STAT 309: MATHEMATICAL COMPUTATIONS I
FALL 2023
LECTURE 3

1. EIGENVALUES AND EIGENVECTORS

- recall: $A \in \mathbb{C}^{n\times n}$, if there exists $\lambda \in \mathbb{C}$ and $\mathbf{0} \neq \mathbf{x} \in \mathbb{C}^n$ such that
$$A\mathbf{x} = \lambda\mathbf{x},$$
we call λ an *eigenvalue* of A and $\mathbf{x}$ an *eigenvector* of A corresponding to λ or λ-eigenvector
- we review some basic properties and terminologies for eigenvalues and eigenvectors
- real matrices can have complex eigenvalues and eigenvectors, an example is
$$S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$
which has eigenvectors $[-i, 1]^\mathsf{T}$ and $[i, 1]^\mathsf{T}$ corresponding to eigenvalues i and $-i$ respectively
- eigenvector is a scale invariant notion, if $\mathbf{x}$ is a λ-eigenvector, then so is $c\mathbf{x}$ for any $c \in \mathbb{C}^\times$
- we usually, but not always, require that $\mathbf{x}$ be a unit vector, i.e., $\|\mathbf{x}\|_2 = 1$
- note that if $\mathbf{x}_1$ and $\mathbf{x}_2$ are λ-eigenvectors, then so is $\mathbf{x}_1 + \mathbf{x}_2$
- for an eigenvalue λ, the subspace
$$V_\lambda := \{\mathbf{x} \in \mathbb{C}^n : A\mathbf{x} = \lambda\mathbf{x}\}$$
is called the λ-*eigenspace* of A and is the set of all λ-eigenvectors of A together with $\mathbf{0}$
- the set of all eigenvalues of A is called its *spectrum* and often denoted $\lambda(A)$, i.e.,
$$\lambda(A) := \{\lambda \in \mathbb{C} : A\mathbf{x} = \lambda\mathbf{x} \text{ for some } \mathbf{x} \neq \mathbf{0}\}$$
- an $n \times n$ matrix always have n eigenvalues in $\mathbb{C}$ counted with multiplicty
- however an $n \times n$ matrix may not have n linear independent eigenvectors, an example is
$$J = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{1.1}$$
which has eigenvalue 0 with multiplicity 2 but only one eigenvector (up to scaling) $\mathbf{x} = [1, 0]^\mathsf{T}$
- normally we will sort the eigenvalues in descending order of magnitude
$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$$
- λ_1, also denoted $\lambda_{\max}$, is called the *principle eigenvalue* of A and a $\lambda_{\max}$-eigenvector is called a *principal eigenvector*
- the eigenvalues of A and A^T are identical, i.e., $\lambda(A) = \lambda(A^\mathsf{T})$
- the eigenvectors of A^T are called *left eigenvectors* of A (and if one needs to make a distinction, the usual eigenvectors are called *right eigenvectors*) and sometimes defined directly via
$$\mathbf{y}^\mathsf{T} A = \lambda\mathbf{y}^\mathsf{T}$$
- in general, for a nonsymmetric matrix $A \in \mathbb{R}^{n\times n}$, left and right eigenvectors corresponding to the same eigenvalue λ are different

Date: October 4, 2023, version 1.0. Comments, bug reports: `lekheng@uchicago.edu`.

- an $n \times n$ matrix A that has n linear independent eigenvectors $\mathbf{x}_1, \dots, \mathbf{x}_n$ is called a *diagonalizable matrix* since if we write these as columns of a matrix $X = [\mathbf{x}_1, \dots, \mathbf{x}_n]$, then X is necessarily nonsingular and

$$AX = [A\mathbf{x}_1, \dots, A\mathbf{x}_n] = [\lambda_1\mathbf{x}_1, \dots, \lambda_n\mathbf{x}_n] = X \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} =: X\Lambda \tag{2.1}$$

  and so

$$A = X\Lambda X^{-1} \tag{2.2}$$

  where $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ is a diagonal matrix of eigenvalues
- the decomposition (2.2) is called the *eigenvalue decomposition* (EVD) of A
- not every matrix has an EVD, an example is J in (1.1)
- summary: a matrix has an EVD iff it has n linearly independent eigenvectors iff it is diagonalizable
- since $\mathbf{x}_1, \dots, \mathbf{x}_n$ form a basis for the domain of A, we call this an *eigenbasis*
- note that the matrix of eigenvectors X in (2.2) is only required to be non-singular (a.k.a. invertible)
- in general it is difficult to check whether a matrix is diagonalizable
- however there is a special class of matrices for which we check diagonalizability easily, namely, the normal matrices
- a normal matrix is one that commutes with its adjoint, i.e. $A^*A = AA^*$
- recall that $A^* = \bar{A}^\mathsf{T}$ is the *adjoint* or Hermitian conjugate of A
- the matrix J above is *not* normal

Theorem 1 (Spectral Theorem for Normal Matrices). *Let $A \in \mathbb{C}^{n\times n}$. Then A is unitarily diagonalizable iff A has an orthonormal eigenbasis iff A is a normal matrix, i.e.*

$$A^*A = AA^*,$$

iff A has an EVD *of the form*

$$A = V\Lambda V^* \tag{2.3}$$

where $V \in \mathbb{C}^{n\times n}$ is unitary and $\Lambda \in \mathbb{C}^{n\times n}$ is diagonal.

- as in (2.1), $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ consists of the eigenvalues of A and the columns of $V = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ are the eigenvectors of A and are mutually orthonormal, i.e.

$$\mathbf{v}_i^*\mathbf{v}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- note that by the standard formula for matrix-matrix product, saying the column vectors of $V = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ are mutually orthonormal is the same as saying $V^*V = I = VV^*$ and is the same as saying that V is unitary
- note that by sum-of-rank-one formula for matrix-matrix product, an eigenvalue decomposition (2.3) may also be written as

$$A = \lambda_1\mathbf{v}_1\mathbf{v}_1^\mathsf{T} + \cdots + \lambda_n\mathbf{v}_n\mathbf{v}_n^\mathsf{T}$$

- a special class of normal matrices are the ones that are equal to their adjoint, i.e. $A^* = A$,and these are called *Hermitian* or self-adjoint matrices
- for Hermitian matrices, we can say more — the diagonal matrix Λ in (2.3) is real

Theorem 2 (Spectral Theorem for Hermitian Matrices). *Let $A \in \mathbb{C}^{n\times n}$. Then A is unitarily diagonalizable with a real diagonal matrix iff A has an orthonormal eigenbasis and all eigenvalues real iff A is a Hermitian matrix, i.e.*

$$A^* = A,$$

iff A has an EVD *of the form*

$$A = V\Lambda V^*$$

where $V \in \mathbb{C}^{n\times n}$ is unitary and $\Lambda \in \mathbb{R}^{n\times n}$ is diagonal.

- if we had start from a real matrix $A \in \mathbb{R}^{n\times n}$, then Theorem 2 holds true with 'Hermitian' replaced by *symmetric* (i.e., $A^{\mathsf{T}} = A$) and 'unitary' replaced by *orthogonal* (i.e., $V^{\mathsf{T}}V = I = VV^{\mathsf{T}}$)
- we have strict inclusions

$$\{\text{real symmetric}\} \subsetneq \{\text{Hermitian}\} \subsetneq \{\text{normal}\} \subsetneq \{\text{diagonalizable}\} \subsetneq \mathbb{C}^{n\times n}$$

3. Jordan canonical form

- if A is not diagonalizable and we want something like a diagonalization, then the best we could do is a *Jordan canonical form* or Jordan normal form where we get

$$A = XJX^{-1} \tag{3.1}$$

  - the matrix J has the following characteristics
    * it is not diagonal but it is the next best thing to diagonal, namely, *bidiagonal*, i.e. only the entries a_{ii} and $a_{i,i+1}$ can be non-zero, every other entry in J is 0
    * the diagonal entries of J are precisely the eigenvalues of A, counted with multiplicity
    * the superdiagonal entries $a_{i,i+1}$ are as simple as they can be — they can take one of two possible values $a_{i,i+1} = 0$ or 1
    * if $a_{i,i+1} = 0$ for all i, then J is in fact diagonal and (3.1) reduces to the eigenvalue decomposition
  - the matrix J is more commonly viewed as a block diagonal matrix

$$J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{bmatrix}$$

    * each block J_r, for $r = 1, \dots, k$, has the form

$$J_r = \begin{bmatrix} \lambda_r & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_r \end{bmatrix}$$

      where J_r is $n_r \times n_r$
    * clearly $\sum_{r=1}^{k} n_r = n$
  - the set of column vectors of X are called a *Jordan basis* of A
- in general the Jordan basis X include all eigenvectors of A but also additional vectors that are not eigenvectors of A
- the Jordan canonical form provides valuable information about the eigenvalues of A
- the values λ_j, for $j = 1, \dots, k$, are the eigenvalues of A
- for each distinct eigenvalue λ, the number of Jordan blocks having λ as a diagonal element is equal to the number of linearly independent eigenvectors associated with λ, this number is called the *geometric multiplicity* of the eigenvalue λ

- the sum of the sizes of all of these blocks is called the *algebraic multiplicity* of λ
- we now consider J_r's eigenvalues,

$$\lambda(J_r) = \lambda_r, \ldots, \lambda_r$$

where λ_r is repeated n_r times, but because

$$J_r - \lambda_r I = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}$$

is a matrix of rank $n_r - 1$, it follows that the homogeneous system $(J_r - \lambda_r I)\mathbf{x} = \mathbf{0}$ has only one vector (up to a scalar multiple) for a solution, and therefore there is only one eigenvector associated with this Jordan block
- the unique unit vector that solves $(J_r - \lambda_r I)\mathbf{x} = \mathbf{0}$ is the vector $\mathbf{e}_1 = [1, 0, \ldots, 0]^\mathsf{T}$
- now consider the matrix

$$(J_r - \lambda_r I)^2 = \begin{bmatrix} 0 & 1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & \ddots & 0 \\ & & & & 0 \end{bmatrix}$$

- it is easy to see that $(J_r - \lambda_r I)^2 \mathbf{e}_2 = 0$
- continuing in this fashion, we can conclude that

$$(J_r - \lambda_r I)^k \mathbf{e}_k = \mathbf{0}, \quad k = 1, \ldots, n_r - 1$$

- the Jordan form can be used to easily compute powers of a matrix
- for example,

$$A^2 = XJX^{-1}XJX^{-1} = XJ^2X^{-1}$$

and, in general,

$$A^k = XJ^kX^{-1}$$

- due to its structure, it is easy to compute powers of a Jordan block J_r:

$$J_r^k = \begin{bmatrix} \lambda_r & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_r \end{bmatrix}^k = (\lambda_r I + N)^k = \sum_{j=0}^{k} \binom{k}{j} \lambda_r^{k-j} N^j$$

where

$$N = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}$$

is a *nilpotent matrix*, i.e., $N^d = 0$ for some power d

- the binomial expansion above yields, for $k > n_r$,

$$J_r^k = \begin{bmatrix} \lambda_r^k & \binom{k}{1}\lambda_r^{k-1} & \binom{k}{2}\lambda_r^{k-2} & \cdots & \binom{k}{n_r-1}\lambda_r^{k-(n_r-1)} \\ & \ddots & \ddots & & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \vdots \\ & & & & \lambda_r^k \end{bmatrix} \tag{3.2}$$

- for example,

$$\begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}^3 = \begin{bmatrix} \lambda^3 & 3\lambda^2 & 3\lambda \\ 0 & \lambda^3 & 3\lambda^2 \\ 0 & 0 & \lambda^3 \end{bmatrix}$$

- we now consider an application of the Jordan canonical form
  - consider the system of differential equations

    $$\mathbf{y}'(t) = A\mathbf{y}(t), \quad \mathbf{y}(t_0) = \mathbf{y}_0$$

  - using the Jordan form, we can rewrite this system as

    $$\mathbf{y}'(t) = XJX^{-1}\mathbf{y}(t)$$

  - ultiplying through by X^{-1} yields

    $$X^{-1}\mathbf{y}'(t) = JX^{-1}\mathbf{y}(t)$$

    which can be rewritten as

    $$\mathbf{z}'(t) = J\mathbf{z}(t)$$

    where $\mathbf{z} = X^{-1}\mathbf{y}(t)$
  - this new system has the initial condition

    $$\mathbf{z}(t_0) = \mathbf{z}_0 = X^{-1}\mathbf{y}_0$$

  - if we assume that J is a diagonal matrix (which is true in the case where A has a full set of linearly independent eigenvectors), then the system decouples into scalar equations of the form

    $$z_i'(t) = \lambda_i z_i(t),$$

    where λ_i is an eigenvalue of A
  - this equation has the solution

    $$z_i(t) = e^{\lambda_i(t-t_0)}z_i(0),$$

    and therefore the solution to the original system is

    $$\mathbf{y}(t) = X \begin{bmatrix} e^{\lambda_1(t-t_0)} & & \\ & \ddots & \\ & & e^{\lambda_n(t-t_0)} \end{bmatrix} X^{-1}\mathbf{y}_0$$

- Jordan canonical form suffers however from one major defect that makes them useless in practice: they cannot be computed in finite precision or in the presence of rounding errors in general, a result of Golub and Wilkinson
- that is why you won't find a MATLAB function for Jordan canonical form

- matrix 2-norm is also known as the *spectral norm*
- name is connected to the fact that the norm is given by the square root of the largest eigenvalue of $A^{\mathsf{T}}A$, i.e., largest singular value of A (more on this later)
- in general, the *spectral radius* $\rho(A)$ of a matrix $A \in \mathbb{C}^{n\times n}$ is defined in terms of its largest eigenvalue
$$\rho(A) = \max\{|\lambda_i| : A\mathbf{x}_i = \lambda_i \mathbf{x}_i,\ \mathbf{x}_i \neq \mathbf{0}\}$$
- note that the spectral radius does not define a norm on $\mathbb{C}^{n\times n}$
- for example the non-zero matrix
$$J = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$
has $\rho(J) = 0$ since both its eigenvalues are 0
- there are some relationships between the norm of a matrix and its spectral radius
- the easiest one is that
$$\rho(A) \leq \|A\|$$
for any matrix norm that satisfies the inequality $\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{C}^n$, i.e., consistent norm
  - here's a proof:
$$A\mathbf{x}_i = \lambda_i \mathbf{x}_i$$
taking norms,
$$\|A\mathbf{x}_i\| = \|\lambda_i \mathbf{x}_i\| = |\lambda_i|\|\mathbf{x}_i\|$$
therefore
$$|\lambda_i| = \frac{\|A\mathbf{x}_i\|}{\|\mathbf{x}_i\|} \leq \|A\|$$
since this holds for any eigenvalue of A, it follows that
$$\max_i |\lambda_i| = \rho(A) \leq \|A\|$$
  - in particular this is true for any operator norm
  - this is in general not true for norms that do not satisfy the consistency inequality $\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|$ (thanks to Likai Chen for pointing out); for example the matrix
$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$
is orthogonal and therefore $\rho(A) = 1$ but $\|A\|_{\mathsf{H},\infty} = 1/\sqrt{2}$ and so $\rho(A) > \|A\|_{\mathsf{H},\infty}$
  - exercise: show that any eigenvalue of a unitary or an orthogonal matrix must have absolute value 1
- on the other hand, the following characterization is true for *any* matrix norm, even the inconsistent ones
$$\rho(A) = \lim_{m\to\infty} \|A^m\|^{1/m}$$
- we can also get an upper bound for any particular matrix (but not for all matrices)

Theorem 3. *Let $A \in \mathbb{C}^{n\times n}$ and $\varepsilon > 0$. There exists an operator norm $\|\cdot\|_\alpha$ of the form*
$$\|A\|_\alpha = \max_{\mathbf{x}\neq 0} \frac{\|A\mathbf{x}\|_\alpha}{\|\mathbf{x}\|_\alpha},$$
where $\|\cdot\|_\alpha$ is a norm on $\mathbb{C}^n$, such that
$$\|A\|_\alpha \leq \rho(A) + \epsilon.$$

The norm $\|\cdot\|_\alpha$ is dependent on A and ε.

- this result suggests that the largest eigenvalue of a matrix can be easily approximated
- here is an example, let

$$A = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & 1 \\ & & & -1 & 2 \end{bmatrix}$$

  - the eigenvalues of this matrix, which arises frequently in numerical methods for solving differential equations, are known to be

$$\lambda_j = 2 + 2\cos\frac{j\pi}{n+1}, \quad j = 1, 2, \ldots, n$$

    the largest eigenvalue is

$$|\lambda_1| = 2 + 2\cos\frac{\pi}{n+1} \le 4$$

    and $\|A\|_\infty = 4$, so in this case, the ∞-norm provides an excellent approximation
  - on the other hand, suppose

$$A = \begin{bmatrix} 1 & 10^6 \\ 0 & 1 \end{bmatrix}$$

    we have $\|A\|_\infty = 10^6 + 1$, but $\rho(A) = 1$, so in this case the norm yields a poor approximation
  - however, suppose

$$D = \begin{bmatrix} \varepsilon & 0 \\ 0 & 1 \end{bmatrix}$$

    then

$$DAD^{-1} = \begin{bmatrix} 1 & 10^6\varepsilon \\ 0 & 1 \end{bmatrix}$$

    and $\|DAD^{-1}\|_\infty = 1 + 10^{-6}\epsilon$, which for sufficiently small ε, yields a much better approximation to $\rho(DAD^{-1}) = \rho(A)$.
- if $\|A\| < 1$ for some submultiplicative norm, then $\|A^m\| \le \|A\|^m \to 0$ as $m \to \infty$
- since $\|A\|$ is a continuous function of the elements of A, it follows that $A^m \to O$, i.e., every entry of A^m goes to 0
- however, if $\|A\| > 1$, it does not follow that $\|A^m\| \to \infty$
- note that the submultiplicative property implies that

$$\|A^n\| \le \|A\|^n$$

- so if $\|A\| < 1$, then, as $n \to \infty$, $\|A^n\| \to 0$
- if $\|A^n\| \to 0$, then $A^n \to O$, i.e. each entry of A^n converges to 0, by the continuity of norms
- the condition $\|A\| < 1$ is not necessary for $A^n \to 0$
  - example:

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

    has $\|A\|_1 = 1$ but $A^2 = A^3 = \cdots = O$
  - the above example is a *nilpotent* matrix, i.e., has the property that $A^n = O$ for some finite $n \in \mathbb{N}$

– another example:

$$A = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 \end{bmatrix}$$

has $\|A\|_\infty = 1$ but $A^n \to O$

- for example, suppose

$$A = \begin{bmatrix} 0.99 & 10^6 \\ 0 & 0.99 \end{bmatrix}$$

in this case, $\|A\|_\infty > 1$, we claim that because $\rho(A) < 1$, $A^m \to O$ and so $\|A^m\| \to 0$

- let us prove this more generally, in fact we claim the following

Lemma 1. $\lim_{m\to\infty} A^m = O$ *if and only if* $\rho(A) < 1$.

Proof. ($\Rightarrow$) Let $A\mathbf{x} = \lambda\mathbf{x}$ with $\mathbf{x} \neq \mathbf{0}$. Then $A^m\mathbf{x} = \lambda^m\mathbf{x}$. Taking limits

$$\left(\lim_{m\to\infty} \lambda^m\right)\mathbf{x} = \lim_{m\to\infty} \lambda^m\mathbf{x} = \lim_{m\to\infty} A^m\mathbf{x} = \left(\lim_{m\to\infty} A^m\right)\mathbf{x} = O\mathbf{x} = \mathbf{0}.$$

Since $\mathbf{x} \neq \mathbf{0}$, we must have $\lim_{m\to\infty} \lambda^m = 0$ and thus $|\lambda| < 1$. Hence $\rho(A) < 1$.

($\Leftarrow$) Since $\rho(A) < 1$, there exists some operator norm $\|\cdot\|_\alpha$ such that $\|A\|_\alpha < 1$ by Theorem 3. So $\|A^m\|_\alpha \leq \|A\|_\alpha^m \to 0$ and so $A^m \to O$. $\square$

- alternatively, the second part above may also be proved directly via the Jordan form of A and the expression

$$J_r^k = \begin{bmatrix} \lambda_r^k & \binom{k}{1}\lambda_r^{k-1} & \binom{k}{2}\lambda_r^{k-2} & \cdots & \binom{k}{n_r-1}\lambda_r^{k-(n_r-1)} \\ & \ddots & \ddots & & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \vdots \\ & & & & \lambda_r^k \end{bmatrix}$$

for sufficiently large k without using Theorem 3).

- in Homework 1 we will see that if for some operator norm, $\|A\| < 1$, then $I - A$ is nonsingular