MA1522 Linear Algebra for Computing

Lecture 11: Diagonalization

Yang Yue

Department of Mathematics
National University of Singapore

3 April, 2025

Outline

Question one in Section 6.2

Suppose $\mathbf{A}$ is a square matrix of order n with distinct eigenvalues $\lambda_1, ..., \lambda_p$, and algebraic multiplicities $r_1, ..., r_p$, respectively. What can you conclude about the sum

$$r_1 + r_2 + \cdots + r_p?$$

Hint: Suppose $\mathbf{A}$ is a square matrix of order n such that the characteristic polynomial splits into linear factors

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_p)^{r_p}.$$

What can you conclude about the sum

$$r_1 + r_2 + \cdots + r_p?$$

Slide 13: Algebraic Multiplicity

Let λ be an eigenvalue of $\mathbf{A}$. The algebraic multiplicity of λ is the largest integer r_λ such that

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda)^{r_\lambda} p(x),$$

for some polynomial $p(x)$. Alternatively, r_λ is the positive integer such that in the above equation, λ is not a root of $p(x)$.

Suppose $\mathbf{A}$ is an order n square matrix such that $\det(x\mathbf{I} - \mathbf{A})$ can be factorize into linear factors completely. Then we can write

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_k)^{r_k}$$

where $r_1 + r_2 + \cdots + r_k = n$, and $\lambda_1, \lambda_2, ..., \lambda_k$ are the distinct eigenvalues of $\mathbf{A}$. Then the algebraic multiplicity of λ_i is r_i for $i = 1, ..., k$.

Answer to Question one in Section 6.2

Q: Suppose $\mathbf{A}$ is a square matrix of order n with distinct eigenvalues $\lambda_1, \ldots, \lambda_p$, and algebraic multiplicities $r_1, \ldots, r_p$, respectively. What can you conclude about the sum

$$r_1 + r_2 + \cdots + r_p?$$

Answer: We have seen that when the characteristic polynomial splits into linear factors

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_p)^{r_p},$$

we have that

$$n = \text{the degree of } \det(x\mathbf{I} - \mathbf{A}) = r_1 + r_2 + \cdots + r_p.$$

In general,

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_p)^{r_p} q_1(x) \ldots q_k(x),$$

where $q_i(x)$ are quadratic polynomials, we have that $r_1 + r_2 + \cdots + r_p \leq n$.

Question two in Section 6.2

Is it possible for the geometric multiplicity to be 0, $\dim(E_\lambda) = 0$?

Slide 17: Eigenspace

Recall that eigenvectors of $\mathbf{A}$ associated to eigenvalue λ are nontrivials solution to

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

Since the system is homogeneous, the set of all solutions is a subspace. We will call it the eigenspace of $\mathbf{A}$ associated to eigenvalue λ.

Definition

Let $\mathbf{A}$ be an order n square matrix. The *eigenspace* associated to an eigenvalue λ of $\mathbf{A}$ is

$$E_\lambda = \left\{ \ \mathbf{v} \in \mathbb{R}^n \ \middle| \ \mathbf{A}\mathbf{v} = \lambda\mathbf{v} \ \right\} = \mathrm{Null}(\lambda\mathbf{I} - \mathbf{A}).$$

The *geometric multiplicity* of an eigenvalue λ is the dimension of its eigenspace,

$$\dim(E_\lambda) = \mathrm{nullity}(\lambda\mathbf{I} - \mathbf{A}).$$

Answer to Question two in Section 6.2

Q: Is it possible for the geometric multiplicity to be 0, $\dim(E_\lambda) = 0$?

Answer: No. If λ is an eigenvalue of $\mathbf{A}$, there is some $\mathbf{v} \neq \mathbf{0}$, such that $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Hence $\mathbf{v} \in E_\lambda$, thus $\dim(E_\lambda) \neq 0$.

Slide 32: Bounds for Geometric Multiplicity

The geometric multiplicity is bounded above by the algebraic multiplicity. The proof can be found in the appendix.

Theorem (Geometric Multiplicity is no greater than Algebraic multiplicity)

The geometric multiplicity of an eigenvalue λ of a square matrix $\mathbf{A}$ *is no greater than the algebraic multiplicity, that is,*

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

See the lower bound.

Exercise in Section 6.2

Suppose $\mathbf{A}$ is an $n \times n$ matrix with $n > 1$. Show that if $\mathbf{A}$ has only 1 eigenvalue λ, then $\mathbf{A}$ is diagonalizable if and only if $\mathbf{A}$ is the scalar matrix, $\mathbf{A} = \lambda \mathbf{I}_n$.

Hence, all non-scalar matrices with only 1 eigenvalue are not diagonalizable.

Slide 21: Diagonalization

Definition

A square matrix $\mathbf{A}$ of order n is *diagonalizable* if there exists an invertible matrix $\mathbf{P}$ such that

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

is a diagonal matrix, OR

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

Slides 24 and 25: Diagonalization

Theorem (Diagonalizability)

An $n \times n$ square matrix $\mathbf{A}$ is diagonalizable if and only if $\mathbf{A}$ has n linearly independent eigenvectors.

That is, $\mathbf{A}$ is diagonalizable if and only if we can find

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

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

$\mathbf{P}$ is invertible if and only if $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is a basis for $\mathbb{R}^n$.

Note that μ_i may not be distinct.

Slide 34: Equivalent Statements for Diagonalizability

Theorem

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

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

(ii) *There exists a basis* $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ *of* $\mathbb{R}^n$ *of eigenvectors of* $\mathbf{A}$.

(iii) *The characteristic polynomial of* $\mathbf{A}$ *splits into 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_{λ_i} *is the algebraic multiplicity of* λ_i*, for* $i = 1, ..., k$*, and the eigenvalues are distinct,* $\lambda_i \neq \lambda_j$ *for all* $i \neq j$*, and the geometric multiplicity is equal to the algebraic multiplicity for each eigenvalue* λ_i*,*

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

Answer to Exercise in Section 6.2

Suppose $\mathbf{A}$ is an $n \times n$ matrix with $n > 1$. Show that if $\mathbf{A}$ has only 1 eigenvalue λ, then $\mathbf{A}$ is diagonalizable if and only if $\mathbf{A}$ is the scalar matrix, $\mathbf{A} = \lambda \mathbf{I}_n$.

Proof: $(\Rightarrow)$ Suppose $\mathbf{A}$ is diagonalizable, there are invertible matrix $\mathbf{P}$ and diagonal matrix $\mathbf{D}$ such that $\mathbf{A} = \mathbf{PDP}^{-1}$. Since $\mathbf{A}$ only has one eigenvalue λ, $\mathbf{D} = \lambda \mathbf{I}$. Thus $\mathbf{A} = \mathbf{P}\lambda\mathbf{IP}^{-1} = \lambda\mathbf{I}$, i.e., it is a scalar matrix.

$(\Rightarrow)$ If $\mathbf{A}$ is a scalar matrix, then itself is a diagonal matrix, we can take $\mathbf{P} = \mathbf{I}$.

Question Three in Section 6.2

Suppose $\mathbf{A}$ is diagonalizable. Which of the following statement(s) is/are true?

(i) If the diagonal matrix $\mathbf{D}$ is fixed, then the invertible matrix $\mathbf{P}$ is fixed.

(ii) If the invertible matrix $\mathbf{P}$ is fixed, then the diagonal matrix $\mathbf{D}$ is fixed.

Recall that the j-th column vector in $\mathbf{P}$ are the associated eigenvector of λ_j which is the (j, j)-entry of $\mathbf{D}$.

Thus, (ii) is true, because each eigenvector can only be associated with a unique eigenvalue.

Answer to Question Three in Section 6.2, part (i)

Suppose $\mathbf{A}$ is diagonalizable. Which of the following statement(s) is/are true?

(i) If the diagonal matrix $\mathbf{D}$ is fixed, then the invertible matrix $\mathbf{P}$ is fixed.

This is false. For example, if $\mathbf{A} = \mathbf{D} = \mathbf{I}$, then $\mathbf{P}$ can be any invertible matrix.

Challenge in Section 6.4

Theorem

Let $\mathbf{P}$ *be an* $n \times n$ *stochastic matrix and*

$$\mathbf{x}_0, \quad \mathbf{x}_1 = \mathbf{P}\mathbf{x}_0, \quad \dots \quad , \quad \mathbf{x}_k = \mathbf{P}\mathbf{x}_{k-1}$$

be a Markov chain for some probability vector $\mathbf{x}_0$*. If the Markov chain converges, it will converge to an equilibrium vector.*

Exercise. Hint:

(i) Show that 1 is always an eigenvalue of a stochastic matrix.

(ii) Show that if $\mathbf{v}$ is a probability vector and $\mathbf{P}$ a stochastic matrix, then $\mathbf{P}\mathbf{v}$ is also a probability vector.

(iii) Show that if the Markov chain do converge, then the state vectors will converge to an equilibrium vector.

Slide 62: Markov Chain

Definition

(i) A vector $\mathbf{v} = (v_i)_n$ with nonnegative coordinates that add up to 1, $\sum_{i=1}^{n} v_i = 1$, is called a *probability vector*.

(ii) A *stochastic* matrix is a square matrix whose columns are probability vectors.

(iii) A *Markov chain* is a sequence of probability vectors $\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_k, ...$, together with a stochastic matrix $\mathbf{P}$ such that

$$\mathbf{x}_1 = \mathbf{P}\mathbf{x}_0, \quad \mathbf{x}_2 = \mathbf{P}\mathbf{x}_1, \quad \dots \quad , \quad \mathbf{x}_k = \mathbf{P}\mathbf{x}_{k-1}, \dots$$

(iv) An *equilibrium vector* for a stochastic matrix $\mathbf{P}$ is a probability vector $\mathbf{v}$ with $\mathbf{P}\mathbf{v} = \mathbf{v}$.

Observe that $\mathbf{x}_k = \mathbf{P}^k \mathbf{x}_0$.

Answer to the Challenge in Section 6.4, part 1

(i) Show that 1 is always an eigenvalue of a stochastic matrix.

Proof. Let $\mathbf{P} = (p_{ij})$ be an $n \times n$ stochastic matrix. Consider its characteristic polynomial

$$\det(x\mathbf{I} - \mathbf{P}) = \begin{vmatrix} x - p_{11} & -p_{12} & \dots & -p_{1n} \\ -p_{21} & x - p_{22} & \dots & -p_{2n} \\ \vdots & \vdots & & \vdots \\ -p_{n1} & -p_{n2} & \dots & x - p_{nn} \end{vmatrix}.$$

Performing row operations $R_1 + R_2$, $R_1 + R_3$, ... $R_1 + R_n$, we get

$$\det(x\mathbf{I} - \mathbf{P}) = \begin{vmatrix} x - \sum_{i=1}^n p_{i1} & x - \sum_{i=1}^n p_{i2} & \dots & x - \sum_{i=1}^n p_{in} \\ -p_{21} & x - p_{22} & \dots & -p_{2n} \\ \vdots & \vdots & & \vdots \\ -p_{n1} & -p_{n2} & \dots & x - p_{nn} \end{vmatrix}.$$

Since $\mathbf{P}$ is a stochastic matrix, $\sum_{i=1}^n p_{ij} = 1$ for each j. We can take out a common factor $(x - 1)$. Thus 1 is an eigenvalue of $\mathbf{P}$.

Answer to the Challenge in Section 6.4, part 2

(ii) Show that if **v** is a probability vector and **P** a stochastic matrix, then **Pv** is also a probability vector.

Proof. Let $\mathbf{P} = (p_{ij})$ and $\mathbf{v} = (v_1, \ldots, v_n)^T$. Then

$$\mathbf{Pv} = (\sum_{j=1}^{n} p_{1j}v_j, \sum_{j=1}^{n} p_{2j}v_j, \ldots, \sum_{j=1}^{n} p_{nj}v_j)^T.$$

Adding all the entries, we have

$$\sum_{i=1}^{n}(\sum_{j=1}^{n} p_{ij}v_j) = \sum_{j=1}^{n}\sum_{i=1}^{n} p_{ij}v_j = \sum_{j=1}^{n} v_j(\sum_{i=1}^{n} p_{ij}) = \sum_{j=1}^{n} v_j = 1,$$

where the third $=$ comes from **P** is stochastic, and the last comes from **v** is a probability vector.

Answer to the Challenge in Section 6.4, part 3

(iii) Show that if the Markov chain does converge, then the state vectors will converge to an equilibrium vector.

Proof. Let the Markov chain be $\mathbf{x}_k = \mathbf{P}^k \mathbf{x}_0$ for some stochastic matrix $\mathbf{P}$ and probability vector $\mathbf{x}_0$.

By (ii), each $\mathbf{x}_k$ is also a probability vector, i.e., $\sum_{i=1}^{n} x_{ki} = 1$ for $k = 1, 2, \ldots$. Let's assume $\lim_k \mathbf{x}_k = \mathbf{a} = (a_1, \ldots, a_n)^T$. Then

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} \lim_k x_{ki} = \lim_k \sum_{i=1}^{n} x_{ki} = 1.$$

Finally, let $k \to \infty$ in $x_{k+1} = \mathbf{P} x_k$, we get $\mathbf{a} = \mathbf{P}\mathbf{a}$. In other words, $\mathbf{a}$ is an equilibrium vector.

Slides 48 and 50: Orthogonally Diagonalizable

Definition

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 $\mathbf{D}$.

Theorem

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

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

(ii) *There exists an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ of $\mathbb{R}^n$ of eigenvectors of* $\mathbf{A}$.

(iii) $\mathbf{A}$ *is a symmetric matrix.*

Slide 56: Algorithm to Orthogonal Diagonalization

Let $\mathbf{A}$ be an order n symmetric matrix. Since $\mathbf{A}$ is symmetric, it is orthogonally diagonalizable.

1. Compute the characteristic polynomial
$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}}(x - \lambda_2)^{r_{\lambda_2}} \cdots (x - \lambda_k)^{r_{\lambda_k}}.$$
2. For each eigenvalue λ_i of $\mathbf{A}$, $i = 1, ..., k$, find a basis S_{λ_i} for the eigenspace, that is, find a basis S_{λ_i} for the solution space of the following linear system,
$$(\lambda_i\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}.$$
3. Apply Gram-Schmidt process to each basis S_{λ_i} of the eigenspace E_{λ_i} to obtain an orthonormal basis T_{λ_i}. Let $T = \bigcup_{i=1}^{k} T_{\lambda_i}$. Then $T = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is an orthonormal basis for $\mathbb{R}^n$.
4. Let $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$, and $\mathbf{D} = \mathrm{diag}(\mu_1, \mu_2, ..., \mu_n)$, where μ_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
$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T.$$

Question about Section 6.3

Which of the follow matrices are orthogonally diagonalizable?

(a) $\mathbf{AB}$, for orthogonally diagonalizable matrices $\mathbf{A}$ and $\mathbf{B}$ of the same size.

(b) $\mathbf{A} + \mathbf{B}$, for orthogonally diagonalizable matrices $\mathbf{A}$ and $\mathbf{B}$ of the same size.

(c) An orthogonal matrix $\mathbf{A}$.

(d) $\mathbf{A} + \mathbf{A}^T$, where $\mathbf{A}$ is any square matrix.

(e) $\mathbf{A}^T\mathbf{A}$ for any matrix $\mathbf{A}$.

Answer to Question about Section 6.3, part (a)–(c)

By Theorem on Slide 50, we only need to answer which matrices are symmetric.

(a) $\mathbf{AB}$, for orthogonally diagonalizable matrices $\mathbf{A}$ and $\mathbf{B}$ of the same size.
Answer: $\mathbf{AB}$ may not be symmetric, for example, try
$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$.

(b) $\mathbf{A} + \mathbf{B}$, for orthogonally diagonalizable matrices $\mathbf{A}$ and $\mathbf{B}$ of the same size.
Answer: Sum of symmetric matrices is symmetric.

(c) An orthogonal matrix $\mathbf{A}$.
Answer: Orthogonal matrices may not be symmetric. For example, $\mathbf{A} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$.

Answer to Question about Section 6.3, part (d) and (e)

(d) $\mathbf{A} + \mathbf{A}^T$, where $\mathbf{A}$ is any square matrix.
Answer: It is symmetric.
$(\mathbf{A} + \mathbf{A}^T)^T = \mathbf{A}^T + (\mathbf{A}^T)^T = \mathbf{A}^T + \mathbf{A} = \mathbf{A} + \mathbf{A}^T$.

(e) $\mathbf{A}^T\mathbf{A}$ for any matrix $\mathbf{A}$.
Answer: For any matrix $\mathbf{A}$, $(\mathbf{A}^T\mathbf{A})^T = \mathbf{A}^T(\mathbf{A}^T)^T = \mathbf{A}^T\mathbf{A}$.
This is related to the singular value decomposition of $\mathbf{A}$.

Example

Let $\mathbf{A} = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. Find its orthogonal diagonalization.

We follow the algorithm on Slide 56.
First, compute its characteristic polynomial:

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-3 & -1 & 0 \\ -1 & x-3 & 0 \\ 0 & 0 & x-2 \end{vmatrix} = (x-2)^2(x-4).$$

Thus $\mathbf{A}$ has two eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 4$.

Example (conti.)

For $\lambda_1 = 2$, the eigenspace

$$E_2 = \left\{ s \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} : s, t \in \mathbb{R} \right\}.$$

The basis is already orthogonal, we only need to normalize them and get

$$\mathbf{p}_1 = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)^T, \mathbf{p}_2 = (0, 0, 1)^T.$$

Example (conti.)

For $\lambda_2 = 4$, the eigenspace

$$E_4 = \{(t, t, 0)^T : t \in \mathbb{R}\}$$

We get $\mathbf{p}_3 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)^T$. Thus
Let

$$\mathbf{P} = \begin{pmatrix} -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{D} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

We get $\mathbf{A} = \mathbf{PDP}^{-1}$, where

$$\mathbf{P}^{-1} = \mathbf{P}^T = \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{pmatrix}.$$