# Math 4571 (Advanced Linear Algebra)

## Lecture #27

---

Applications of Diagonalization and the Jordan Canonical Form (Part 1):

- Spectral Mapping and the Cayley-Hamilton Theorem
- Transition Matrices and Markov Chains
- The Spectral Theorem for Hermitian Operators

This material represents §4.4.1 + §4.4.4 +§4.4.5 from the course notes.

# Overview

In this lecture and the next, we discuss a variety of applications of diagonalization and the Jordan canonical form. This lecture will discuss three essentially unrelated topics:

- A proof of the Cayley-Hamilton theorem for general matrices
- Transition matrices and Markov chains, used for modeling iterated changes in systems over time
- The spectral theorem for Hermitian operators, in which we establish that Hermitian operators (i.e., operators with $T^* = T$) are diagonalizable

In the next lecture, we will discuss another fundamental application: solving systems of linear differential equations.

# Cayley-Hamilton, I

First, we establish the Cayley-Hamilton theorem for arbitrary matrices:

**Theorem (Cayley-Hamilton)**

*If $p(x)$ is the characteristic polynomial of a matrix $A$, then $p(A)$ is the zero matrix $\mathbf{0}$.*

The same result holds for the characteristic polynomial of a linear operator $T : V \to V$ on a finite-dimensional vector space.

# Cayley-Hamilton, II

Proof:

- Since the characteristic polynomial of a matrix does not depend on the underlying field of coefficients, we may assume that the characteristic polynomial factors completely over the field (i.e., that all of the eigenvalues of $A$ lie in the field) by replacing the field with its algebraic closure.
- Then by our results, the Jordan canonical form of $A$ exists.
- Let $J = Q^{-1}AQ$ with $J$ in Jordan canonical form, and $p(x) = (x - \lambda_1)^{d_1} \cdots (x - \lambda_k)^{d_k}$ be the char. poly. of $A$.
- We first claim that for a $d \times d$ Jordan block matrix $J_i$ with associated eigenvalue $\lambda_i$, we have $(J_i - \lambda_i I)^d = \mathbf{0}$.
- To see this, let $\{\mathbf{v}_0, \mathbf{v}_1 \ldots, \mathbf{v}_{d-1}\}$ be a Jordan basis and $S = A - \lambda_i I$: then $\mathbf{v}_{i+1} = S\mathbf{v}_i$ and $S\mathbf{v}_{d-1} = \mathbf{0}$.
- It is then easy to see that $S^d$ is zero on all the $\mathbf{v}_i$.

# Cayley-Hamilton, III

Proof (continued):

- Now, if $J_i$ is any $d \times d$ Jordan block in $J$ of eigenvalue $\lambda_i$, the characteristic polynomial of $A$ is divisible by $(t - \lambda_i)^d$, since $\lambda_i$ occurs as an eigenvalue with multiplicity at least $d$
- Therefore, $p(J_i) = (J_i - \lambda_1 I)^{d_1} \cdots (J_i - \lambda_i I)^{d_i} \cdots (J_i - \lambda_k I)^{d_k}$.
- But by the calculation on the previous slide, $(J_i - \lambda_i I)^{d_i} = \mathbf{0}$, so that means $p(J_i) = \mathbf{0}$.
- We then see $p(J) = \begin{bmatrix} p(J_1) & & \\ & \ddots & \\ & & p(J_n) \end{bmatrix} = \mathbf{0}$.
- Finally, $p(A) = Q[p(J)]Q^{-1} = Q(\mathbf{0})Q^{-1} = \mathbf{0}$, as required.

# Spectral Mapping, I

Related to the ideas in our proof is the following useful result about eigenvalues of polynomials of operators:

**Theorem (Spectral Mapping)**

*If $T : V \to V$ is a linear operator on an n-dimensional vector space having eigenvalues $\lambda_1, \dots, \lambda_n$ (counted with multiplicity), then for any polynomial $q(x)$, the eigenvalues of $q(T)$ are $q(\lambda_1), \dots, q(\lambda_n)$.*

The set of eigenvalues of an operator is called its <u>spectrum</u>. The spectral mapping theorem, then, tells us how the spectrum of an operator is transformed by applying a polynomial map to the operator (whence the name). In fact, it still holds if $q$ is any convergent power series (e.g., the exponential).

# Spectral Mapping, II

Proof:

- As in the proof of Cayley-Hamilton we may assume that the underlying field is algebraically closed.
- Now let $\beta$ be a basis for $V$ such that $[T]_\beta^\beta = J$ is in Jordan canonical form. Then $[q(T)]_\beta^\beta = q(J)$, so it suffices to find the eigenvalues of $q(J)$.
- Observe that if $B$ is any upper-triangular matrix with diagonal entries $b_{1,1}, \dots, b_{n,n}$, then $q(B)$ is also upper-triangular and has diagonal entries $q(b_{1,1}), \dots, q(b_{n,n})$.
- Applying this to the Jordan canonical form $J$, we see that the diagonal entries of $q(J)$ are $q(\lambda_1), \dots, q(\lambda_n)$, and the diagonal entries of any upper-triangular matrix are its eigenvalues (counted with multiplicity).

# Spectral Mapping, III

<u>Example</u>: Verify the spectral mapping theorem for $A = \begin{bmatrix} 3 & 5 \\ 7 & 5 \end{bmatrix}$ and $q(t) = t^2 - 2t$.

# Spectral Mapping, III

Example: Verify the spectral mapping theorem for $A = \begin{bmatrix} 3 & 5 \\ 7 & 5 \end{bmatrix}$ and $q(t) = t^2 - 2t$.

- The characteristic polynomial of $A$ is $p_A(t) = (t-3)(t-5) - 35 = (t-10)(t+2)$ so the eigenvalues are $\lambda = -2, 10$.
- Also, $q(A) = \begin{bmatrix} 38 & 30 \\ 42 & 50 \end{bmatrix}$ with characteristic polynomial $p_{q(A)}(t) = (t-38)(t-50) - 1260 = (t-8)(t-80)$.
- Thus, the eigenvalues of $q(A)$ are $\lambda = 8, 80$.
- Indeed, we see that $q(-2) = 8$ and $q(10) = 80$, in agreement with the spectral mapping theorem.

# Transition Matrices, I

We next discuss how to use linear algebra to model the behavior of iterated systems. Our treatment will be fairly simplistic, but there are many ways to extend these basic ideas to create much better models.

The basic setup common to many such models is that we have a system in an initial state, and an iteration map that tells us how the state of the system changes (evolves) over time.

We will discuss the situation where the iteration map is a linear transformation, in which case we are studying the behavior of the system after we apply this linear transformation repeatedly.

# Transition Matrices, II

To give a specific example, consider a state with two cities $A$ and $B$ whose populations flow back and forth over time.

- Suppose that after one year passes, a resident of city $A$ has a 10% chance of moving to city $B$ and a 90% chance of staying in city $A$, while a resident of city $B$ has a 30% change of moving to $A$ and a 70% chance of staying in $B$.
- We would like to know what will happen to the relative populations of cities $A$ and $B$ over a long period of time.
- If city $A$ has a population of $A_{\text{old}}$ and city $B$ has a population of $B_{\text{old}}$, then one year later, we can see that city $A$'s population will be $A_{\text{new}} = 0.9A_{\text{old}} + 0.3B_{\text{old}}$, while $B$'s population will be $B_{\text{new}} = 0.1A_{\text{old}} + 0.7B_{\text{old}}$.

# Transition Matrices, III

- Observe that we can describe this map using matrix multiplication: $\begin{bmatrix} A_{\text{new}} \\ B_{\text{new}} \end{bmatrix} = \begin{bmatrix} 0.9 & 0.3 \\ 0.1 & 0.7 \end{bmatrix} \begin{bmatrix} A_{\text{old}} \\ B_{\text{old}} \end{bmatrix}$.
- Since the population one year into the future is obtained by left-multiplying the population vector by $M = \begin{bmatrix} 0.9 & 0.3 \\ 0.1 & 0.7 \end{bmatrix}$, the population $k$ years into the future can then be obtained by left-multiplying the population vector by $M^k$.
- By diagonalizing this matrix, we can easily compute $M^k$, and thus analyze the behavior of the population as time extends forward. (If it is not diagonalizable, we would have to resort to using its Jordan canonical form.)

# Transition Matrices, IV

- In this case, $M$ is diagonalizable: $M = QDQ^{-1}$ with $D = \begin{bmatrix} 1 & 0 \\ 0 & 3/5 \end{bmatrix}$ and $Q = \begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix}$.
- Then $M^k = QD^kQ^{-1}$, and as $k \to \infty$, we see that $D^k \to \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, so $M^k \to Q \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} Q^{-1} = \begin{bmatrix} 3/4 & 3/4 \\ 1/4 & 1/4 \end{bmatrix}$.
- From this calculation, we can see that as time extends on, the cities' populations will approach the situation where 3/4 of the residents live in city $A$ and 1/4 of the residents live in city $B$, regardless of the cities' original populations!
- Notice that this "steady-state" solution where the cities' populations both remain constant represents an eigenvector of the original matrix with eigenvalue $\lambda = 1$.

# Transition Matrices, V

Systems in which members of a set (in our example, residents of the cities) are identified as belonging to one of several states that can change over time, is known as a stochastic process.

- If, as in our example, the probabilities of changing from one state to another are independent of time, the system is called a Markov chain.
- Markov chains and their continuous analogues (known as Markov processes) arise (for example) in probability problems involving repeated wagers or random walks, in economics modeling the flow of goods among industries and nations, in biology modeling the gene frequencies in populations, and in civil engineering modeling the arrival of people to buildings.
- A Markov chain model was also used for one of the original versions of the PageRank algorithm used by Google to rank internet search results.

# Transition Matrices, VI

**Definition**

A square matrix whose entries are nonnegative and whose columns sum to 1 is called a <u>transition matrix</u> (or a <u>stochastic matrix</u>).

- Equivalently, a square matrix $M$ is a transition matrix precisely when $M^T\mathbf{v} = \mathbf{v}$, where $\mathbf{v}$ is the column vector of all 1s.
- From this description, we can see that $\mathbf{v}$ is an eigenvector of $M^T$ of eigenvalue 1, and since $M^T$ and $M$ have the same characteristic polynomial, we conclude that $M$ has 1 as an eigenvalue.

## Transition Matrices, VII

- If it were true that $M$ were diagonalizable and every eigenvalue of $M$ had absolute value less than 1 (except for the eigenvalue 1), then we could apply the same argument as we did in the example to conclude that the powers of $M$ approached a limit.
- Unfortunately, this is not true in general: the transition matrix $M = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ has $M^2$ equal to the identity matrix, so odd powers of $M$ are equal to $M$ while even powers are equal to the identity. (In this case, the eigenvalues of $M$ are 1 and $-1$.)

# Transition Matrices, VIII

Fortunately, the argument does apply to many transition matrices:

**Theorem (Markov Chains)**

*If $M$ is a transition matrix, then every eigenvalue $\lambda$ of $M$ has $|\lambda| \leq 1$. Furthermore, if some power of $M$ has all entries positive, then the only eigenvalue of $M$ of absolute value 1 is $\lambda = 1$, and the 1-eigenspace has dimension 1. In such a case, the "matrix limit" $\lim_{k\to\infty} M^k$ exists and has all columns equal to a "steady-state" solution of the Markov chain whose transition matrix is $M$.*

# Transition Matrices, IX

Examples:

- For $A = \begin{bmatrix} 1/2 & 1/3 \\ 1/2 & 2/3 \end{bmatrix}$, the eigenvalues are $\lambda = 1, 1/6$, and the matrix limit $\lim_{k\to\infty} A^k = \begin{bmatrix} 2/5 & 2/5 \\ 3/5 & 3/5 \end{bmatrix}$ as can be verified by diagonalizing $A$. (Note $(2/5, 3/5)$ is a 1-eigenvector of $A$.)

# Transition Matrices, IX

Examples:

- For $A = \begin{bmatrix} 1/2 & 1/3 \\ 1/2 & 2/3 \end{bmatrix}$, the eigenvalues are $\lambda = 1, 1/6$, and the matrix limit $\lim_{k\to\infty} A^k = \begin{bmatrix} 2/5 & 2/5 \\ 3/5 & 3/5 \end{bmatrix}$ as can be verified by diagonalizing $A$. (Note $(2/5, 3/5)$ is a 1-eigenvector of $A$.)
- For $B = \begin{bmatrix} 1/4 & 1/3 & 1/5 \\ 1/2 & 2/3 & 1/5 \\ 1/4 & 0 & 3/5 \end{bmatrix}$, eigenvalues $\lambda = 1, \frac{31\pm\sqrt{1201}}{120}$, and $\lim_{k\to\infty} B^k = \frac{1}{28}\begin{bmatrix} 8 & 8 & 8 \\ 15 & 15 & 15 \\ 5 & 5 & 5 \end{bmatrix}$ as can be verified by diagonalizing. (Note $(8, 15, 5)/28$ is a 1-eigenvector of $B$.)

# Hermitian Operators, I

We now use establish fundamental result about the diagonalizability of self-adjoint operators known as the spectral theorem. First, we define the necessary family of operators:

**Definition**

If $T : V \to V$ is a linear transformation and $T^*$ exists, we say that $T$ is Hermitian (or self-adjoint) if $T^* = T$, and we say that $T$ is skew-Hermitian if $T^* = -T$.

Examples:

- Left-multiplication by a real symmetric matrix is Hermitian.
- Left-multiplication by a real skew-symmetric matrix is skew-Hermitian.
- The identity transformation is Hermitian.

# Hermitian Operators, II

We extend this definition to matrices in the natural way:

- A matrix $A$ is (skew)-Hermitian if $A = [T]_\beta^\beta$ for some basis $\beta$ of $V$ and some (skew)-Hermitian linear transformation $T$.
- As we showed before, the matrix associated to $T^*$ is $A^*$, the conjugate-transpose of $A$, so $A$ is Hermitian precisely when $A = A^*$ and $A$ is skew-Hermitian precisely when $A = -A^*$.
- In particular, if $A$ is a matrix with real entries, then $A$ is Hermitian if and only if $A = A^T$ (i.e., $A$ is a symmetric matrix), and $A$ is skew-Hermitian if and only if $A = -A^T$ (i.e., $A$ is a skew-symmetric matrix).

# Hermitian Operators, III

Hermitian linear operators (and Hermitian matrices) have a variety of very nice properties. Here are some fundamental ones:

**Theorem (Properties of Hermitian Operators)**

*Suppose $V$ is a finite-dimensional inner product space and $T : V \to V$ is Hermitian. Then the following hold:*

1. *For any $\mathbf{v} \in V$, $\langle T(\mathbf{v}), \mathbf{v} \rangle$ is a real number.*
2. *All eigenvalues of $T$ are real numbers.*
3. *Eigenvectors of $T$ with different eigenvalues are orthogonal.*
4. *Every generalized eigenvector of $T$ is an eigenvector of $T$.*

# Hermitian Operators, IV

Proofs:

1. For any $\mathbf{v} \in V$, $\langle T(\mathbf{v}), \mathbf{v} \rangle$ is a real number.
   - We have $\langle T(\mathbf{v}), \mathbf{v} \rangle = \langle \mathbf{v}, T^*(\mathbf{v}) \rangle = \langle \mathbf{v}, T(\mathbf{v}) \rangle = \overline{\langle T(\mathbf{v}), \mathbf{v} \rangle}$.
   - Thus, $\langle T(\mathbf{v}), \mathbf{v} \rangle$ is equal to its complex conjugate, hence is real.
2. All eigenvalues of $T$ are real numbers.
   - Proof: Suppose $\lambda$ is an eigenvalue with eigenvector $\mathbf{v} \neq \mathbf{0}$.
   - Then $\langle T(\mathbf{v}), \mathbf{v} \rangle = \langle \lambda \mathbf{v}, \mathbf{v} \rangle = \lambda \langle \mathbf{v}, \mathbf{v} \rangle$ is real.
   - Since $\mathbf{v}$ is not the zero vector we conclude that $\langle \mathbf{v}, \mathbf{v} \rangle$ is a nonzero real number, so $\lambda = \langle T(\mathbf{v}), \mathbf{v} \rangle / \langle \mathbf{v}, \mathbf{v} \rangle$ is also real.
3. Eigenvectors of $T$ with different eigenvalues are orthogonal.
   - Proof: Suppose that $T\mathbf{v}_1 = \lambda_1 \mathbf{v}_1$ and $T\mathbf{v}_2 = \lambda_2 \mathbf{v}_2$. Then $\lambda_1 \langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle T\mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, T^*\mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \lambda_2 \mathbf{v}_2 \rangle = \lambda_2 \langle \mathbf{v}_1, \mathbf{v}_2 \rangle$ since $\lambda_2$ is real. But since $\lambda_1 \neq \lambda_2$, this means $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$.

# Hermitian Operators, V

Proofs:

4. Every generalized eigenvector of $T$ is an eigenvector of $T$.
   - We show by induction that if $(T - \lambda I)^k \mathbf{w} = \mathbf{0}$ then in fact $(T - \lambda I)\mathbf{w} = \mathbf{0}$, so $\mathbf{w}$ is in fact an eigenvector.
   - For the base case we take $k = 2$, so that $(\lambda I - T)^2 \mathbf{w} = \mathbf{0}$. Then since $\lambda$ is an eigenvalue of $T$ and therefore real, we have

$$\begin{aligned} \mathbf{0} = \left\langle (T - \lambda I)^2 \mathbf{w}, \mathbf{w} \right\rangle &= \left\langle (T - \lambda I)\mathbf{w}, (T - \lambda I)^* \mathbf{w} \right\rangle \\ &= \left\langle (T - \lambda I)\mathbf{w}, (T^* - \overline{\lambda} I)\mathbf{w} \right\rangle \\ &= \left\langle (T - \lambda I)\mathbf{w}, (T - \lambda I)\mathbf{w} \right\rangle \end{aligned}$$

     and thus the inner product of $(T - \lambda I)\mathbf{w}$ with itself is zero, so $(T - \lambda I)\mathbf{w}$ must be zero.
   - For the inductive step, observe that $(T - \lambda I)^{k+1} \mathbf{w} = \mathbf{0}$ implies $(T - \lambda I)^k \left[ (T - \lambda I)\mathbf{w} \right] = \mathbf{0}$, and therefore by the inductive hypothesis this means $(T - \lambda I) \left[ (T - \lambda I)\mathbf{w} \right] = \mathbf{0}$, or equivalently, $(T - \lambda I)^2 \mathbf{w} = \mathbf{0}$. Applying the result for $k = 2$ from above yields $(T - \lambda I)\mathbf{w} = \mathbf{0}$, as required.

# Spectral Theorem, I

Using these basic properties, we can prove that Hermitian operators are diagonalizable, and in fact that they are diagonalizable in a particularly nice way:

**Theorem (Spectral Theorem)**

*Suppose $V$ is a finite-dimensional inner product space over $\mathbb{R}$ or $\mathbb{C}$ and $T : V \to V$ is a Hermitian linear transformation. Then $V$ has an orthonormal basis $\beta$ of eigenvectors of $T$, so in particular, $T$ is diagonalizable.*

The equivalent formulation for Hermitian matrices is: every Hermitian matrix $A$ can be written as $A = U^{-1}DU$ where $D$ is a real diagonal matrix and $U$ is a unitary matrix (i.e., a matrix satisfying $U^* = U^{-1}$).

# Spectral Theorem, II

Proof:

- By our properties, every eigenvalue of $T$ is real hence lies in the scalar field.
- Then every generalized eigenvector of $T$ is an eigenvector of $T$, and so since $V$ has a basis of generalized eigenvectors, it has a basis of eigenvectors and is therefore diagonalizable.
- For the orthonormal basis, start with a basis for each eigenspace, and then apply Gram-Schmidt, yielding an orthonormal basis for each eigenspace.
- Since $T$ is diagonalizable, the union of these bases is a basis for $V$: furthermore, each of the vectors has norm 1, and they are all orthogonal by the orthogonal result above.
- Thus, the union is an orthonormal basis of eigenvectors of $T$.

# Spectral Theorem, III

We will make a few remarks about the spectral theorem:

- The converse of this theorem is not quite true: if $V$ has an orthonormal basis of eigenvectors of $T$, then $T$ is not necessarily Hermitian.
- The correct general converse theorem is that $V$ has an orthonormal basis of eigenvectors of $T$ if and only if $T$ is a normal operator, meaning that $T^*T = TT^*$.
- The spectral theorem shows that $V$ is the direct sum of the eigenspaces of $T$, meaning that the action of $T$ on $V$ can be decomposed into simple pieces (acting as scalar multiplication), with one piece coming from each piece of the spectrum. (This is the reason for the name of the theorem.)
- Most of these results also extend to skew-Hermitian operators upon noting that $T$ is skew-Hermitian iff $iT$ is Hermitian.

# Spectral Theorem, IV

As a corollary, we obtain the following extremely useful computational fact:

**Corollary (Real Symmetric Matrices Are Diagonalizable)**

*Every real symmetric matrix has real eigenvalues and is diagonalizable over the real numbers.*

Proof: This follows immediately from the spectral theorem since a real symmetric matrix is Hermitian.

# Spectral Theorem, V

Examples:

- The real symmetric matrix $A = \begin{bmatrix} 3 & 6 \\ 6 & 8 \end{bmatrix}$ has eigenvalues $\lambda = -1, 12$ and has $A = UDU^{-1}$ where $D = \begin{bmatrix} -1 & 0 \\ 0 & 12 \end{bmatrix}$ and $U = \frac{1}{\sqrt{13}} \begin{bmatrix} -3 & 2 \\ 2 & 3 \end{bmatrix}$.

# Spectral Theorem, V

Examples:

- The real symmetric matrix $A = \begin{bmatrix} 3 & 6 \\ 6 & 8 \end{bmatrix}$ has eigenvalues $\lambda = -1, 12$ and has $A = UDU^{-1}$ where $D = \begin{bmatrix} -1 & 0 \\ 0 & 12 \end{bmatrix}$ and $U = \frac{1}{\sqrt{13}} \begin{bmatrix} -3 & 2 \\ 2 & 3 \end{bmatrix}$.
- The Hermitian matrix $A = \begin{bmatrix} 6 & 2-i \\ 2+i & 2 \end{bmatrix}$ has eigenvalues $\lambda = 1, 7$ and has $A = UDU^{-1}$ where $D = \begin{bmatrix} 1 & 0 \\ 0 & 7 \end{bmatrix}$ and $U = \frac{1}{\sqrt{30}} \begin{bmatrix} 5 & 2-i \\ 2+i & -5 \end{bmatrix}$.

# Spectral Theorem, VI

As a final remark, we will note that although real symmetric matrices are diagonalizable (and complex Hermitian matrices are diagonalizable), it is *not* true that complex symmetric matrices are always diagonalizable.

# Spectral Theorem, VI

As a final remark, we will note that although real symmetric matrices are diagonalizable (and complex Hermitian matrices are diagonalizable), it is *not* true that complex symmetric matrices are always diagonalizable.

<u>Non-Example</u>:

- The complex symmetric matrix $\begin{bmatrix} 1 & i \\ i & -1 \end{bmatrix}$ is not diagonalizable.
- This follows from the observation that its trace and determinant are both zero, but since it is not the zero matrix, the only possibility for its Jordan form is $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$.
- Since its Jordan form is not diagonal, it is not diagonalizable.

# Summary

We proved the general Cayley-Hamilton theorem and the related spectral mapping theorem.

We discussed transition matrices and Markov chains, and some of their applications to modeling the behavior of iterated systems.

We established some basic properties of Hermitian operators, and then proved the spectral theorem and deduced the important consequence that real symmetric matrices are diagonalizable.

Next lecture: Applications of the Jordan Canonical Form (part 2)