag}(c_1, c_2, \ldots, c_n)$ where $c_1 \geq c_2 \geq \cdots \geq c_n \geq 0$ are the eigenvalues of $A^T A$. Let $r = \text{rank}(A)$. By Theorem 3.3.14(6), $\text{rank}(A^T A) = r$ and by Theorem 3.8.12, $r$ is the number of nonzero eigenvalues of $A^T A$. Define the $r \times r$ diagonal matrix of rank $r$, $D_1 = \text{diag}(\sqrt{c_1}, \sqrt{c_2}, \ldots, \sqrt{c_r})$. Since $D_1$ is full rank then $D_1^{-1}$ exists.

# Theorem 3.8.16 (continued 1)

**Proof (continued).** Partition $Q$ as $Q = [Q_1 \;\; Q_2]$ where $Q_1$ is $m \times r$. Now define $n \times r$ matrix $P_1$ as $P_1 = AQD_1^{-1}$ and let $P_2$ be *any* $n \times (n-r)$ matrix such that $P_1^T P_2 = 0$ (where 0 is the $r \times (n-r)$ zero matrix; one such choice for $P_2$ is the $n \times (n-r)$ zero matrix but we make a particular choice of $P_2$ later). Create $n \times n$ matrix $P$ as $P = [P_1 \;\; P_2]$.

Notice that $A^T A = QCQ^T$ implies $Q^T A^T AQ = C = \begin{bmatrix} D_1^2 & 0 \\ 0 & 0 \end{bmatrix}$. Also

$$Q^T A^T AQ = \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} A^T A [Q_1 \;\; Q_2] = \begin{bmatrix} Q_1^T A^T AQ_1 & Q_1^T A^T AQ_2 \\ Q_2^T A^T AQ_1 & Q_2^T A^T AQ_2 \end{bmatrix}$$

where $Q_1^T A^T AQ_1$ is $r \times r$.

# Theorem 3.8.16 (continued 1)

**Proof (continued).** Partition $Q$ as $Q = [Q_1 \ \ Q_2]$ where $Q_1$ is $m \times r$. Now define $n \times r$ matrix $P_1$ as $P_1 = AQD_1^{-1}$ and let $P_2$ be *any* $n \times (n-r)$ matrix such that $P_1^T P_2 = 0$ (where 0 is the $r \times (n-r)$ zero matrix; one such choice for $P_2$ is the $n \times (n-r)$ zero matrix but we make a particular choice of $P_2$ later). Create $n \times n$ matrix $P$ as $P = [P_1 \ \ P_2]$.

Notice that $A^T A = QCQ^T$ implies $Q^T A^T A Q = C = \begin{bmatrix} D_1^2 & 0 \\ 0 & 0 \end{bmatrix}$. Also

$$Q^T A^T A Q = \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} A^T A [Q_1 \ \ Q_2] = \begin{bmatrix} Q_1^T A^T A Q_1 & Q_1^T A^T A Q_2 \\ Q_2^T A^T A Q_1 & Q_2^T A^T A Q_2 \end{bmatrix}$$

where $Q_1^T A^T A Q_1$ is $r \times r$. So $Q_1^T A^T A Q_1 = D_1^2$ and $Q_2^T A^T A Q_2 = (AQ_2)^T A Q_2 = 0$. The second equation implies $AQ_2 = 0$ by Theorem 3.3.14(1). Now $P_1 = AQ_1 D_1^{-1}$ by definition, so $P_1^T = D_1^{-1} Q_1^T A^T$ and hence $Q_1^T A^T = D_1 P_1^T$ or $AQ_1 = P_1 D_1$.

# Theorem 3.8.16 (continued 1)

**Proof (continued).** Partition $Q$ as $Q = [Q_1 \ \ Q_2]$ where $Q_1$ is $m \times r$. Now define $n \times r$ matrix $P_1$ as $P_1 = AQD_1^{-1}$ and let $P_2$ be *any* $n \times (n-r)$ matrix such that $P_1^T P_2 = 0$ (where 0 is the $r \times (n-r)$ zero matrix; one such choice for $P_2$ is the $n \times (n-r)$ zero matrix but we make a particular choice of $P_2$ later). Create $n \times n$ matrix $P$ as $P = [P_1 \ \ P_2]$.

Notice that $A^T A = QCQ^T$ implies $Q^T A^T AQ = C = \begin{bmatrix} D_1^2 & 0 \\ 0 & 0 \end{bmatrix}$. Also

$$Q^T A^T AQ = \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} A^T A [Q_1 \ \ Q_2] = \begin{bmatrix} Q_1^T A^T AQ_1 & Q_1^T A^T AQ_2 \\ Q_2^T A^T AQ_1 & Q_2^T A^T AQ_2 \end{bmatrix}$$

where $Q_1^T A^T AQ_1$ is $r \times r$. So $Q_1^T A^T AQ_1 = D_1^2$ and $Q_2^T A^T AQ_2 = (AQ_2)^T AQ_2 = 0$. The second equation implies $AQ_2 = 0$ by Theorem 3.3.14(1). Now $P_1 = AQ_1 D_1^{-1}$ by definition, so $P_1^T = D_1^{-1} Q_1^T A^T$ and hence $Q_1^T A^T = D_1 P_1^T$ or $AQ_1 = P_1 D_1$.

# Theorem 3.8.16 (continued 2)

**Proof (continued).** So

$$\begin{aligned}
P^TAQ &= \begin{bmatrix} P_1^T \\ P_2^T \end{bmatrix} A[Q_1 \;\; Q_2] = \begin{bmatrix} P_1^TAQ_1 & P_1^TAQ_2 \\ P_2^TAQ_1 & P_2^TAQ_2 \end{bmatrix} \\
&= \begin{bmatrix} (D_1^{-1}Q_1^TA^T)AQ_1 & P_1^T(0) \\ P_2^T(P_1D_1) & P_2^T(0) \end{bmatrix} \text{ since } P_1^T = D_1^{-1}Q_1^TA^T, \\
&\quad AQ_1 = P_1D_1, \text{ and } AQ_2 = 0 \\
&= \begin{bmatrix} D_1^{-1}(D_1^2) & 0 \\ (P_1^TP_2)^TD_1 & 0 \end{bmatrix} \text{ since } Q_1^TA^TAQ_1 = D_1^2 \\
&= \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} \text{ since } P_1^TP_2 = 0. \qquad (*)
\end{aligned}$$

Notice that $P^TAQ$ is an $n \times m$ matrix. Now

$$P_1^TP_1 = (D_1^{-1}Q_1^TA^T)(D_1^{-1}Q_1^TA^T)^T \text{ since } P_1^T = D_1^{-1}Q_1^TA^T$$

$$= D_1^{-1}Q_1^TA^TAQ_1D_1^{-1} = D_1^{-1}D_1^2D_1^{-1}(\text{ since } Q_1^TA^TAQ_1 = D_1^2) = \mathcal{I}_r,$$

# Theorem 3.8.16 (continued 2)

**Proof (continued).** So

$$
\begin{aligned}
P^TAQ &= \begin{bmatrix} P_1^T \\ P_2^T \end{bmatrix} A[Q_1 \;\; Q_2] = \begin{bmatrix} P_1^TAQ_1 & P_1^TAQ_2 \\ P_2^TAQ_1 & P_2^TAQ_2 \end{bmatrix} \\
&= \begin{bmatrix} (D_1^{-1}Q_1^TA^T)AQ_1 & P_1^T(0) \\ P_2^T(P_1D_1) & P_2^T(0) \end{bmatrix} \text{ since } P_1^T = D_1^{-1}Q_1^TA^T, \\
&\quad AQ_1 = P_1D_1, \text{ and } AQ_2 = 0 \\
&= \begin{bmatrix} D_1^{-1}(D_1^2) & 0 \\ (P_1^TP_2)^TD_1 & 0 \end{bmatrix} \text{ since } Q_1^TA^TAQ_1 = D_1^2 \\
&= \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} \text{ since } P_1^TP_2 = 0. \qquad (*)
\end{aligned}
$$

Notice that $P^TAQ$ is an $n \times m$ matrix. Now

$$P_1^TP_1 = (D_1^{-1}Q_1^TA^T)(D_1^{-1}Q_1^TA^T)^T \text{ since } P_1^T = D_1^{-1}Q_1^TA^T$$

$$= D_1^{-1}Q_1^TA^TAQ_1D_1^{-1} = D_1^{-1}D_1^2D_1^{-1}(\text{ since } Q_1^TA^TAQ_1 = D_1^2) = \mathcal{I}_r,$$

# Theorem 3.8.16 (continued 3)

**Proof (continued).** ...so by Theorem 3.7.1, $P_1$ is orthogonal. By Theorem 3.5.4, $\dim(\mathcal{N}(P_1^T)) = n - \text{rank}(P_1^T)$. By Theorem 3.3.14(6), $\text{rank}(P_1^T P_1) = \text{rank}(P_1)$ and by Theorem 3.3.2, $\text{rank}(P_1) = \text{rank}(P_1^T)$. So $\text{rank}(P_1^T) = \text{rank}(P_1^T P_1) = \text{rank}(\mathcal{I}_r) = r$. Hence $\dim(\mathcal{N}(P_1^T)) = n - \text{rank}(P_1^T) = n - r$. Let $P_2$ be any $n \times (n-r)$ matrix whose columns form an orthonormal basis of $\mathcal{N}(P_1^T)$. Then $P_1^T P_2 = 0$ as required above and $P_2^T P_2 = I_{n-r}$ since $P_2$ is orthogonal (Theorem 3.7.1). So

$$P^T P = \begin{bmatrix} P_1^T \\ P_2^T \end{bmatrix} [P_1 \ \ P_2] = \begin{bmatrix} P_1^T P_1 & P_1^T P_2 \\ P_2^T P_1 & P_2^T P_2 \end{bmatrix} = \begin{bmatrix} \mathcal{I}_r & 0 \\ 0 & \mathcal{I}_{n-r} \end{bmatrix} = \mathcal{I}_n,$$

and $P$ is orthogonal (Theorem 3.7.1).

# Theorem 3.8.16 (continued 3)

**Proof (continued).** ...so by Theorem 3.7.1, $P_1$ is orthogonal. By Theorem 3.5.4, $\dim(\mathcal{N}(P_1^T)) = n - \text{rank}(P_1^T)$. By Theorem 3.3.14(6), $\text{rank}(P_1^T P_1) = \text{rank}(P_1)$ and by Theorem 3.3.2, $\text{rank}(P_1) = \text{rank}(P_1^T)$. So $\text{rank}(P_1^T) = \text{rank}(P_1^T P_1) = \text{rank}(\mathcal{I}_r) = r$. Hence $\dim(\mathcal{N}(P_1^T)) = n - \text{rank}(P_1^T) = n - r$. Let $P_2$ be any $n \times (n-r)$ matrix whose columns form an orthonormal basis of $\mathcal{N}(P_1^T)$. Then $P_1^T P_2 = 0$ as required above and $P_2^T P_2 = I_{n-r}$ since $P_2$ is orthogonal (Theorem 3.7.1). So

$$P^T P = \begin{bmatrix} P_1^T \\ P_2^T \end{bmatrix} [P_1 \;\; P_2] = \begin{bmatrix} P_1^T P_1 & P_1^T P_2 \\ P_2^T P_1 & P_2^T P_2 \end{bmatrix} = \begin{bmatrix} \mathcal{I}_r & 0 \\ 0 & \mathcal{I}_{n-r} \end{bmatrix} = \mathcal{I}_n,$$

and $P$ is orthogonal (Theorem 3.7.1). By $(*)$, $P^T AQ = \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} := D$, or $A = PDQ^T$ where $P$ is an $n \times n$ orthogonal matrix and $Q$ is an $m \times m$ orthogonal matrix. With $U = P$ and $V = Q$, we see that $A$ has a singular value decomposition, as claimed. □

# Theorem 3.8.16 (continued 3)

**Proof (continued).** ...so by Theorem 3.7.1, $P_1$ is orthogonal. By Theorem 3.5.4, $\dim(\mathcal{N}(P_1^T)) = n - \text{rank}(P_1^T)$. By Theorem 3.3.14(6), $\text{rank}(P_1^T P_1) = \text{rank}(P_1)$ and by Theorem 3.3.2, $\text{rank}(P_1) = \text{rank}(P_1^T)$. So $\text{rank}(P_1^T) = \text{rank}(P_1^T P_1) = \text{rank}(\mathcal{I}_r) = r$. Hence $\dim(\mathcal{N}(P_1^T)) = n - \text{rank}(P_1^T) = n - r$. Let $P_2$ be any $n \times (n-r)$ matrix whose columns form an orthonormal basis of $\mathcal{N}(P_1^T)$. Then $P_1^T P_2 = 0$ as required above and $P_2^T P_2 = I_{n-r}$ since $P_2$ is orthogonal (Theorem 3.7.1). So

$$P^T P = \begin{bmatrix} P_1^T \\ P_2^T \end{bmatrix} [P_1 \ \ P_2] = \begin{bmatrix} P_1^T P_1 & P_1^T P_2 \\ P_2^T P_1 & P_2^T P_2 \end{bmatrix} = \begin{bmatrix} \mathcal{I}_r & 0 \\ 0 & \mathcal{I}_{n-r} \end{bmatrix} = \mathcal{I}_n,$$

and $P$ is orthogonal (Theorem 3.7.1). By $(*)$, $P^T A Q = \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} := D$, or $A = PDQ^T$ where $P$ is an $n \times n$ orthogonal matrix and $Q$ is an $m \times m$ orthogonal matrix. With $U = P$ and $V = Q$, we see that $A$ has a singular value decomposition, as claimed. □

# Theorem 3.8.17

**Theorem 3.8.17.** Let $A$ be an $n \times m$ matrix with spectral decomposition $A = UDV^T = \sum_{i=1}^{r} d_i u_i v_i^T$. Then $\langle u_i v_i^T, u_j v_j^T \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$ and $d_i = \langle A, u_i v_i^T \rangle$. That is, the spectral decomposition is a Fourier expansion of $A$.

**Proof.** Recall for matrices, $\langle A, B \rangle = \sum_{k=1}^{m} a_k^T b_k = \sum_{k=1}^{n} \langle a_k, b_k \rangle$ (see Section 3.2), and the $k$th column of $u_i v_i^T$ is $\begin{bmatrix} u_i^1 v_i^k \\ u_i^2 v_i^k \\ \vdots \\ u_i^n v_i^k \end{bmatrix}$ where we use superscripts to indicate entries in a column vector. Notice that $u_i v_i^T$ is $n \times m$ and so has $m$ columns, each of length $n$.

# Theorem 3.8.17

**Theorem 3.8.17.** Let $A$ be an $n \times m$ matrix with spectral decomposition $A = UDV^T = \sum_{i=1}^r d_i u_i v_i^T$. Then $\langle u_i v_i^T, u_j v_j^T \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$ and $d_i = \langle A, u_i v_i^T \rangle$. That is, the spectral decomposition is a Fourier expansion of $A$.

**Proof.** Recall for matrices, $\langle A, B \rangle = \sum_{k=1}^m a_k^T b_k = \sum_{k=1}^n \langle a_k, b_k \rangle$ (see Section 3.2), and the $k$th column of $u_i v_i^T$ is $\begin{bmatrix} u_i^1 v_i^k \\ u_i^2 v_i^k \\ \vdots \\ u_i^n v_i^k \end{bmatrix}$ where we use superscripts to indicate entries in a column vector. Notice that $u_i v_i^T$ is $n \times m$ and so has $m$ columns, each of length $n$.

# Theorem 3.8.17 (continued 1)

**Proof (continued).** So

$$\langle u_i v_i^T, u_i v_i^T \rangle = \sum_{k=1}^{m} \langle [u_i^1 v_i^k, u_i^2 v_i^k, \ldots, u_i^n v_i^k]^T, [u_i^1 v_i^k, u_i^2 v_i^k, \ldots, u_i^n v_i^k]^T \rangle$$

$$= \sum_{k=1}^{m} \left( (u_i^1 v_i^k)^2 + (u_i^2 v_i^k)^2 + \cdots + (u_i^n v_i^k)^2 \right)$$

$$= \left( (u_i^1)^2 + (u_i^2)^2 + \cdots + (u_i^n)^2 \right) \sum_{k=1}^{m} (v_i^k)^2 = \|u_i\|^2 \|v_i\|^2 = 1.$$

Next,

$$\langle u_i v_i^T, u_j v_j^T \rangle = \sum_{k=1}^{m} \langle [u_i^1 v_i^k, u_i^2 v_i^k, \ldots, u_i^n v_i^k]^T, [u_j^1 v_j^k, u_j^2 v_j^k, \ldots, u_j^n v_j^k]^T \rangle$$

$$= \sum_{k=1}^{m} \left( u_i^1 v_i^k u_j^1 v_j^k + u_i^2 v_i^k u_j^2 v_j^k + \cdots + u_i^n v_i^k u_j^n v_j^k \right)$$

# Theorem 3.8.17 (continued 1)

**Proof (continued).** So

$$\langle u_i v_i^T, u_i v_i^T \rangle = \sum_{k=1}^{m} \langle [u_i^1 v_i^k, u_i^2 v_i^k, \ldots, u_i^n v_i^k]^T, [u_i^1 v_i^k, u_i^2 v_i^k, \ldots, u_i^n v_i^k]^T \rangle$$

$$= \sum_{k=1}^{m} \left( (u_i^1 v_i^k)^2 + (u_i^2 v_i^k)^2 + \cdots + (u_i^n v_i^k)^2 \right)$$

$$= \left( (u_i^1)^2 + (u_i^2)^2 + \cdots + (u_i^n)^2 \right) \sum_{k=1}^{m} (v_i^k)^2 = \|u_i\|^2 \|v_i\|^2 = 1.$$

Next,

$$\langle u_i v_i^T, u_j v_j^T \rangle = \sum_{k=1}^{m} \langle [u_i^1 v_i^k, u_i^2 v_i^k, \ldots, u_i^n v_i^k]^T, [u_j^1 v_j^k, u_j^2 v_j^k, \ldots, u_j^n v_j^k]^T \rangle$$

$$= \sum_{k=1}^{m} \left( u_i^1 v_i^k u_j^1 v_j^k + u_i^2 v_i^k u_j^2 v_j^k + \cdots + u_i^n v_i^k u_j^n v_j^k \right)$$

# Theorem 3.8.17 (continued 2)

**Theorem 3.8.17.** Let $A$ be an $n \times m$ matrix with spectral decomposition $A = UDV^T = \sum_{i=1}^r d_i u_i v_i^T$. Then $\langle u_i v_i^T, u_j v_j^T \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$ and $d_i = \langle A, u_i v_i^T \rangle$. That is, the spectral decomposition is a Fourier expansion of $A$.

**Proof (continued).** . . .

$$= \sum_{k=1}^{m} \left( u_i^1 v_i^k u_j^1 v_j^k + u_i^2 v_i^k u_j^2 v_j^k + \cdots + u_i^n v_i^k u_j^n v_j^k \right)$$

$$= \sum_{k=1}^{m} v_i^k v_j^k (u_i^1 u_j^1 + u_i^2 u_j^2 + \cdots + u_i^n u_j^n) = \sum_{k=1}^{m} v_i^k v_j^k \langle u_i, u_j \rangle = 0.$$

The proof that $d_i = \langle A, u_i v_i^T \rangle$ is left as a Exercise 3.8.D. □

# Theorem 3.8.17 (continued 2)

**Theorem 3.8.17.** Let $A$ be an $n \times m$ matrix with spectral decomposition $A = UDV^T = \sum_{i=1}^{r} d_i u_i v_i^T$. Then $\langle u_i v_i^T, u_j v_j^T \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$ and $d_i = \langle A, u_i v_i^T \rangle$. That is, the spectral decomposition is a Fourier expansion of $A$.

**Proof (continued).** ...

$$= \sum_{k=1}^{m} \left( u_i^1 v_i^k u_j^1 v_j^k + u_i^2 v_i^k u_j^2 v_j^k + \cdots + u_i^n v_i^k u_j^n v_j^k \right)$$

$$= \sum_{k=1}^{m} v_i^k v_j^k (u_i^1 u_j^1 + u_i^2 u_j^2 + \cdots + u_i^n u_j^n) = \sum_{k=1}^{m} v_i^k v_j^k \langle u_i, u_j \rangle = 0.$$

The proof that $d_i = \langle A, u_i v_i^T \rangle$ is left as a Exercise 3.8.D. □