Math 4571 (Advanced Linear Algebra)

Lecture #24

Diagonalization:

- Review of Eigenvalues and Eigenvectors
- Diagonalization
- The Cayley-Hamilton Theorem

This material represents §4.1-§4.2 from the course notes.

Eigenvalues, I

Recall our definition of eigenvalues and eigenvectors for linear transformations:

Definition

If $T : V \to V$ is a linear transformation, a nonzero vector $\mathbf{v}$ with $T(\mathbf{v}) = \lambda \mathbf{v}$ is called an <u>eigenvector</u> of T, and the corresponding scalar $\lambda \in F$ is called an <u>eigenvalue</u> of T.

By convention, the zero vector $\mathbf{0}$ is not an eigenvector.

Definition

If $T : V \to V$ is a linear transformation, then for any fixed value of $\lambda \in F$, the set E_λ of vectors in V satisfying $T(\mathbf{v}) = \lambda \mathbf{v}$ is a subspace of V called the <u>λ-eigenspace</u>.

Eigenvalues, II

Examples:

- $T : \mathbb{R}^2 \to \mathbb{R}^2$ is the map with $T(x, y) = \langle 2x + 3y, x + 4y \rangle$, then $\mathbf{v} = \langle 3, -1 \rangle$ is a 1-eigenvector since $T(\mathbf{v}) = \langle 3, -1 \rangle = \mathbf{v}$, and $\mathbf{w} = \langle 1, 1 \rangle$ is an 5-eigenvector since $T(\mathbf{w}) = \langle 5, 5 \rangle = 5\mathbf{w}$.

- If $T : M_{n \times n}(\mathbb{R}) \to M_{n \times n}(\mathbb{R})$ is the transpose map, then the 1-eigenspace of T consists of the symmetric matrices, while the (-1)-eigenspace consists of the skew-symmetric matrices.

- If V is the space of infinitely-differentiable functions and $D : V \to V$ is the derivative map, then for any real r, the function $f(x) = e^{rx}$ is an eigenfunction with eigenvalue r since $D(e^{rx}) = re^{rx}$.

- If $I : \mathbb{R}[x] \to \mathbb{R}[x]$ is the integration map $I(p) = \int_0^x p(t)\, dt$, then I has no eigenvectors, since $I(p)$ always has a larger degree than p, so $I(p)$ cannot equal λp for any $\lambda \in \mathbb{R}$.

Eigenvalues, III

We can also compute eigenvalues and eigenvectors of matrices:

Definition

If $A \in M_{n \times n}(F)$, then a nonzero vector $\mathbf{x}$ with $A\mathbf{x} = \lambda\mathbf{x}$ is called a (left) <u>eigenvector</u> of A, and the corresponding scalar $\lambda \in F$ is called an <u>eigenvalue</u> of A.

This procedure is essentially equivalent to computing eigenvalues and eigenvectors of arbitrary linear transformations:

Proposition (Eigenvalues and Matrices)

Suppose V is a finite-dimensional vector space with ordered basis β and that $T : V \to V$ is linear. Then $\mathbf{v}$ is an eigenvector of T with eigenvalue λ if and only if $[\mathbf{v}]_\beta$ is an eigenvector of left-multiplication by $[T]_\beta^\beta$ with eigenvalue λ.

The proof is immediate from our results on associated matrices.

Eigenvalues, IV

Examples:

- If $A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$, then $\mathbf{x} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$ is a 1-eigenvector of A,

  because $A\mathbf{x} = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \mathbf{x}$.

- If $B = \begin{bmatrix} 2 & -4 & 5 \\ 2 & -2 & 5 \\ 2 & 1 & 2 \end{bmatrix}$, then $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ is a 4-eigenvector of

  B, because $B\mathbf{x} = \begin{bmatrix} 2 & -4 & 5 \\ 2 & -2 & 5 \\ 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \\ 8 \end{bmatrix} = 4\mathbf{x}$.

Eigenvalues, V

For real matrices, eigenvalues may be non-real numbers. Because of this, we will often implicitly assume that the underlying field is algebraically closed (e.g., $\mathbb{C}$) unless otherwise specified.

Example:

- If $A = \begin{bmatrix} 6 & 3 & -2 \\ -2 & 0 & 0 \\ 6 & 4 & 2 \end{bmatrix}$, the vector $\mathbf{x} = \begin{bmatrix} 1 - i \\ 2i \\ 2 \end{bmatrix}$ is an
  eigenvector of A with eigenvalue $1 + i$, because
  $$A\mathbf{x} = \begin{bmatrix} 6 & 3 & -2 \\ -2 & 0 & 0 \\ 6 & 4 & -2 \end{bmatrix} \begin{bmatrix} 1 - i \\ 2i \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 + 2i \\ 2 + 2i \end{bmatrix} = (1 + i)\mathbf{x}.$$

Eigenvalues, VI

We can exploit properties of determinants to give an explicit procedure for calculating eigenvalues:

Definition

For an $n \times n$ matrix A, the degree-n polynomial
$p(t) = \det(tI_n - A)$ is called the <u>characteristic polynomial</u> of A.

Proposition (Computing Eigenvalues)

If A is an $n \times n$ matrix, then the eigenvalues of A are precisely the roots of the characteristic polynomial of A.

<u>Proof</u>:

- First, $A\mathbf{v} = \lambda\mathbf{v}$ is equivalent to $(\lambda I_n - A)\mathbf{v} = \mathbf{0}$.
- Then by our results on determinants, there is a nonzero vector $\mathbf{v}$ in the nullspace of $\lambda I_n - A$ if and only if $\det(\lambda I_n - A) = 0$, which is to say, when $p(\lambda) = 0$.

Eigenvalues, VII

Example: Find the eigenvalues of $A = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix}$.

- First we compute the characteristic polynomial
  $\det(tI - A) = \begin{vmatrix} t - 3 & -1 \\ -2 & t - 4 \end{vmatrix} = t^2 - 7t + 10$.

- The eigenvalues are then the zeroes of this polynomial. Since
  $t^2 - 7t + 10 = (t - 2)(t - 5)$ we see that the zeroes are $t = 2$
  and $t = 5$, meaning that the eigenvalues are $\boxed{2 \text{ and } 5}$.

Once we have found the eigenvalues, it is easy to compute a basis
of each eigenspace, since that is the same as finding a basis for the
nullspace of $\lambda I - A$.

Eigenvalues, VIII

Example: Find a basis for each eigenspace of $A = \begin{bmatrix} 2 & 2 \\ 3 & 1 \end{bmatrix}$.

- We compute $p(t) = \det(tI - A) = t^2 - 3t - 4 = (t-4)(t+1)$ so the eigenvalues are $\lambda = -1, 4$.

- For $\lambda = -1$, we want the kernel of $-I - A = \begin{bmatrix} -3 & -2 \\ -3 & -2 \end{bmatrix}$.

- The row-echelon form is $\begin{bmatrix} -3 & -2 \\ 0 & 0 \end{bmatrix}$, so the (-1)-eigenspace is 1-dimensional and is spanned by $\begin{bmatrix} -2 \\ 3 \end{bmatrix}$.

- Similarly, for $\lambda = 4$, the nullspace of $4 - A = \begin{bmatrix} 2 & -2 \\ -3 & 3 \end{bmatrix}$ is 1-dimensional and is spanned by $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Eigenvalues, IX

Here are some other useful facts about eigenvalues and eigenvectors (proofs of all of these facts are in the notes, along with numerous extra examples of calculations):

- If A is a real matrix and $\mathbf{v}$ is an eigenvector with a complex eigenvalue λ, then the complex conjugate $\overline{\mathbf{v}}$ is an eigenvector with eigenvalue $\overline{\lambda}$.
- The product of the eigenvalues of A is the determinant of A.
- The sum of the eigenvalues of A equals the trace of A.
- If λ is an eigenvalue of the matrix A which appears exactly k times as a root of the characteristic polynomial, then the dimension of the eigenspace corresponding to λ is at least 1 and at most k.

Diagonalization, I

Our motivation for discussing eigenvalues and eigenvectors is to characterize when a linear transformation can be diagonalized:

Definition

A linear operator $T : V \to V$ on a finite-dimensional vector space V is <u>diagonalizable</u> if there exists a basis β of V such that the associated matrix $[T]_\beta^\beta$ is a diagonal matrix.

By writing down explicitly what this means, we see that T is diagonalizable if and only if the vectors in the basis β are all eigenvectors of T.

Diagonalization, II

We can also formulate diagonalization for matrices:

- If A is an $n \times n$ matrix, then A is the associated matrix of $T : F^n \to F^n$ given by left-multiplication by A.
- Then we say A is diagonalizable when T is diagonalizable.
- By our results on change of basis, this is equivalent to saying that there exists an invertible matrix $Q \in M_{n \times n}(F)$, namely the change-of-basis matrix $Q = [I]_{\gamma}^{\beta}$, for which $Q^{-1}AQ = [I]_{\gamma}^{\beta}[T]_{\gamma}^{\gamma}[I]_{\beta}^{\gamma} = [T]_{\beta}^{\beta}$ is a diagonal matrix.

Definition

An $n \times n$ matrix $A \in M_{n \times n}(F)$ is <u>diagonalizable</u> over F if there exists an invertible $n \times n$ matrix $Q \in M_{n \times n}(F)$ for which $Q^{-1}AQ$ is a diagonal matrix.

Interlude on Similarity, I

More compactly, A is diagonalizable if it is similar to a diagonal matrix. Similarity also respects most other properties:

Proposition (Characteristic Polynomials and Similarity)

If A and B are similar, then they have the same characteristic polynomial, determinant, trace, and eigenvalues (and their eigenvalues have the same multiplicities).

Proof:

- Suppose $B = Q^{-1}AQ$. Then $\det(tI - B)$
  $= \det(Q^{-1}(tI)Q - Q^{-1}AQ) = \det(Q^{-1}(tI - A)Q)$
  $= \det(Q^{-1})\det(tI - A)\det(Q) = \det(tI - A)$.

- Thus, A and B have the same characteristic polynomial.

- The determinant, trace, and eigenvalues are all obtained from the characteristic polynomial, so they are also equal.

Interlude on Similarity, II

The eigenvectors for similar matrices are also closely related:

Proposition (Eigenvectors and Similarity)

If $B = Q^{-1}AQ$, then $\mathbf{v}$ is an eigenvector of B with eigenvalue λ if and only if $Q\mathbf{v}$ is an eigenvector of A with eigenvalue λ.

Proof:

- The case where $\mathbf{v} = \mathbf{0}$ is trivial so assume $\mathbf{v} \neq \mathbf{0}$.
- First suppose $\mathbf{v}$ is an eigenvector of B with eigenvalue λ.
- Then $A(Q\mathbf{v}) = Q(Q^{-1}AQ)\mathbf{v} = Q(B\mathbf{v}) = Q(\lambda\mathbf{v}) = \lambda(Q\mathbf{v})$, meaning that $Q\mathbf{v}$ is an eigenvector of A with eigenvalue λ.
- Now suppose $Q\mathbf{v}$ is an eigenvector of A with eigenvalue λ.
- Then $B\mathbf{v} = Q^{-1}A(Q\mathbf{v}) = Q^{-1}\lambda(Q\mathbf{v}) = \lambda(Q^{-1}Q\mathbf{v}) = \lambda\mathbf{v}$, so $\mathbf{v}$ is an eigenvector of B with eigenvalue λ.

Diagonalization, III

Back to our discussion of diagonalization. We now give a characterization of diagonalizable transformations. First step:

Theorem (Independent Eigenvectors)

If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are eigenvectors of T associated to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent.

Proof (induction on n, base case $n = 1$ is trivial):

- For the inductive step, suppose $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$ for eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ with distinct $\lambda_1, \lambda_2, \ldots, \lambda_n$.
- Applying $T - \lambda_1 I$ to both sides yields
  $a_2(\lambda_2 - \lambda_1)\mathbf{v}_2 + a_3(\lambda_3 - \lambda_1)\mathbf{v}_3 + \cdots + a_n(\lambda_n - \lambda_1)\mathbf{v}_n = \mathbf{0}$.
- By induction, all of the coefficients $a_i(\lambda_i - \lambda_1)$ must be zero, so all of the $a_i = 0$ for $2 \leq i \leq n$. But then we just get $a_1\mathbf{v}_1 = 0$ so $a_1 = 0$ also. Thus the $\mathbf{v}_i$ are independent.

Diagonalization, IV

Now we can establish our diagonalizability criterion for matrices:

Theorem (Diagonalizability Criterion)

A matrix $A \in M_{n \times n}(F)$ is diagonalizable (over F) if and only if the characteristic polynomial factors into a product of linear terms in $F[x]$ and, for each eigenvalue λ, the dimension of the λ-eigenspace is equal to the multiplicity of λ as a root of the characteristic polynomial.

More informally, a matrix is diagonalizable if and only if all of the roots of the characteristic polynomial are in F (rather than some larger field) and all of the eigenspaces have the "maximal possible" dimension according to the multiplicity of that eigenvalue as a root.

Diagonalization, V

Proof:

- If A is diagonalizable, then the diagonal entries are the eigenvalues of A, so they must all lie in the scalar field F.
- For each λ_i let b_i be the dimension of the λ_i-eigenspace and d_i be the multiplicity as a root of the char. polynomial.
- Then, since eigenvectors with different eigenvalues are linearly independent, $V = F^n$ has a basis of eigenvectors if and only if the sum Σb_i is equal to n.
- But as we also proved, $b_i \leq d_i$ for each i.
- Also, Σd_i is the sum of the exponents of all the linear terms appearing in the factorization of the char. polynomial. This is at most n (the degree of the char. polynomial) and equals n only when all of the terms in the factorization are linear.
- Therefore, $\Sigma b_i = n$ if and only if $b_i = d_i$ for each i and all of the terms in the factorization are linear.

Diagonalization, VI

In particular, we can give an easy-to-test sufficient condition for
diagonalizability:

Corollary

*If $A \in M_{n \times n}(F)$ has n distinct eigenvalues in F, then A is
diagonalizable over F.*

Proof:

- Every eigenvalue must occur with multiplicity 1 as a root of
  the characteristic polynomial, since there are n eigenvalues
  and the sum of their multiplicities is also n.
- Then the dimension of each eigenspace is equal to 1 (since it
  is always between 1 and the multiplicity)
- Hence by the diagonalizability theorem, A is diagonalizable.

Diagonalization, VII

<u>Example</u>: For $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x, y) = \langle -2y, 3x + 5y \rangle$, show that T is diagonalizable and find a diagonalizing basis.

- The associated matrix is $A = \begin{bmatrix} 0 & -2 \\ 3 & 5 \end{bmatrix}$.

- For the characteristic polynomial, we compute $\det(tI - A) = t^2 - 5t + 6 = (t - 2)(t - 3)$, so the eigenvalues are therefore $\lambda = 2, 3$. Since the eigenvalues are distinct we know that T is diagonalizable.

- A short calculation yields that $\langle 1, -1 \rangle$ is a basis for the 2-eigenspace, and that $\langle -2, 3 \rangle$ is a basis for the 3-eigenspace.

- Thus, for $\beta = \boxed{\{\langle 1, -1 \rangle, \langle -2, 3 \rangle\}}$, we can see that $[T]_\beta^\beta = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ is diagonal.

Diagonalization, VIII

<u>Example</u>: Determine if $A = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$ is diagonalizable.

- We compute $\det(tI - A) = (t - 1)^3$ since $tI - A$ is upper-triangular, and the eigenvalues are $\lambda = 1, 1, 1$.

- The 1-eigenspace is the nullspace of $I - A = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$,

  which (since the matrix is already in row-echelon form) is

  1-dimensional and spanned by $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

- Since the eigenspace for $\lambda = 1$ is 1-dimensional but the eigenvalue appears 3 times as a root of the characteristic polynomial, the matrix A is not diagonalizable.

Diagonalization, IX

Example: For $A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 2 & 0 \\ 0 & 2 & 1 \end{bmatrix}$, determine whether there exists

a diagonal matrix D and an invertible matrix Q with $D = Q^{-1}AQ$, and if so, find them.

- We compute $\det(tI - A) = (t - 1)^2(t - 2)$, so the eigenvalues are $\lambda = 1, 1, 2$.
- Row-reduction shows $\{(1, 0, 0), (0, 0, 1)\}$ is a basis for the 1-eigenspace and $\{(-1, 1, 2)\}$ is a basis for the 2-eigenspace.
- Since $2 + 1 = 3$, A is diagonalizable. We take Q to be the matrix whose columns are eigenvectors and D to be diagonal with the corresponding eigenvalues:
  $Q = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$ and $D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ has $D = Q^{-1}AQ$.

Diagonalization, X: Are We There Yet?

Having a diagonalization of a matrix allows us to do certain computations very quickly.

- For example, if A is diagonalizable with $D = Q^{-1}AQ$, then it is very easy to compute any power of A.

- Explicitly, since we can rearrange to write $A = QDQ^{-1}$, then $A^k = (QDQ^{-1})^k = Q(D^k)Q^{-1}$, since the conjugate of the kth power is the kth power of a conjugate.

- But since D is diagonal, D^k is simply the diagonal matrix whose diagonal entries are the kth powers of the diagonal entries of D.

Diagonalization, XI: Apparently Not....

<u>Example</u>: If $A = \begin{bmatrix} -2 & -6 \\ 3 & 7 \end{bmatrix}$, find a formula for A^k.

- First, we (try to) diagonalize A. Since $\det(tI - A) = (t-1)(t-4)$, the eigenvalues are 1 and 4. Since these are distinct, A is diagonalizable.

- Computing the eigenvectors shows $D = Q^{-1}AQ$ where $D = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$, $Q = \begin{bmatrix} -2 & -1 \\ 1 & 1 \end{bmatrix}$, $Q^{-1} = \begin{bmatrix} -1 & -1 \\ 1 & 2 \end{bmatrix}$.

- Then $D^k = \begin{bmatrix} 1 & 0 \\ 0 & 4^k \end{bmatrix}$, so $A^k = QD^kQ^{-1}$
  $= Q \begin{bmatrix} 1 & 0 \\ 0 & 4^k \end{bmatrix} Q^{-1} = \begin{bmatrix} 2 - 4^k & 2 - 2 \cdot 4^k \\ -1 + 4^k & -1 + 2 \cdot 4^k \end{bmatrix}$.

Cayley-Hamilton, I: But Wait, There's More!

By diagonalizing a given matrix, we can often prove theorems in a much simpler way.

Definition

If $T : V \to V$ is a linear operator and $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ is a polynomial, we define $p(T) = a_0 I + a_1 T + \cdots + a_n T^n$. If A is an $n \times n$ matrix, we similarly define $p(A) = a_0 I_n + a_1 A + \cdots + a_n A^n$.

Since conjugation preserves sums and products, it is easy to see that $Q^{-1} p(A) Q = p(A^{-1} A Q)$ for any invertible Q.

Cayley-Hamilton, II: We're Almost Done

Using diagonalization, we can establish a useful fact relating a
matrix and its characteristic polynomial.

Theorem (Cayley-Hamilton)

If $p(x)$ is the characteristic polynomial of a matrix A, then $p(A)$ is
the zero matrix.

Example:

- For the matrix $A = \begin{bmatrix} 2 & 2 \\ 3 & 1 \end{bmatrix}$, we have $p(t) = t^2 - 3t - 4$.

- We can compute $A^2 = \begin{bmatrix} 10 & 6 \\ 9 & 7 \end{bmatrix}$, and then indeed we have

$$A^2 - 3A - 4I_2 = \begin{bmatrix} 10 & 6 \\ 9 & 7 \end{bmatrix} - \begin{bmatrix} 6 & 6 \\ 9 & 3 \end{bmatrix} - \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Cayley-Hamilton, III: The Last Actual Slide

Proof (if A is diagonalizable):

- If A is diagonalizable, then let $D = Q^{-1}AQ$ with D diagonal, and let $p(x)$ be the characteristic polynomial of A.

- Then, because raising D to a power just raises all of its diagonal entries to that power, we can see that
$$p(D) = p\left(\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}\right) = \begin{bmatrix} p(\lambda_1) & & \\ & \ddots & \\ & & p(\lambda_n) \end{bmatrix}$$
will just yield the zero matrix, since $p(\lambda_i) = 0$ for each i because the λ_i are the eigenvalues of A (which are roots of p).

- Now by conjugating each term and adding the results, we see that $\mathbf{0} = p(D) = p(Q^{-1}AQ) = Q^{-1}[p(A)]Q$.

- Conjugating back yields $p(A) = Q \cdot \mathbf{0} \cdot Q^{-1} = \mathbf{0}$, as claimed.

Cayley-Hamilton, IV: Just Kidding

In the case where A is not diagonalizable, the proof of the Cayley-Hamilton theorem is substantially more difficult.

Our goal over the next few lectures is to discuss what can be done with non-diagonalizable matrices.

We will show (with suitable assumptions about the eigenvalues of the matrix) that such matrices can still be "nearly" diagonalized by putting them into what is called the Jordan canonical form.

Summary

We reviewed eigenvalues, eigenvectors, and eigenspaces, and methods for computing these things.

We discussed diagonalization of matrices and linear transformations, and showed that a matrix is diagonalizable if and only if all of its eigenspaces have the "maximum possible" dimension.

We proved the Cayley-Hamilton theorem (that a matrix satisfies its characteristic polynomial).

Next lecture: Generalized Eigenvectors.