**MA1508E: LINEAR ALGEBRA FOR ENGINEERING**

**Lecture 11 Notes**

# References

1. Elementary Linear Algebra: Application Version, Section 5.2, 5.4, 7.1-7.2
2. Linear Algebra with Application, Section 3.3-3.4, 5.5, 8.2

## 6.2 Diagonalization

A square matrix $\mathbf{A}$ is said to be diagonalizable if there exists an invertible matrix $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$ is a diagonal matrix.

**Remark.** The statement above is equivalent to being able to express $\mathbf{A}$ as $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ for some invertible $\mathbf{P}$ and diagonal matrix $\mathbf{D}$.

**Example.**

1. Any square zero matrix is diagonalizable, $\mathbf{0} = \mathbf{I}\mathbf{0}\mathbf{I}^{-1}$.
2. Any diagonal matrix $\mathbf{D}$ is diagonalizable, $\mathbf{D} = \mathbf{I}\mathbf{D}\mathbf{I}^{-1}$.
3. $\mathbf{A} = \begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix}$ is diagonalizable, with $\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}^{-1}$.

Suppose $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$, for some matrix $\mathbf{D}$ and invertible matrix $\mathbf{P}$. Then the characteristic polynomial of $\mathbf{A}$ is

$$\begin{aligned} \det(x\mathbf{I} - \mathbf{A}) &= \det(x\mathbf{I} - \mathbf{P}\mathbf{D}\mathbf{P}^{-1}) = \det(\mathbf{P}x\mathbf{P}^{-1} - \mathbf{P}\mathbf{D}\mathbf{P}^{-1}) \\ &= \det(\mathbf{P}(x\mathbf{I} - \mathbf{D})\mathbf{P}^{-1}) = \det(\mathbf{P})\det(x\mathbf{I} - \mathbf{D})\det(\mathbf{P}^{-1}) \\ &= \det(\mathbf{P})\det(\mathbf{P})^{-1}\det(x\mathbf{I} - \mathbf{D}) = \det(x\mathbf{I} - \mathbf{D}). \end{aligned} \tag{6}$$

In other words, the characteristics polynomail of $\mathbf{A}$ is the same as $\mathbf{D}$. Furthermore, if $\mathbf{D}$ is a diagonal matrix, then the eigenvalues of $\mathbf{A}$ are exactly the diagonal entries of $\mathbf{D}$, with algebraic multiplicity equals to the number of diagonal entries taking the value. Hence, if $\mathbf{A}$ is diagonalizable, $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$, then the diagonal entries of the matrix $\mathbf{D}$ are exactly the eigenvalues of $\mathbf{A}$, appearing their algebraic multiplicities times. The invertible matrix $\mathbf{P}$ is constructed from the eigenvectors of $\mathbf{A}$.

**Theorem.** *Let $\mathbf{A}$ be a square matrix of order $n$. $\mathbf{A}$ is diagonalizable if and only if tdiag()here exists a basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ of $\mathbb{R}^n$ of eigenvectors of $\mathbf{A}$.*

The invertible matrix $\mathbf{P}$ that diagonalizes $\mathbf{A}$ have the form $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$, where $\mathbf{u}_i$ are eigenvectors of $\mathbf{A}$, and the diagonal matrix is $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$, where $\lambda_i$ is the eigenvalue associated to eigenvector $\mathbf{u}_i$. In other words, the $i$-th column of the invertible matrix $\mathbf{P}$ is an eigenvector of $\mathbf{A}$ with eigenvalue the $i$-th diagonal entry of $\mathbf{D}$. For the details of the proof, readers may refer to the appendix.

**Theorem** (Geometric Multiplicity is no greater than Algebraic multiplicity)**.** *The geometric multiplicity of an eigenvalue $\lambda$ of a square matrix $\mathbf{A}$ is no greater than the algebraic multiplicity, that is,*

$$1 \le \dim(E_\lambda) \le r_\lambda.$$

**Theorem.** *Suppose* $\mathbf{A}$ *is a square matrix such that its characteristic polynomial can be written as a product of linear factors,*

$$\det(x\mathbf{I}-\mathbf{A}) = (x-\lambda_1)^{r_{\lambda_1}}(x-\lambda_2)^{r_{\lambda_2}}\cdots(x-\lambda_k)^{r_{\lambda_k}},$$

*where* $r_{\lambda_i}$ *is the algebraic multiplicity of* $\lambda_i$*, for* $i = 1, ..., k$*, and the eigenvalues are distinct,* $\lambda_i \neq \lambda_j$ *for all* $i \neq j$*. Then* $\mathbf{A}$ *is diagonalizable if and only if for each eigenvalue of* $\mathbf{A}$*, its geometric multiplicity is equal to its algebraic multiplicity,*

$$\dim(E_{\lambda_i}) = r_{\lambda_i}$$

*for all eigenvalues* $\lambda_i$ *of* $\mathbf{A}$*.*

In other words, there are two obstructions to a matrix $\mathbf{A}$ being diagonalizable.

(i) The characteristic polynomial of $\mathbf{A}$ do not split into linear factors.

(ii) There is an eigenvalue of $\mathbf{A}$ where the geometric multiplicity is strictly less than the algebraic multiplicity, $\dim(E_{\lambda_i}) < r_{\lambda_i}$.

For in either cases, there will not be enough linearly independent eigenvectors to form a basis for $\mathbb{R}^n$.

**Corollary.** *If* $\mathbf{A}$ *is a square matrix of order* $n$ *with* $n$ *distinct eigenvalues, then* $\mathbf{A}$ *is diagonalizable.*

*Proof.* If $\mathbf{A}$ has $n$ distinct eigenvalues, then the algebraic multiplicity of each eigenvalue must be 1. Thus

$$1 \le \dim(E_\lambda) \le r_\lambda = 1 \Rightarrow \dim(E_\lambda) = 1 = r_\lambda$$

for every eigenvalue $\lambda$ of $\mathbf{A}$. Therefore $\mathbf{A}$ is diagonalizable. □

### Algorithm to diagonalization

(i) Compute the characteristic polynomial of $\mathbf{A}$

$$\det(x\mathbf{I}-\mathbf{A}).$$

If it cannot be factorized into linear factors, then $\mathbf{A}$ is not diagonalizable.

(ii) Otherwise, write

$$\det(x\mathbf{I}-\mathbf{A}) = (x-\lambda_1)^{r_{\lambda_1}}(x-\lambda_2)^{r_{\lambda_2}}\cdots(x-\lambda_k)^{r_{\lambda_k}},$$

where $r_{\lambda_i}$ is the algebraic multiplicity of $\lambda_i$, for $i = 1, ..., k$, and the eigenvalues are distinct, $\lambda_i \neq \lambda_j$ for all $i \neq j$. For each eigenvalue $\lambda_i$ of $\mathbf{A}$, $i = 1, ..., k$, find a basis for the eigenspace, that is, find the solution space of the following linear system,

$$(\lambda_i\mathbf{I}-\mathbf{A})\mathbf{x} = \mathbf{0}.$$

If there is a $i$ such that $\dim(E_{\lambda_i}) < r_{\lambda_i}$, that is, if the number of parameters in the solution space of the above linear system is not equal to the algebraic multiplicity, then $\mathbf{A}$ is not diagonalizable.

(iii) Otherwise, find a basis $S_{\lambda_i}$ of the eigenspace $E_{\lambda_i}$ for each eigenvalue $\lambda_i$, $i = 1, ..., k$. Necessarily $|S_{\lambda_i}| = r_{\lambda_i}$ for all $i = 1, ..., k$. Let $S = \bigcup_{i=1}^{k} S_{\lambda_i}$. Then

$$|S| = \sum_{i=1}^{k} |S_{\lambda_i}| = \sum_{i=1}^{k} r_{\lambda_i} = n,$$

and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is a basis for $\mathbb{R}^n$.

(iv) Let

$$\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}, \text{ and } \mathbf{D} = \operatorname{diag}(\mu_1, \mu_2, ..., \mu_n) = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n \end{pmatrix},$$

where $\mu_i$ is the eigenvalue associated to $\mathbf{u}_i$, $i = 1, ..., n$, $\mathbf{A}\mathbf{u}_i = \mu_i \mathbf{u}_i$. Then

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}.$$

**Example.** 1. $\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. It has eigenvalues 0 and 2 with multiplicity $r_0 = 1$ and $r_2 = 2$, respectively. Also, $\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$ is a basis for $E_0$ and $\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ is a basis for $E_2$. Then $\dim(E_0) = 1 = r_0$ and $\dim(E_2) = 2 = r_2$. Hence, $\mathbf{A}$ is diagonalizable, with

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}.$$

2. $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}$. $\mathbf{A}$ is a triangular matrix, hence the diagonal entries, $1, 2, 3$ are the eigenvalues, each with algebraic multiplicity 1. Therefore $\mathbf{A}$ is diagonalizable. We will need to find a basis for each of the eigenspace.

$\lambda = 1$: $\begin{pmatrix} 1-1 & -1 & -1 \\ 0 & 1-2 & -2 \\ 0 & 0 & 1-3 \end{pmatrix} = \begin{pmatrix} 0 & -1 & -1 \\ 0 & -1 & -2 \\ 0 & 0 & -2 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. So $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$ is a basis for $E_1$.

$\lambda = 2$: $\begin{pmatrix} 2-1 & -1 & -1 \\ 0 & 2-2 & -2 \\ 0 & 0 & 2-3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & -1 \\ 0 & 0 & -2 \\ 0 & 0 & -1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. So $\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\}$ is a basis for $E_2$.

$\lambda = 3$: $\begin{pmatrix} 3-1 & -1 & -1 \\ 0 & 3-2 & -2 \\ 0 & 0 & 3-3 \end{pmatrix} = \begin{pmatrix} 2 & -1 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 2 & 0 & -3 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$. So $\left\{ \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix} \right\}$ is a basis for $E_3$.

$$\Rightarrow \mathbf{A} = \begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 2 \end{pmatrix}^{-1}.$$

3. $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$. $\lambda = 1$ is the only eigenvalue with algebraic multiplicity $r_1 = 2$.

$\lambda = 1$: $\begin{pmatrix} 0 & -2 \\ 0 & 0 \end{pmatrix}$. There is only one non-pivot columns, hence $\dim(E_1) = 1 < 2 = r_1$.

This shows that $\mathbf{A}$ is not diagonalizable.

## 6.3 Orthogonal Diagonalization

A square matrix $\mathbf{A}$ of order $n$ is an orthogonal matrix if $\mathbf{A}^T = \mathbf{A}^{-1}$, equivalently, $\mathbf{A}^T\mathbf{A} = \mathbf{I} = \mathbf{A}\mathbf{A}^T$.

**Theorem.** *Let $\mathbf{A}$ be a square matrix of order $n$. The following statements are equivalent.*

*(i) $\mathbf{A}$ is orthogonal.*

*(ii) The columns of $\mathbf{A}$ forms an orthonormal basis for $\mathbb{R}^n$.*

*(iii) The rows of $\mathbf{A}$ forms an orthonormal basis for $\mathbb{R}^n$.*

*Proof.* Write

$$\mathbf{A} = \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{pmatrix} = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_n \end{pmatrix},$$

where for $i = 1, ..., n$, $\mathbf{c}_i$ and $\mathbf{r}_i$ are the columns of and rows of $\mathbf{A}$, respectively. Then

$$\begin{aligned} \mathbf{A}^T\mathbf{A} &= \begin{pmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \vdots \\ \mathbf{c}_n^T \end{pmatrix} \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1^T\mathbf{c}_1 & \mathbf{c}_1^T\mathbf{c}_2 & \cdots & \mathbf{c}_1^T\mathbf{c}_n \\ \mathbf{c}_2^T\mathbf{c}_1 & \mathbf{c}_2^T\mathbf{c}_2 & \cdots & \mathbf{c}_2^T\mathbf{c}_n \\ \vdots & & \ddots & \vdots \\ \mathbf{c}_n^T\mathbf{c}_1 & \mathbf{c}_n^T\mathbf{c}_2 & \cdots & \mathbf{c}_n^T\mathbf{c}_n \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{c}_1 \cdot \mathbf{c}_1 & \mathbf{c}_1 \cdot \mathbf{c}_2 & \cdots & \mathbf{c}_1 \cdot \mathbf{c}_n \\ \mathbf{c}_2 \cdot \mathbf{c}_1 & \mathbf{c}_2 \cdot \mathbf{c}_2 & \cdots & \mathbf{c}_2 \cdot \mathbf{c}_n \\ \vdots & & \ddots & \vdots \\ \mathbf{c}_n \cdot \mathbf{c}_1 & \mathbf{c}_n \cdot \mathbf{c}_2 & \cdots & \mathbf{c}_n \cdot \mathbf{c}_n \end{pmatrix}, \end{aligned}$$

and

$$\begin{aligned} \mathbf{A}\mathbf{A}^T &= \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_n \end{pmatrix} \begin{pmatrix} \mathbf{r}_1^T & \mathbf{r}_2^T & \cdots & \mathbf{r}_n^T \end{pmatrix} = \begin{pmatrix} \mathbf{r}_1\mathbf{r}_1^T & \mathbf{r}_1\mathbf{r}_2^T & \cdots & \mathbf{r}_1\mathbf{r}_n^T \\ \mathbf{r}_2\mathbf{r}_1^T & \mathbf{r}_2\mathbf{r}_2^T & \cdots & \mathbf{r}_2\mathbf{r}_n^T \\ \vdots & & \ddots & \vdots \\ \mathbf{r}_n\mathbf{r}_1^T & \mathbf{r}_n\mathbf{r}_2^T & \cdots & \mathbf{r}_n\mathbf{r}_n^T \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{r}_1 \cdot \mathbf{r}_1 & \mathbf{r}_1 \cdot \mathbf{r}_2 & \cdots & \mathbf{r}_1 \cdot \mathbf{r}_n \\ \mathbf{r}_2 \cdot \mathbf{r}_1 & \mathbf{r}_2 \cdot \mathbf{r}_2 & \cdots & \mathbf{r}_2 \cdot \mathbf{r}_n \\ \vdots & & \ddots & \vdots \\ \mathbf{r}_n \cdot \mathbf{r}_1 & \mathbf{r}_n \cdot \mathbf{r}_2 & \cdots & \mathbf{r}_n \cdot \mathbf{r}_n \end{pmatrix}. \end{aligned}$$

So $\mathbf{A}^T\mathbf{A} = \mathbf{I}$ if and only if $\mathbf{c}_i \cdot \mathbf{c}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$, and $\mathbf{A}\mathbf{A}^T = \mathbf{I}$ if and only if $\mathbf{r}_i \cdot \mathbf{r}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$ □

An order $n$ square matrix $\mathbf{A}$ is orthogonally diagonalizable if

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T$$

for some orthogonal matrix $\mathbf{P}$ and diagonal matrix $\mathbf{D}$.

Supppose $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T$ for some orthogonal matrix $\mathbf{P}$ and diagonal matrix $\mathbf{D}$. Then

$$\mathbf{A}^T = (\mathbf{P}\mathbf{D}\mathbf{P}^T)^T = (\mathbf{P}^T)^T\mathbf{D}^T\mathbf{P}^T = \mathbf{P}\mathbf{D}\mathbf{P}^T = \mathbf{A},$$

since $\mathbf{D}$ is diagonal, and hence symmetric. This shows that if $\mathbf{A}$ is orthogonally diagonalizable, it is symmetric. The converse is also true, but the proof is beyond the scope of this course.

**Theorem.** *An order $n$ square matrix is orthogonally diagonalizable if and only if it is symmetric.*

The algorithm to orthogonally diagonalize a matrix is the same as the usual diagonalization, except until the last step, instead of using a basis of eigenvectors to form the matrix $\mathbf{P}$, we have to turn it into an orthonormal basis of eigenvectors, that is, to use the Gram-Schmidt process. However, we do not need to use the Gram-Schmidt process for the whole basis, but only among those eigenvectors that belong to the same eigenspace. This follows from the fact that the eigenspaces are already orthogonal to each other.

**Theorem.** *If $\mathbf{A}$ is orthogonally diagonalizable, then the eigenspaces are orthogonal to each other. That is, suppose $\lambda_1$ and $\lambda_2$ are distinct eigenvalues of a symeetric matrix $\mathbf{A}$, $\lambda_1 \neq \lambda_2$. Let $E_{\lambda_i}$ denote the eigenspace associated to eigenvalue $\lambda_i$, for $i = 1, 2$. Then for any $\mathbf{v}_1 \in E_{\lambda_1}$ and $\mathbf{v}_2 \in E_{\lambda_2}$, $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.*

### Algorithm to orthogonal diagonalization

Follow step (i) to (iii) in algorithm to diagonalization.

(iv) Apply Gram-Schmidt process to each basis $S_{\lambda_i}$ of the eigenspace $E_{\lambda_i}$ to obtain an orthonormal basis $T_{\lambda_i}$. Let $T = \bigcup_{i=1}^{k} T_{\lambda_i}$, it is an orthonormal set. Similarly, we have $|T_{\lambda_i}| = r_{\lambda_i}$, and so

$$|T| = \sum_{i=1}^{k} |T_{\lambda_i}| = \sum_{i=1}^{k} r_{\lambda_i} = n,$$

which shows that $T = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is an orthonormal basis for $\mathbb{R}^n$.

(v) Let

$$P = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}, \text{ and } \mathbf{D} = \operatorname{diag}(\mu_1, \mu_2, ..., \mu_n) = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n \end{pmatrix},$$

where $\mu_i$ is the eigenvalue associated to $\mathbf{u}_i$, $i = 1, ..., n$, $\mathbf{A}\mathbf{u}_i = \mu_i \mathbf{u}_i$. Then $\mathbf{P}$ is an orthogonal matrix, and

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T.$$

**Example.** Let $\mathbf{A} = \begin{pmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{pmatrix}$. $\mathbf{A}$ is symmetric, hence it is orthogonally diagonalizable.

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-5 & 1 & 1 \\ 1 & x-5 & 1 \\ 1 & 1 & x-5 \end{vmatrix} = (x-3)(x-6)^2.$$

$\mathbf{A}$ has eigenvalues $\lambda = 3, 6$, with miltiplicity $r_3 = 1$, $r_6 = 2$. Let us now compute the eigenspaces.

$\lambda = 3$: $\begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$. So $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$ is a basis for $E_3$.

$\lambda = 6$: $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. So $\left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$ is a basis for $E_6$.

Observe that $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is orthogonal to $\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$, but the set $\left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$ is not orthogonal. So we only need to apply the Gram-Schmidt process to the set $\left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$.

$$\begin{aligned} \mathbf{v}_1 &= \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \\ \mathbf{v}_2 &= \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = -\frac{1}{2} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}. \end{aligned}$$

Indeed, $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \right\}$ is an orthogonal basis. Normalizing, we get an orthonormal basis

$$\left\{ \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \right\}.$$

So

$$\mathbf{A} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}^T.$$

## 6.4 Application of Diagonalization: Recursion Formula

**Lemma.** *Suppose* $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. *Then* $\mathbf{A}^m = \mathbf{P}\mathbf{D}^m\mathbf{P}^{-1}$.

*Proof.* Exercise $\square$

**Lemma.** *If* $\mathbf{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$, *then* $\mathbf{D}^m = \begin{pmatrix} d_1^m & 0 & \cdots & 0 \\ 0 & d_2^m & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n^m \end{pmatrix}$.

*Proof.* Exercise $\square$

Combining the two lemmas, we have the following statement. Suppose $\mathbf{A}$ is diagonalizable, $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ with $\mathbf{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$. Then

$$\mathbf{A}^m = \mathbf{P} \begin{pmatrix} d_1^m & 0 & \cdots & 0 \\ 0 & d_2^m & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n^m \end{pmatrix} \mathbf{P}^1.$$

### Linear Recurrence

Consider the Fibonacci sequence

$$a_0 = 0, a_1 = 1, a_2 = 1, a_3 = 2, a_4 = 3, a_5 = 5, a_6 = 8, ...$$

where $a_{n+1} = a_n + a_{n-1}$. We can represent the linear recurrence relation as a matrix equation,

$$\begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a_{n-1} \\ a_n \end{pmatrix}.$$

Let $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. Observe that

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \mathbf{A} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}, \begin{pmatrix} a_2 \\ a_3 \end{pmatrix} = \mathbf{A} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \mathbf{A}^2 \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}, ..., \begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} = \mathbf{A}^n \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \mathbf{A}^n \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Hence, by computing $\mathbf{A}^n$, we are able to find the general formula $a_n$ for the Fibonacci sequence. We will diagonalize $\mathbf{A}$.

$$\begin{vmatrix} x & -1 \\ -1 & x-1 \end{vmatrix} = x^2 - x - 1.$$

The eigenvalues are $\frac{-1 \pm \sqrt{5}}{2}$. Let $\lambda_\alpha = \frac{-1+\sqrt{5}}{2}$ and $\lambda_\beta = \frac{-1-\sqrt{5}}{2}$. We will now compute the eigenvectors. For each $\lambda = \lambda_\alpha, \lambda_\beta$,

$$\lambda\mathbf{I} - \mathbf{A} = \begin{pmatrix} \lambda & -1 \\ -1 & \lambda - 1 \end{pmatrix}.$$

So the eigenvector is $\mathbf{v} = \begin{pmatrix} 1 \\ \lambda \end{pmatrix}$, for $\lambda = \lambda_\alpha, \lambda_\beta$. Hence,

$$\mathbf{A} = \mathbf{PDP}^{-1} = \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix} \begin{pmatrix} \lambda_\alpha & 0 \\ 0 & \lambda_\beta \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix}^{-1}.$$

Then

$$\mathbf{A}^n = \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix} \begin{pmatrix} \lambda_\alpha^n & 0 \\ 0 & \lambda_\beta^n \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix}^{-1}.$$

So,

$$\begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix} \begin{pmatrix} \lambda_\alpha^n & 0 \\ 0 & \lambda_\beta^n \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

We can avoid computing $\begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix}^{-1}$ but compute $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ directly, $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, that is, we only need to solve the linear system.

$$\left( \begin{array}{cc|c} 1 & 1 & 0 \\ \lambda_\alpha & \lambda_\beta & 1 \end{array} \right) \xrightarrow{R_2 - \lambda_\alpha R_1} \left( \begin{array}{cc|c} 1 & 1 & 0 \\ 0 & \lambda_\beta - \lambda_\alpha & 1 \end{array} \right) \xrightarrow{\frac{1}{\lambda_\beta - \lambda_\alpha} R_2} \left( \begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & \frac{1}{\lambda_\beta - \lambda_\alpha} \end{array} \right)$$
$$\xrightarrow{R_1 - R_2} \left( \begin{array}{cc|c} 1 & 0 & \frac{1}{\lambda_\alpha - \lambda_\beta} \\ 0 & 1 & \frac{1}{\lambda_\beta - \lambda_\alpha} \end{array} \right)$$

And so $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/(\lambda_\alpha - \lambda_\beta) \\ 1/(\lambda_\beta - \lambda_\alpha) \end{pmatrix} = \frac{1}{\lambda_\alpha - \lambda_\beta} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, which gives us

$$\begin{aligned} \begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} &= \frac{1}{\lambda_\alpha - \lambda_\beta} \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix} \begin{pmatrix} \lambda_\alpha^n & 0 \\ 0 & \lambda_\beta^n \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\ &= \frac{1}{\lambda_\alpha - \lambda_\beta} \begin{pmatrix} 1 & 1 \\ \lambda_\alpha & \lambda_\beta \end{pmatrix} \begin{pmatrix} \lambda_\alpha^n \\ -\lambda_\beta^n \end{pmatrix} = \frac{1}{\lambda_\alpha - \lambda_\beta} \begin{pmatrix} \lambda_\alpha^n - \lambda_\beta^n \\ \lambda_\alpha^{n+1} - \lambda_\beta^{n+1} \end{pmatrix}. \end{aligned}$$

Reading off the first coordinates, we have

$$a_n = \frac{\lambda_\alpha^n - \lambda_\beta^n}{\lambda_\alpha - \lambda_\beta} = \frac{(-1 + \sqrt{5})^n - (-1 - \sqrt{5})^n}{2^n \sqrt{5}}.$$

### Markov Chain

Sheldon only patronizes three stalls in the school canteen, the mixed rice, noodle, and mala hotpot stall for lunch everyday. He never buys from same stall two days in a row. If he buys from the mixed rice stall on a certain day, there is a 40% chance he will eat from the noodles stall the next day. If he buys from the noodle stall on a certain day, there is a 50% chance he will eat mala hotpot the next day. If he buys eats mala hotpot on a certain day, there is a 60% chance he will buy mixed rice the next day.

(a.) Suppose Sheldon had noodles today. What is the probability that he patronizes each of the stalls 3 days later?

(b.) Show that regardless of what he had today, in the long run, he is most likely to patronize the mixed rice and mala hotpot stall equally.

Let $a_n$, $b_n$, and $c_n$ be the probability that Sheldon eats from the mixed rice, noodles, and mala hotpot stalls $n$ days later, respectively. Then the probabilities $a_n, b_n, c_n$ depends on the probabilities for the previous day as such

$$\begin{array}{rcrcrcr} a_n & = & & & 0.5b_{n-1} & + & 0.6c_{n-1} \\ b_n & = & 0.4a_{n-1} & & & + & 0.4c_{n-1} \\ c_n & = & 0.6a_{n-1} & + & 0.5b_{n-1} & & \end{array},$$

or,

$$\begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix} = \mathbf{A} \begin{pmatrix} a_{n-1} \\ b_{n-1} \\ c_{n-1} \end{pmatrix},$$

where $\mathbf{A} = \begin{pmatrix} 0 & 0.5 & 0.6 \\ 0.4 & 0 & 0.4 \\ 0.6 & 0.5 & 0 \end{pmatrix}$.

We shall now diagonalize $\mathbf{A}$.

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x & -0.5 & -0.6 \\ -0.4 & x & -0.4 \\ -0.6 & -0.5 & x \end{vmatrix} = (x-1)(x+0.6)(x+0.4).$$

The eigenvalues of $\mathbf{A}$ are $\lambda = 1, -0.6, -0.4$, with algebraic multiplicity 1 each.

$$\lambda = 1\text{:}\quad \begin{pmatrix} 1 & -0.5 & -0.6 \\ -0.4 & 1 & -0.4 \\ -0.6 & -0.5 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -0.8 \\ 0 & 0 & 0 \end{pmatrix}. \text{ Let } \mathbf{v}_1 = \begin{pmatrix} 1 \\ 0.8 \\ 1 \end{pmatrix}.$$

$$\lambda = -0.6\text{:}\quad \begin{pmatrix} -0.6 & -0.5 & -0.6 \\ -0.4 & -0.6 & -0.4 \\ -0.6 & -0.5 & -0.6 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \text{ Let } \mathbf{v}_{-0.6} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

$$\lambda = -0.4\text{:}\quad \begin{pmatrix} -0.4 & -0.5 & -0.6 \\ -0.4 & -0.4 & -0.4 \\ -0.6 & -0.5 & -0.4 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}. \text{ Let } \mathbf{v}_{-0.4} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}.$$

So $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$, where $\mathbf{P} = \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix}$ and $\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -0.6 & 0 \\ 0 & 0 & -0.4 \end{pmatrix}$, and

$$\begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix} = \mathbf{A}^n \begin{pmatrix} a_0 \\ b_0 \\ c_0 \end{pmatrix}.$$

(a.) Since he had noodles today, $\begin{pmatrix} a_0 \\ b_0 \\ c_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. So, the probabilities 3 days later will be

$$\mathbf{A}^3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & (-0.6)^3 & 0 \\ 0 & 0 & (-0.4)^3 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.38 \\ 0.24 \\ 0.38 \end{pmatrix}.$$

(b.) In the long run, $\mathbf{A}^k = \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & (-0.6)^k & 0 \\ 0 & 0 & (-0.4)^k \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix}^{-1}$

tends to

$$\begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix}^{-1} = \frac{1}{14} \begin{pmatrix} 5 & 5 & 5 \\ 4 & 4 & 4 \\ 5 & 5 & 5 \end{pmatrix}.$$

Notice that the columns are the same, so the probabilities in the long run is the same regardless of what he had today. Moreover, he is most likely, with $5/14$ chance of patronizing the mixed rice and mala hotpot stall.

## Appendix to Lecture 11

**Theorem.** *Let $\mathbf{A}$ be a square matrix of order $n$. $\mathbf{A}$ is diagonalizable if and only if there exists a basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ of $\mathbb{R}^n$ of eigenvectors of $\mathbf{A}$.*

*Proof.*

($\Rightarrow$) Suppose $\mathbf{A}$ is diagonalizable, $\mathbf{A} = \mathbf{PDP}^{-1}$ for some diagonal matrix $\mathbf{D}$ and invertible matrix $\mathbf{P}$. Write $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$ and $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$, where $\mathbf{u}_i$ is the $i$-th column of $\mathbf{P}$, for $i = 1, ..., n$. Since $\mathbf{P}$ is invertible, $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ forms a basis for $\mathbb{R}^n$. Then we have

$$\begin{aligned}
&\mathbf{A} = \mathbf{PDP}^{-1} \Leftrightarrow \mathbf{AP} = \mathbf{PD} \\
&\Leftrightarrow \mathbf{A}\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \qquad (7)\\
&\Leftrightarrow \begin{pmatrix} \mathbf{Au}_1 & \mathbf{Au}_2 & \cdots & \mathbf{Au}_n \end{pmatrix} = \begin{pmatrix} \lambda_1\mathbf{u}_1 & \lambda_2\mathbf{u}_2 & \cdots & \lambda_n\mathbf{u}_n \end{pmatrix} \\
&\Leftrightarrow \mathbf{Au}_i = \lambda_i\mathbf{u}_i \text{ for all } i = 1, ..., n.
\end{aligned}$$

The third equivalence follows from that fact that $\text{diag}(\lambda_1, \lambda_2, ..., \lambda_n) = \text{diag}(\lambda_1, 1, ..., 1)$ $\text{diag}(1, \lambda_2, ..., 1) \cdots \text{diag}(1, 1, ..., \lambda_n)$, and post-multiplying a matrix by $\text{diag}(1, ..., \lambda_i, ..., 1)$ is multiplying the $i$-th column of the matrix by $\lambda_i$. Since $\mathbf{P}$ is invertible, $\mathbf{u}_i \neq \mathbf{0}$ for all $i = 1, ..., n$. This shows that $\mathbf{u}_i$ are eigenvectors of $\mathbf{A}$ associated to eigenvalues $\lambda_i$. So $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is a basis of $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$.

($\Leftarrow$) Suppose there is a basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ of $\mathbb{R}^n$ of eigenvectors of $\mathbf{A}$. Let $\lambda_i$ be the eigevalues of $\mathbf{A}$ associated to $\mathbf{u}_i$ for $i = 1, ..., n$, that is, $\mathbf{Au}_i = \lambda\mathbf{u}_i$ for $i = 1, ..., n$. Let $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$ and $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$. Then by (7), $\mathbf{A}$ is diagonalizable. $\square$

Recall we have shown that if $\mathbf{A}$ is diagonalizable, then the algebraic multiplicity of an eigenvalue $\lambda$ of $\mathbf{A}$ is the number of diagonal entries taking the value $\lambda$. Suppose the $i_1, i_2, ..., i_k$ diagonal entries of $\mathbf{D}$ take the value $\lambda$, that is, $\lambda_{i_1} = \lambda_{i_2} = \cdots = \lambda_{i_k} = \lambda$. This would mean that the algebraic multiplicity $r_\lambda$ of $\lambda$ is $k$, $r_\lambda = k$. Then the corresponding columns of $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$ must be eigenvectors associated to $\lambda$, that is, $\mathbf{u}_{i_1}, \mathbf{u}_{i_2}, ..., \mathbf{u}_{i_k}$ are all eigenvectors associated to $\lambda$. Since the columns of $\mathbf{P}$ are linearly independent, this means that $\{\mathbf{u}_{i_1}, \mathbf{u}_{i_2}, ..., \mathbf{u}_{i_k}\}$ is a linearly independent subset of the eigenspace $E_\lambda$ associated to $\lambda$. Hence, the geometric multiplicity of $\lambda$ is less than or equals to its algebraic multiplicity, $\dim(E_\lambda) \leq r_\lambda$.

In general, for a square matrix $\mathbf{A}$, regardless of whether it is diagonalizable, the geometric multiplicity of an eigenvalue is less than or equals to its algebraic multiplicity.

**Theorem** (Geometric multiplicity is no greater than algebraic multiplicity)**.** *The geometric multiplicity of an eigenvalue $\lambda$ of a square matrix $\mathbf{A}$ is no greater than the algebraic multiplicity, that is,*

$$1 \leq \dim(E_\lambda) \leq r_\lambda.$$

*Proof.* Suppose $\mathbf{A}$ has order $n$. Let $\lambda$ be an eigenvalue of $\mathbf{A}$ and $E_\lambda$ be the associated eigenspace. Suppose $\dim(E_\lambda) = k$. Let $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for the eigenspace $E_\lambda$.

Extend this set to be a basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}_{k+1}, ..., \mathbf{u}_n\}$ of $\mathbb{R}^n$. Let $\mathbf{Q} = \begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{pmatrix}$, it is an invertbile matrix. Note that

$$\begin{aligned}\begin{pmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n\end{pmatrix} &= \mathbf{I} = \mathbf{Q}^{-1}\mathbf{Q} = \mathbf{Q}^{-1}\begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{pmatrix} \\ &= \begin{pmatrix}\mathbf{Q}^{-1}\mathbf{u}_1 & \mathbf{Q}^{-1}\mathbf{u}_2 & \cdots & \mathbf{Q}^{-1}\mathbf{u}_n\end{pmatrix},\end{aligned}$$

that is, $\mathbf{Q}^{-1}\mathbf{u}_i = \mathbf{e}_i$ for all $i = 1, ..., n$. Let $\mathbf{B} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$. Then

$$\begin{aligned}\mathbf{B} &= \mathbf{Q}^{-1}\mathbf{A}\begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{pmatrix} = \mathbf{Q}^{-1}\begin{pmatrix}\mathbf{A}\mathbf{u}_1 & \mathbf{A}\mathbf{u}_2 & \cdots & \mathbf{A}\mathbf{u}_n\end{pmatrix} \\ &= \mathbf{Q}^{-1}\begin{pmatrix}\lambda\mathbf{u}_1 & \cdots & \lambda\mathbf{u}_k & \mathbf{A}\mathbf{u}_{k+1}\cdots & \mathbf{A}\mathbf{u}_n\end{pmatrix} \\ &= \begin{pmatrix}\lambda\mathbf{Q}^{-1}\mathbf{u}_1 & \cdots & \lambda\mathbf{Q}^{-1}\mathbf{u}_k & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_{k+1}\cdots & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_n\end{pmatrix} \\ &= \begin{pmatrix}\lambda & 0 & \cdots & 0 & & & \\ 0 & \lambda & \cdots & 0 & & & \\ \vdots & & \ddots & \vdots & & & \\ 0 & 0 & \cdots & \lambda & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_{k+1}\cdots & & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_n \\ \vdots & \vdots & & \vdots & & & \\ 0 & 0 & \cdots & 0 & & & \end{pmatrix}.\end{aligned}$$

This means that $\det(x\mathbf{I} - \mathbf{B}) = (x - \lambda)^k p(x)$ for some polynomial $p(x)$. By (6),

$$\det(x\mathbf{I} - \mathbf{A}) = \det(x\mathbf{I} - \mathbf{B}) = (x - \lambda)^k p(x).$$

This means that the algebraic multiplicity of the eigenvalue $\lambda$ of $\mathbf{A}$ is no less than $k$, that is,

$$r_\lambda \geq k = \dim(E_\lambda).$$

□

**Lemma.** *Suppose $V_i \subseteq \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$, for $i = 1, 2$ with trivial intersection, $V_1 \cap V_2 = \{\mathbf{0}\}$. Let $S_i \subseteq V_i$ be a linearly independent subset for $i = 1, 2$. Then $S_1 \cup S_2$ is linearly independent.*

*Proof.* Let $S_1 = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ and $S_2 = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m\}$. Suppose $c_1, c_2, ..., c_k, d_1, d_2, ..., d_m \in \mathbb{R}$ are real numbers such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k + d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_m\mathbf{v}_m = \mathbf{0}.$$

Let us consider cases.

Case 1: $c_1 = c_2 = \cdots = c_k = 0$. Then the equation is reduced to

$$d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_m\mathbf{v}_m = \mathbf{0},$$

and hence by independence of $S_2$, necessarily $d_1 = d_2 = \cdots = d_m = 0$.

Case 2: $d_1 = d_2 = \cdots = d_m$. Then the equation is reduced to

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{0},$$

and hence by independence of $S_1$, necessarily $c_1 = c_2 = \cdots = c_k = 0$.

Case 3: There is an $i = 1, ..., k$ and $j = 1, ..., m$ such that $c_i \neq 0$ and $d_j \neq 0$. Let

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = -(d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_m\mathbf{v}_m).$$

By assumption, $\mathbf{v} \neq \mathbf{0}$, and $\mathbf{v} \in \text{span}(S_i) \subseteq V_i$ for $i = 1, 2$. But since $V_1 \cap V_2 = \{\mathbf{0}\}$, this means that $\mathbf{v} = \mathbf{0}$, a contradiction. Therefore case 3 is not possible.

Hence, $c_1 = c_2 = \cdots = c_k = d_1 = d_2 = \cdots = d_m = 0$ in every possible cases. This shows that $S_1 \cup S_2$ is linearly independent. $\square$

**Theorem** (Intersection of distinct eigenspaces is trivial)**.** *Let $\lambda_1$ and $\lambda_2$ be distinct eigenvalues, $\lambda_1 \neq \lambda_2$, of a square matrix $\mathbf{A}$. Let $E_{\lambda_i}$ be the associated eigenspace for $\lambda_i$, $i = 1, 2$. Then the intersection of the eigenspaces is trivial,*

$$E_{\lambda_1} \cap E_{\lambda_2} = \{\mathbf{0}\}.$$

*Proof.* Let $\mathbf{v} \in E_{\lambda_1} \cap E_{\lambda_2}$. Then

$$\lambda_1\mathbf{v} = \mathbf{A}\mathbf{v} = \lambda_2\mathbf{v},$$

where the first equality follows from $\mathbf{v} \in E_{\lambda_1}$ and the second equality follows from $\mathbf{v} \in E_{\lambda_2}$. This means that

$$(\lambda_1 - \lambda_2)\mathbf{v} = \mathbf{0}.$$

But since $\lambda_1 \neq \lambda_2$, necessarily $\mathbf{v} = \mathbf{0}$. $\square$

Together with the previous lemma, the theorem shows that if $\lambda_1$ and $\lambda_2$ are two distinct eigenvalues of $\mathbf{A}$, $\lambda_1 \neq \lambda_2$ and $E_{\lambda_i}$ is the eigenspace associated to $\lambda_i$, for $i = 1, 2$, then for any linearly independent subsets $S_1 \subseteq E_{\lambda_1}$ and $S_2 \subseteq E_{\lambda_2}$, $S_1 \cup S_2$ is linearly independent. In particular, if $S_i$ is a basis for $E_{\lambda_i}$, $i = 1, 2$, then $S_1 \cup S_2$ is a basis for $E_{\lambda_1} + E_{\lambda_2}$.

**Theorem.** *Suppose $\mathbf{A}$ is a square matrix such that its characteristic polynomial can be written as a product of linear factors,*

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}}(x - \lambda_2)^{r_{\lambda_2}} \cdots (x - \lambda_k)^{r_{\lambda_k}},$$

*where $r_{\lambda_i}$ is the algebraic multiplicity of $\lambda_i$, for $i = 1, ..., k$, and the eigenvalues are distinct, $\lambda_i \neq \lambda_j$ for all $i \neq j$. Then $\mathbf{A}$ is diagonalizable if and only if for each eigenvalue of $\mathbf{A}$, its geometric multiplicity is equal to its algebraic multiplicity,*

$$\dim(E_{\lambda_i}) = r_{\lambda_i}$$

*for all eigenvalues $\lambda_i$ of $\mathbf{A}$.*

*Proof.* Firstly, note that $r_{\lambda_1} + r_{\lambda_2} + \cdots + r_{\lambda_k} = n$. Let $m_{\lambda_i} = \dim(E_{\lambda_i})$ and $S_{\lambda_i}$ be a basis of $E_{\lambda_i}$, for $i = 1, ..., k$. Then since geometric multiplicity is no greater than the algebraic multiplicity, we have $|S_{\lambda_i}| = m_{\lambda_i} \leq r_{\lambda_i}$. Let $S = \bigcup_{i=1}^{k} S_{\lambda_i}$ be the union of all the basis for the eigenspaces. Then

$$|S| = \sum_{i=1}^{k} |S_{\lambda_i}| = \sum_{i=1}^{k} m_{\lambda_i} \leq \sum_{i=1}^{k} r_{\lambda_i} = n.$$

By the previous lemma, $S$ is linearly independent. So $S$ is a basis for $\mathbb{R}^n$ if and only if $|S| = n$, which is equivalent to $\dim(E_{\lambda_i}) = m_{\lambda_i} = r_{\lambda_i}$ for all $i = 1, ..., k$. $\square$