Lecture 9. Eigenproblem, diagonalization of a matrix, and similarity transformation.

1. Importance of eigenvalues.
   In Ex. 3 and 4 of lecture 1 we showed that eigenvalues and eigenvectors arise naturally when we solve a matrix diff. eqn. $Mx'' + Kx = 0$ (Ex. 3) or study long-term dynamics of a system $x^{(n+1)} = Ax^n$ (Ex. 4). The situation that we saw in both cases can be illustrated by the solution of a first-order matrix IVP (initial value problem):
   \[
   \frac{dx}{dt} = Ax \tag{1a}
   \]
   \[
   x(0) = x^{(0)} \tag{1b}
   \]
   where $A$ is a constant $p \times p$ matrix.
   We look for the solution of (1a) in the form:
   \[
   x(t) = e^{At} \nu, \quad \nu = \text{const.} \tag{2}
   \]
   Substitute (2) into (1a) to get:
   \[
   \frac{d}{dt} (e^{At} \nu) = Ae^{At} \nu
   \]
   \[
   e^{At} \nu = e^{At} A \nu = \lambda^t \nu.
   \]
   Suppose $A$ has $p$ l.i. eigenvectors $\nu_i$ satisfying
   \[
   A \nu_i = \lambda_i \nu_i, \quad i = 1, \ldots, p. \tag{3}
   \]
   Then we get $p$ l.i. solutions of (1a):
   \[
   x(t) = \sum_{i=1}^{p} c_i e^{\lambda_i t} \nu_i, \quad c_i = \text{const.}
   \]
\[ x_i(t) = e^{\lambda_i t} v_i, \quad (4) \]

But all solutions of (1a) form a vector space (Ex. 4 in Lecture 3), \( \Rightarrow \) any lin. combination of solutions of (4) is also a solution of (1a):

\[ x(t) = c_1 e^{\lambda_1 t} v_1 + \ldots + c_p e^{\lambda_p t} v_p. \quad (5) \]

The constants \( c_1, \ldots, c_p \) are found from the initial condition. Since \( v_1, \ldots, v_p \) are p l.i. vectors in \( \mathbb{R}^p \), they form a basis. Then any initial condition \( x^{(0)} \) in (1b) can be expanded over this basis:

\[ x^{(0)} = c_1 v_1 + \ldots + c_p v_p, \quad (6a) \]

i.e. \( c_1, \ldots, c_p \) can be found from

\[ [v_1, \ldots, v_p] \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} = x^{(0)}, \quad \Rightarrow \]

\[ \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} = p^{-1} x^{(0)}. \quad (6b) \]

Thus, Eqs. (5), (3), (6b) form the solution of the IVP (4).

**To summarize:** If matrix \( A \) has \( p \) l.i. eigenvectors, we can solve the IVP (1).

Then we can ask questions similar to those we asked in Lecture 1.
New Q1: Are there any properties of A that guarantee that it has p l.i.e. eigenvectors?
New Q2: What is the solution of (1) if A has fewer than p l.i.e. eigenvectors?
New Q3: Are there any properties of A that guarantee that all \( \Re(\lambda_i) < 0 \) (since then all \( e^{\lambda_it} \to 0 \) as \( t \to \infty \), and the system is called stable)? Or that all \( \lambda_i \) are real?
New Q4: How can we solve a non-homogeneous version of (1) (with \( Ax \) being replaced by \( Ax + u(t) \))? 

We will be answering these questions. See also Sec. 7.1 for different aspects of why eigenvectors and eigenvalues are important.

(2) Basic properties of eigenvalues & eigenvectors.

Eigenvalues \( \lambda \) are defined as

\[
A\mathbf{v} = \lambda \mathbf{v}, \quad \mathbf{v} \neq \mathbf{0}
\]

\[
A\mathbf{v} - \lambda \mathbf{v} = \mathbf{0}
\]

\[
(A - \lambda I)\mathbf{v} = \mathbf{0}
\]

\[\Rightarrow \text{by definition, } (A - \lambda I) = \text{singular} \]

\[\Rightarrow \text{(Thm. 6 of Lecture 2)} \]

\[
\det (A - \lambda I) = 0
\]

(7)

If \( A = p \times p \), then the l.h.s. of (7) is a \( p \)-th degree polynomial (characteristic polynomial of \( A \)).
\[
\det (A - \lambda I) = (-1)^p (\lambda^p + q_1 \lambda^{p-1} + \cdots + q_p \lambda + q_p)
\]
where \(\lambda_i \neq \lambda_j\) for \(i \neq j\), \(r \leq p\), and \(m_1 + \cdots + m_r = p\). The integers \(m_i \geq 1\) are called the algebraic multiplicity of \(\lambda_i\). An eigenvalue with \(m_i = 1\) is called a simple eigenvalue.

Note \(\text{Eq. (7)}\) is a useful theoretical tool to study properties of eigenvalues, but it should not be used in practice to find eigenvalues of a large matrix (say 10x10 or larger). There are 2 reasons:

1) As we saw in Lecture 2 and HW2, it is impractical (and often impossible) to compute the determinant of a large matrix.

2) Even if one finds the polynomial in (8), then it is still a bad idea to try to solve an eqn. polynomial \(p(\lambda) = 0\), because for a high-degree polynomial (i.e. when \(p\) is large), the computed roots may be extremely sensitive to tiny changes in its coefficients. Google "Wilkinson's polynomial" for the classic example of this bad behavior.

But we still deduce some theoretical results from \(\det (A - \lambda I) = 0\).
Thm. 1 \((\lambda = 0 \text{ is an eigenvalue of } A)\)
\[\Leftrightarrow (A \text{ is singular})\]
Follows from \(\det(A - 0I) = 0\) and Thm. 6 of Sec. 2.
Thm. 2(a) The eigenvalues of \(A\) and \(A^T\) are the same: \(\lambda_{A^T} = \lambda_A\).
(b) \(\lambda_{A^T} = \lambda_A^*\).

Proof for (a) (for (b) it is similar).
Given: \(\det(A - \lambda I) = 0\).
Want: \(\det(A^T - \lambda I) = 0\).
\[\det(A^T - \lambda I) = \det((A - \lambda I)^T) = \det((A - \lambda I)^T)\]
Thm. 5 of Lecture 9
\[\Rightarrow \det(A - \lambda I) = 0\], q.e.d.

Note: The eigenvectors of \(A^T\) are in general different from the eigenvectors of \(A\) with the same eigenvalues.

Ex. 1 \[A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}\]
\[\lambda_1 = 1, \quad \mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\]
\[\lambda_2 = 2, \quad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\]
\[A^T = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}\]
\[\lambda_1 = 1, \quad \mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\]
\[\lambda_2 = 2, \quad \mathbf{u}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\]

Despite \(\mathbf{v}_{A^T} \neq \mathbf{v}_A\), there are still important relations between \(\mathbf{v}_{A^T}\) and \(\mathbf{v}_A\) that we will exhibit later.

Thm. 3 If any of the coefficients \(a_1, \ldots, a_p\) in (9) is non-positive, then there is at least one \(\lambda_i\) with \(\Re(\lambda_i) > 0\) (and then the solution (5) of IVP (1) will \(\to \infty\) as \(t \to \infty\)).
The proof of this Thm. is elementary and is posted online. Note that this Thm.
does not give a sufficient condition for all
\( \text{Re}(\lambda_i) < 0 \) (when the solution \( \to 0 \) and is stable).
Both necessary and sufficient conditions
\( \text{to have } \text{Re}(\lambda_i) < 0 \) are given by
the Routh-Hurwitz criterion, which is also
posted online. It is practical only for
small \( p \). (You should be aware that in
Control Theory, the notations of the Routh-Hurwitz
criterion are different from those I posted online.)

Note: Thm. 3 \& \text{R-H criterion partially answer New Q3.}

Thm. 4 (Key Thm. 7.10(c,d)) \text{ see (a,b,e) on}
your own.

(i) If \( \lambda_1, \ldots, \lambda_r \) are distinct
\text{eigenvalues}
of matrix \( A \), then the corresponding
\text{eigenvectors} \( \mathbf{v}_1, \ldots, \mathbf{v}_r \) are l.i.

(ii) = corollary of (i).
\text{If } A = p \times p \text{ has } p \text{ distinct eigenvalues, then}
its \( p \) \text{ eigenvectors } \mathbf{v}_1, \ldots, \mathbf{v}_p \text{ are l.i. and}
hence form a basis in } \mathbb{R}^p.
Proof: \text{ p. 288 (see it - may help with Text 1).}

Note: Thm. 4(ii) partially answers New Q1.
3. **Repeated eigenvalues**

Let \( A \) have a repeated eigenvalue (it appears as a factor \((\lambda - \lambda_i)^{m_i}\) with \(m_i \geq 2\) in (9)). This eigenvalue may have \(m_i\) or fewer eigenvectors.

**Ex. 2** Let \( A = (\begin{smallmatrix} 0 & 0 \\ 1 & 1 \end{smallmatrix}) \) and \( B = (\begin{smallmatrix} 1 & a \\ 0 & 1 \end{smallmatrix}) \), \( a \neq 0 \).

They both have the double eigenvalue \( \lambda = 1 \).

How many l.i. eigenvectors are there for \( A \& B \)?

**Sol'n:** (a) for \( A = (\begin{smallmatrix} 0 & 0 \\ 1 & 1 \end{smallmatrix}) \),

\[
A \mathbf{v} = \lambda I \mathbf{v} \Rightarrow (\begin{smallmatrix} 0 & 0 \\ 1 & 1 \end{smallmatrix}) \mathbf{v} = (\begin{smallmatrix} 0 & 0 \\ 1 & 1 \end{smallmatrix}) \mathbf{v}
\]

\[
\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \mathbf{v} = 0 \Rightarrow (\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix})(\begin{smallmatrix} \alpha \\ \beta \end{smallmatrix}) = (\begin{smallmatrix} 0 \\ 0 \end{smallmatrix})
\]

\[]

no restrictions on \( \alpha \) and \( \beta \), \( \Rightarrow \)

\[
\mathbf{v} = (\begin{smallmatrix} \alpha \\ \beta \end{smallmatrix}) = \alpha (\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}) + \beta (\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}) \equiv \alpha \mathbf{v}_1 + \beta \mathbf{v}_2
\]

\( \Rightarrow \) there are two l.i. eigenvectors \((\begin{smallmatrix} 1 \\ 0 \end{smallmatrix})\) and \((\begin{smallmatrix} 0 \\ 1 \end{smallmatrix})\).

**Note:** # of l.i. eigenvectors = \( p - \text{rank}(A - \lambda I) \).

(b) for \( B = (\begin{smallmatrix} 1 & a \\ 0 & 1 \end{smallmatrix}) \), \( a \neq 0 \),

\[
B \mathbf{v} = \lambda I \mathbf{v} \Rightarrow (\begin{smallmatrix} 1 & a \\ 0 & 1 \end{smallmatrix}) \mathbf{v} = (\begin{smallmatrix} 0 & 0 \\ 1 & 1 \end{smallmatrix}) \mathbf{v}
\]

\[
(\begin{smallmatrix} 1 & a \\ 0 & 1 \end{smallmatrix})(\begin{smallmatrix} \alpha \\ \beta \end{smallmatrix}) = (\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}) \Rightarrow (\begin{smallmatrix} 0 & a \\ 0 & 0 \end{smallmatrix})(\begin{smallmatrix} \alpha \\ \beta \end{smallmatrix}) = (\begin{smallmatrix} 0 \\ 0 \end{smallmatrix})
\]

\( \Rightarrow \)

\[
\begin{cases} a \cdot \beta &= 0 \\ \alpha &= 0 \\ \beta &= 0 \end{cases} \Rightarrow \beta = 0 \text{ (since } a \neq 0) .
\]

Thus, there is only one eigenvector \( \mathbf{v} = (\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}) \).

**Note:** # of l.i. eigenvectors = \( p - \text{rank}(A - \lambda I) \).
Def. (Ex. 12 in book) The geometric multiplicity \( m \) of an eigenvalue \( \lambda \) is the max. \# of e.i. eigenvectors corresponding to this eigenvector.

So, in Ex. 2, \( m_{\lambda_1}(A) = 2 \), \( m_{\lambda_1}(B) = 1 \), while the algebraic multiplicities \( M_{\lambda_1}(A) = 2 \), \( M_{\lambda_1}(B) = 2 \).

(4) **Similarity transformation and its properties.**

Suppose \( A = p \times p \) has \( p \) e.i. eigenvectors

\[
A v_1 = \lambda_1 v_1
\]

(3) repeated.

Consider \( p = 3 \).

\[
A v_1 = \lambda_1 v_1 = v_1 \cdot \lambda_1 + v_2 \cdot 0 + v_3 \cdot 0 = [v_1, v_2, v_3] \begin{pmatrix} \lambda_1 \\ 0 \\ 0 \end{pmatrix} (\star)
\]

\[
A v_2 = \lambda_2 v_2 = v_1 \cdot 0 + v_2 \cdot \lambda_2 + v_3 \cdot 0 = [v_1, v_2, v_3] \begin{pmatrix} 0 \\ \lambda_2 \\ 0 \end{pmatrix} (**)
\]

\[
A v_3 = \lambda_3 v_3 = v_1 \cdot 0 + v_2 \cdot 0 + v_3 \cdot \lambda_3 = [v_1, v_2, v_3] \begin{pmatrix} 0 \\ 0 \\ \lambda_3 \end{pmatrix} (**)
\]

\[
A [v_1, v_2, v_3] = [v_1, v_2, v_3] \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}
\]

\[
A P = P \Lambda
\]

(10)

Since \( v_1, v_2, v_3 \) are e.i. \( \Rightarrow P \) is nonsingular

\( \Rightarrow P^{-1} \) exists \( \Rightarrow 

A = P \Lambda P^{-1}

(11a)

or \( P^{-1} A P = \Lambda \equiv \text{diag}(\lambda_1, \ldots, \lambda_p) \),

(11b)

where \( P \) is the matrix made up of the eigenvectors of \( A \). If \( A \) can be represented as in (11), it is called diagonalizable.
Thm. 5 (key Thm. (7.14) or (7.28))
\[ (A = p \times p \text{ is diagonalizable}) \iff \text{A has p.e.i. eigenvectors}. \]

Corollary of Thm. 5 + Thm. 4
\[ (A = p \times p \text{ has } p \text{ distinct eigenvalues}) \Rightarrow (A \text{ is diagonalizable}). \]

Note: The converse is not always true: a diagonalizable matrix may have repeated eigenvalues (matrix \( A = (1, 1) \) in Ex. 2).

Def: Let \( A \) and \( B \) be square matrices of the same size. Suppose there is \( P \) n.t.
\[ A = PBP^{-1} \]
(\( B \) is not necessarily diagonal). Then \( A \) and \( B \) are said to be similar, or related by the similarity transformation:
\[ A = PBP^{-1} \quad \text{or} \quad P^{-1}AP = B, \]

Thm. (7.22) - on your own.

Thm. 6 (Thm. (7.25)) For similar matrices \( A \) and \( B \):
\( \det A = \det B. \) In particular, \( A \) and \( B \)
are either both singular or both nonsingular.

For any \( k \),
\[ A^k = PB^kP^{-1} \quad \text{and} \quad B^k = P^{-1}A^kP. \]
In particular, if $B = \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$, then

$$A^k = P \cdot \text{diag}(\lambda_1^k, \ldots, \lambda_p^k) \cdot P^{-1} \quad (12)$$

So it's very easy to compute $A^k$ if we know its diagonalization $(11)$.

Moreover, if we know $(11)$, it is very easy to compute any function $f(x)$ that can be represented as

$$f(x) = \ldots \frac{a_{-n+1}}{x^n} + \frac{a_{n+1}}{x^{n-1}} + \ldots + a_0 + a_1 x + a_2 x^2 + \ldots$$

$$f(A) = P \cdot \text{diag}(f(\lambda_1), \ldots, f(\lambda_p)) \cdot P^{-1} \quad (13)$$

In particular, for the constant and diagonalizable $A$

$$e^{(At)} = P \cdot \text{diag}(e^{At_1}, \ldots, e^{At_p}) \cdot P^{-1} \quad (14)$$

where $e^{(At)}$ is defined by its Maclaurin series:

$$e^{At} = I + At + \frac{t^2}{2!} A^2 + \frac{t^3}{3!} A^3 + \ldots \quad (15)$$

(again, this is for a constant matrix $A$ only).

**Note 1:** Thm. 6(iii) says that it is very easy to find long-term evolution of $X(t) = A X(t)$ if we know the diagonalization (i.e., eigenvalues and eigenvectors) of $A$. Indeed, the solution is

$$X(t) = A^t X(0)$$
\[ x^{(k)} = P \Lambda^k P^{-1} x^{(0)} = \begin{bmatrix} v_1 & \ldots & v_p \end{bmatrix} \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p^k \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} \]

coordinates of \( x^{(0)} \) in basis \( \{v_1, \ldots, v_p\} \) (see (6))

HW2

# 1(a) \[ \mathbf{I} = [\lambda_1^k v_1, \ldots, \lambda_p^k v_p] \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} = c_1 \lambda_1^k v_1 + \cdots + c_p \lambda_p^k v_p. \]

This is the same as the answer in Ex. 4 of Sec. 1 (derived differently).

Note 2 Thm. 6(iii) says that it is very easy to solve the IVP

\[ \frac{dx}{dt} = Ax, \quad x(0) = x^{(0)} \] (1)

if we know the diagonalization of \( A \). Indeed, the solution of (1) is

\[ x(t) = e^{At} x^{(0)} \] (16)

where \( A \) is formally defined in (15). (You'll verify (16) in a few problems.) Substitute \( e^{At} \) from (14) into (16):

\[ \overbrace{e^{At}}^{P \Lambda^t P^{-1}} x^{(0)} = \begin{bmatrix} \lambda_1 t v_1 & \cdots & \lambda_p t v_p \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} \]

coordinates of \( x^{(0)} \) in basis \( \{v_1, \ldots, v_p\} \)

\[ = c_1 e^{\lambda_1 t} v_1 + \cdots + c_p e^{\lambda_p t} v_p, \]

which is the same as (5).

Note 3 Since it is so easy to work with diagonalizable matrices, they are sometimes called perfect (and non-diagonalizable ones are called defective).
Thm. 7 (Thm. (7.23)+(7.25))

Let \( A = SBS^{-1} \), where \( S \) is not necessarily \([v_1, \ldots, v_p]\).

Then:

(i) \( A \) and \( B \) have the same eigenvalues; \( \lambda_A = \lambda_B \).

(ii) \( \det A = \det B \).

(iii) If \( A v = \lambda v \), then \( B (S^{-1} v) = \lambda (S^{-1} v) \).

Note: The converse statement is not always true (see Ex. 2); i.e., (i) and (ii) do not always imply that \( A \) and \( B \) are similar.

Proof:

\[
\begin{align*}
A & = SBS^{-1} \quad v = \lambda v \\
B(S^{-1} v) & = \lambda (S^{-1} v)
\end{align*}
\]

eigenvector of \( B \)