Module 9: Linear Algebra III

Operational math bootcamp

Statistical Sciences
UNIVERSITY OF TORONTO

Ichiro Hashimoto

University of Toronto

July 24, 2024

Outline

- Adjoints, unitaries and orthogonal matrices
- Orthogonal decomposition
- Spectral theory
  - Eigenvalues and eigenvectors
  - Algebraic and geometric multiplicity of eigenvalues
  - Matrix diagonalization

Recall

Definition

Let V be an $\mathbb{F}$-vector space. A function $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{F}$ is called *inner product* on V if the following holds:

1. (Conjugate) symmetry: $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$ for all $\mathbf{x}, \mathbf{y} \in V$, where $\overline{a}$ denotes the complex conjugate for $a \in \mathbb{C}$
2. Linearity in the first argument: $\langle \alpha \mathbf{x} + \beta \mathbf{y}, \mathbf{z} \rangle = \alpha \langle \mathbf{x}, \mathbf{z} \rangle + \beta \langle \mathbf{y}, \mathbf{z} \rangle$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $\alpha, \beta \in \mathbb{F}$
3. Positive definiteness: $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$

A vector space equipped with an inner product is called an *inner product space*.

Recall

Example

- Standard inner product on $\mathbb{R}^n$: $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
- Standard inner product on $\mathbb{C}^n$: $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i \bar{y}_i$ for $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$
- On the space of polynomials $\mathbb{P}_n(\mathbb{R})$: $\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_{-1}^{1} p(x)q(x)\mathrm{d}x$ for $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{P}_n(\mathbb{R})$

Proposition

Let V be an inner product space. Then

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}\sqrt{\langle \mathbf{y}, \mathbf{y} \rangle}$$

for all $\mathbf{x}, \mathbf{y} \in V$.

Proposition

Let V be an inner product space. Then $\langle \cdot, \cdot \rangle$ induces a norm on V via $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ for all $\mathbf{x} \in V$.

Proof.

Note: With this identification the Cauchy-Schwarz inequality can be restated as:
$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in V$.

Example

The norm introduced by the standard inner product on $\mathbb{R}^n$ is the Euclidean distance.

Adjoint

Definition

Let U, V be inner product spaces and $S\colon U \to V$ be a linear map. The *adjoint* S^* of S is the linear map $S^*: V \to U$ defined such that

$$\langle S\mathbf{u}, \mathbf{v}\rangle_V = \langle \mathbf{u}, S^*\mathbf{v}\rangle_U \qquad \text{for all } \mathbf{u} \in U, \mathbf{v} \in V.$$

Proposition

Let U, V be inner product spaces and $S\colon U \to V$ be a linear map. Then S^* is unique and linear.

Proof.

Example

Define $S\colon \mathbb{R}^3 \to \mathbb{R}^2$ by $S\mathbf{x} = (2x_1 + x_3, -x_2)$. What is the adjoint operator S^*?

Proposition

Let $A \in M_{m\times n}(\mathbb{F})$ be a matrix and $T_A \colon \mathbb{F}^n \to F^m \colon \mathbf{x} \mapsto A\mathbf{x}$. Then, $T_A^*(\mathbf{x}) = A^*\mathbf{x}$, where $A^* \in M_{n\times m}(\mathbb{F})$ with $(A^*)_{ij} = \overline{A_{ji}}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

In particular, if $\mathbb{F} = \mathbb{R}$, the adjoint of the matrix is given by its transpose,denoted A^T, and if $\mathbb{F} = \mathbb{C}$, it is given by its conjugate transpose, denoted A^*.

Proof for $\mathbb{R}$:

Definition

A matrix $O \in M_n(\mathbb{R})$ is called *orthogonal* if its inverse is given by its transpose, i.e. $O^T O = OO^T = I$.

A matrix $U \in M_n(\mathbb{C})$ is called *unitary* if the inverse is given by the conjugate transpose, i.e. $U^* U = UU^* = I$.

Example

- Let $\varphi \in [0, 2\pi]$. Then
$$\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}$$
is an orthogonal matrix. What does it describe geometrically?
- The following is a unitary matrix:
$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

Definition

Let $A \in M_n(\mathbb{F})$. We call A *self-adjoint* if $A^* = A$. In the case $\mathbb{F} = \mathbb{R}$, such an A is called *symmetric* and if $\mathbb{F} = \mathbb{C}$, such an A is called *Hermitian*.

Orthogonality and Gram-Schmidt

Definition

Two vectors $\mathbf{x}, \mathbf{y} \in V$ are called *orthogonal* if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, denoted $\mathbf{x} \perp \mathbf{y}$. We call them *orthonormal* if additionally the vectors are normalized, i.e. $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$. A basis $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of V is called *orthonormal basis (ONB)*, if the vectors are pairwise orthogonal and normalized.

Proposition

Let $\mathbf{x}_1, \ldots, \mathbf{x}_k \in V$ be orthonormal. Then the system of vectors is linearly independent.

Proof.

Proposition (Orthogonal Decomposition)

Let $\mathbf{x}, \mathbf{y} \in V$ with $\mathbf{y} \neq 0$. Then, there exist $c \in F$ and $\mathbf{z} \in V$ such that $\mathbf{x} = c\mathbf{y} + \mathbf{z}$ with $\mathbf{y} \perp \mathbf{z}$.

Given a basis, we can obtain an ONB from it using the Gram-Schmidt algorithm by repeating this orthogonal decomposition.

Proposition (Gram-Schmidt Algorithm)

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in V$ be a system of linearly independent vectors. Define $\mathbf{y}_1 = \mathbf{x}_1 / \|\mathbf{x}_1\|$. For $i = 2, \ldots, n$ define $\mathbf{y}_j$ inductively by

$$\mathbf{y}_i = \frac{\mathbf{x}_i - \sum_{k=1}^{i-1} \langle \mathbf{x}_i, \mathbf{y}_k \rangle \mathbf{y}_k}{\|\mathbf{x}_i - \sum_{k=1}^{i-1} \langle \mathbf{x}_i, \mathbf{y}_k \rangle \mathbf{y}_k\|}.$$

Then the $\mathbf{y}_1, \ldots, \mathbf{y}_n$ are orthonormal and

$$\operatorname{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} = \operatorname{span}\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}.$$

The proof is omitted but can be found in the book.

Recall: connection between matrices and linear maps

Multiplication by a matrix defines a linear map

Let $A \in M_{m\times n}$ be a fixed matrix. Then, we can define a linear map $T_A \colon \mathbb{F}^n \to \mathbb{F}^m$ via $T_A(\mathbf{v}) = A\mathbf{v}$, where we recall matrix vector multiplication $(A\mathbf{v})_i = \sum_{k=1}^{n} A_{ik}v_k$ for $i = 1, \dots, m$.

Given a bases for U and V, $T : U \to V$ can be written as a matrix

Let $T \in \mathcal{L}(U, V)$ where U and V are vector spaces. Let $\mathbf{u}_1, \dots, \mathbf{u}_n$ and $\mathbf{v}_1, \dots, \mathbf{v}_m$ be bases for U and V respectively. The matrix of T with respect to these bases is the $m \times n$ matrix $\mathcal{M}(T)$ with entries A_{ij}, $i = 1, \dots, m$, $j = 1, \dots, n$ defined by

$$T\mathbf{u}_k = A_{1k}\mathbf{v}_1 + \cdots + A_{mk}\mathbf{v}_m.$$

Eigenvalues

Definition

Given an operator $A \colon V \to V$ and $\lambda \in \mathbb{F}$, λ is called an *eigenvalue* of A if there exists a non-zero vector $\mathbf{v} \in V \setminus \{\mathbf{0}\}$ such that

$$A\mathbf{v} = \lambda\mathbf{v}.$$

We call such $\mathbf{v}$ an *eigenvector* of A with eigenvalue λ. We call the set of all eigenvalues of A *spectrum* of A and denote it by $\sigma(A)$.

Motivation in terms of linear maps: Let $T \colon V \to V$ be a linear map, where V is a vector space. We would like to describe the action of this linear map in a particularly "nice" way: such that T acts only by scaling, i.e. $T\mathbf{v}_i = \lambda_i\mathbf{v}_i$ where $\lambda_i \in \mathbb{F}$ for $i = 1, \ldots, n$.

Finding eigenvalues

Note: here we will assume $\mathbb{F} = \mathbb{C}$, so that we are working on an algebraically closed field.

- Rewrite $A\mathbf{v} = \lambda\mathbf{v}$ as
- Thus, if λ is an eigenvalue, we can find the corresponding eigenvectors by finding the null space of $A - \lambda I$.
- The subspace null$(A - \lambda I)$ is called the *eigenspace*
- To find the eigenvalues of A, one must find the scalars λ such that null$(A - \lambda I)$ contains non-trivial vectors (i.e. not $\mathbf{0}$)
- Recall: We saw that $T \in \mathcal{L}(U, V)$ is injective if and only if null $T = \{\mathbf{0}\}$.
- Thus λ is an eigenvalue if and only if $A - \lambda I$ is not invertible.
- Recall: $|A| \neq 0$ if and only if A is invertible.
- Thus λ is an eigenvalue if and only if

Theorem

The following are equivalent

1. $\lambda \in \mathbb{F}$ *is an eigenvalue of* A,
2. $(A - \lambda I)\mathbf{v} = 0$ *has a non-trivial solution,*
3. $|A - \lambda I| = 0$.

Characteristic polynomial

Definition

If A is an $n \times n$ matrix, $p_A(\lambda) = |A - \lambda I|$ is a polynomial of degree n called the *characteristic polynomial* of A.

To find the eigenvectors of A, one needs to find the roots of the characteristic polynomial.

Example

Find the eigenvalues of

$$\begin{bmatrix} 4 & -2 \\ 5 & -3 \end{bmatrix}.$$

Multiplicity

Definition

The multiplicity of the root λ in the characteristic polynomial is called the *algebraic multiplicity* of the eigenvalue λ. The dimension of the eigenspace $\text{null}(A - \lambda I)$ is called the *geometric multiplicity* of the eigenvalue λ.

Definition (Similar matrices)

Square matrices A and B are called *similar* if there exists an invertible matrix S such that

$$A = SBS^{-1}.$$

Similar matrices have the same characteristic polynomials and hence the same eigenvalues (see exercise).

Theorem

Suppose A is a square matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be eigenvectors corresponding to these eigenvalues. Then $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent.

Proof.

Corollary

If a $A \in M_n(\mathbb{C})$ has n distinct eigenvalues, then A is diagonalizable. That is there exists an invertible matrix $S \in M_n(\mathbb{C})$ such that $A = SDS^{-1}$, where D is a diagonal matrix with the eigenvalues of A in the diagonal.

Theorem

Let $A : V \to V$ be an operator with n eigenvalues. A is diagonalizable if and only if for each eigenvalue λ, the geometric multiplicity of λ and the algebraic multiplicity of λ are the same.

Example: a diagonalizable matrix

$$\begin{bmatrix} 1 & 2 \\ 8 & 1 \end{bmatrix} \text{ is diagonalizable.}$$

Example continued

Example continued

Example: a matrix that is not diagonalizable

$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ is *not* diagonalizable.

References

Howard Anton and Chris Rorres. Elementary Linear Algebra. 11th ed. Wiley, 2014

Axler S. *Linear Algebra Done Right*. 3rd ed. Undergraduate Texts in Mathematics. Springer, 2015. Available from: https://link.springer.com/book/10.1007/978-3-319-11080-6

Treil S. *Linear Algebra Done Wrong*. 2017. Available from: https://www.math.brown.edu/streil/papers/LADW/LADW.html