Module 9: Linear Algebra III

Operational math bootcamp

Statistical Sciences
UNIVERSITY OF TORONTO

Emma Kroell

University of Toronto

July 25, 2022

Outline

Spectral theory and matrix decompositions:

- Eigenvalues and eigenvectors
- Algebraic and geometric multiplicity of eigenvalues
- Jordan canonical form
- Singular value decomposition
- LU and QR decompositions

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 $i = 1, \ldots, m$.

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 $\mathbf{u}_1, \ldots, \mathbf{u}_n$ and $\mathbf{v}_1, \ldots, \mathbf{v}_m$ 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}, $i = 1, \ldots, m$, $j = 1, \ldots, n$ defined by

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

Statistical Sciences
UNIVERSITY OF TORONTO

Eigenvalues

Definition

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

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

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

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

- Rewrite $A\mathbf{v} = \lambda\mathbf{v}$ as $(A - \lambda I)v = 0$.
- 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 $T \in \mathcal{L}(U, \mathbf{V})$ is injective if and only if null $T = \{\mathbf{0}\}$.
- Thus λ is an eigenvector 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 eigenvector if and only if $|A - \lambda I| = 0$.

$\det(A - \lambda I) = 0$.

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

$$|A - \lambda I| = 0$$

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

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

$$\begin{aligned} 0 &= |A - \lambda I| \\ &= \begin{vmatrix} 4-\lambda & -2 \\ 5 & -3-\lambda \end{vmatrix} \\ &= (4-\lambda)(-3-\lambda) + 10 \\ &= \lambda^2 - \lambda - 2 \\ &= (\lambda - 2)(\lambda + 1) \end{aligned}$$

$$\therefore \lambda = -1, 2$$

Multiplicity

$$p(\lambda) = (\lambda - 1)^2 (\lambda - 2)(\lambda + 3)^4$$

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 $\lambda_1, \dots, \lambda_n$. Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be eigenvectors corresponding to these eigenvalues. Then $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent.

Proof

By induction on n.

Base case: $n=1$. So there is 1 eigenvalue λ_1 & 1 eigenvector v_1. This is trivial, since any non-zero vector is linearly independent.

Proof continued

Inductive hypothesis: Suppose the claim holds for $k \geq 1$. Then $v_1, \ldots, v_k$ corresponding to $\lambda_1, \ldots, \lambda_k$ (which are distinct) are linearly independent.

Suppose λ_{k+1} is an eigenvalue for A with $\lambda_1, \ldots, \lambda_k \neq \lambda_{k+1}$ and v_{k+1} is corresponding eigenvector.

Let $0 = \sum_{i=1}^{k+1} \alpha_i v_i \quad \forall \alpha_i \in \mathbb{F}$

Apply $(A - \lambda_{k+1} I)$

$$\Rightarrow 0 = \sum_{i=1}^{k+1} \alpha_i (A - \lambda_{k+1} I) v_i$$

Proof continued

$$\Rightarrow 0 = \sum_{i=1}^{k} \alpha_i (Av_i - \lambda_{k+1} v_i) + \alpha_{k+1}(A - \lambda_{k+1} I) v_{k+1}$$

$$= \sum_{i=1}^{k} \alpha_i (\lambda_i - \lambda_{k+1}) v_i + \alpha_{k+1} (\lambda_{k+1} - \lambda_{k+1}) v_{k+1}$$

$$= \sum_{i=1}^{k} \alpha_i (\lambda_i - \lambda_{k+1}) v_i$$

$\therefore\ \alpha_i \underbrace{(\lambda_i - \lambda_{k+1})}_{\neq 0} = 0 \ \forall i$ since $v_1, \ldots, v_k$ are lin ind

$\Rightarrow \alpha_i = 0 \quad \forall i = 1, \ldots, k$

$\Rightarrow 0 = \alpha_{k+1} v_{k+1} \overset{\text{by } (*)}{\Rightarrow} \alpha_{k+1} = 0$

$\therefore v_1, \dots, v_k, v_{k+1}$ are lin. ind.

$0 = \alpha_{k+1} v_{k+1}$

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.

A is nxn matrix

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.

$\text{null}(A - \lambda I)$

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_1 & \\ 0 & & \lambda_2 \end{pmatrix}$$

Example: a diagonalizable matrix

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

Find eigenvalues:

$$0 = |A - \lambda I| = \begin{vmatrix} 1-\lambda & 2 \\ 8 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 16$$

$$= \lambda^2 - 2\lambda - 15$$

$$= (\lambda - 5)(\lambda + 3)$$

$$\therefore \lambda = -3, 5$$

Next: find eigenvectors

Statistical Sciences
UNIVERSITY OF TORONTO

Example continued

$$A + 3I = \begin{pmatrix} 4 & 2 \\ 8 & 4 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$(1, -2)$ spans $\text{null}(A + 3I)$

$$A - 5I = \begin{pmatrix} -4 & 2 \\ 8 & -4 \end{pmatrix} \rightsquigarrow \begin{pmatrix} -2 & 1 \\ 0 & 0 \end{pmatrix}$$

$(1, 2)$ spans $\text{null}(A - 5I)$

Example continued

$$A = \underbrace{\begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix}}_{S} \underbrace{\begin{pmatrix} 5 & 0 \\ 0 & -3 \end{pmatrix}}_{D} \underbrace{\begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix}^{-1}}_{S^{-1}}$$

Statistical Sciences
UNIVERSITY OF TORONTO

Example: a matrix that is not diagonalizable

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

$$0 = |B - \lambda I| = \begin{vmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{vmatrix} = (1-\lambda)^2$$

$\lambda = 1$, multiplicity 2

$A - I = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \Rightarrow \text{null}(A - I)$ is spanned by $(0,1)$

$\therefore \lambda = 1$ has geometric multiplicity of 1

Theorem

Let $A \in M_n(\mathbb{R})$ be a symmetric matrix. Then, there exists an orthogonal matrix $O \in M_n(\mathbb{R})$ such that $A = ODO^T$, where D is a diagonal matrix with the eigenvalues of A in the diagonal. Furthermore, all eigenvalues of A are real.

We can also state this for $M_n(\mathbb{C})$:
Let $A \in M_n(\mathbb{C})$ be a Hermitian matrix. Then, there exists a unitary matrix $U \in M_n(\mathbb{C})$ such that $A = UDU^*$, where D is a diagonal matrix with the eigenvalues of A in the diagonal. Furthermore, all eigenvalues of A are real.

Block matrices

Definition

A block matrix is a matrix that can be broken into sections called blocks, which are smaller matrices.

Example

$$\left[\begin{array}{cc|cc} 2 & 1 & 0 & 1 \\ 0 & 2 & 1 & 1 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 1 \end{array}\right] = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

$$A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad C = 0$$

Definition

A square matrix is called *block diagonal* if it can be written as a block matrix where the main-diagonal blocks are all square matrices and the off-diagonal blocks are all zero.

Example

The matrix

$$\begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 1 \end{bmatrix}$$

is block diagonal.

Definition

A vector $\mathbf{v}$ is called a *generalized eigenvector* of A corresponding to an eigenvalue λ if there exists $k \geq 1$ such that

$$(A - \lambda I)^k \mathbf{v} = 0.$$

The set of generalized eigenvectors of an eigenvalue λ (plus $\mathbf{0}$) is called the *generalized eigenspace* of λ.

Proposition

The algebraic multiplicity of an eigenvalue λ is the same as the dimension of the corresponding generalized eigenspace.

Theorem (Jordan decomposition theorem)

For any operator A there exists a basis such that A is block diagonal with blocks that have eigenvalues on the diagonal and 1s on the upper off-diagonal. In other words, A can be written in the form

n×n

S invertible

$$A = \begin{bmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_k \end{bmatrix}$$

where the blocks J_i on the main diagonal are Jordan block of the form

$$[\lambda], \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}, \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}, \textit{ etc.}$$

This form is called Jordan canonical form.

Connection to algebraic and geometric multiplicity:

- The algebraic multiplicity of an eigenvalue λ is the number of times λ appears on the diagonal.
- The geometric multiplicity of λ is the number of Jordan blocks associated with λ.

Why is Jordan form useful?

- every square matrix has JCF
- JCF $= D + N$ $\quad \exists k \geq 1$ s.t.

D: diagonal $\quad$ N: nilpotent $\rightarrow$ $N^k = 0$

useful in ODEs

Singular value decomposition

- $A^T A$ is symmetric

$$A^T A = ODO^{-1}$$

- therefore it is orthogonally diagonalizable and has real eigenvalues
- In fact, the eigenvalues are non-negative (exercise)

Definition

Let A be an $m \times n$ matrix. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $A^T A$. Then the *singular values* of A are defined as

$$\sigma_1 = \sqrt{\lambda_1}, \ldots, \sigma_n = \sqrt{\lambda_n}.$$

Theorem (Singular value decomposition)

If A is an $m \times n$ matrix of rank k, then we can write

$$A = U\Sigma V^T$$

where Σ is an $m \times n$ matrix of the form

$$\begin{bmatrix} D_{k\times k} & 0_{k\times(n-k)} \\ 0_{(m-k)\times k} & 0_{(m-k)\times(n-k)} \end{bmatrix},$$

D is a diagonal matrix with the singular values of A, $\sigma_1, \ldots, \sigma_n$, on the diagonal and U and V are both orthogonal matrices (of size $m \times n$ and $n \times n$, respectively).

Uses of SVD:

- numerical applications
- U, V are orthogonal so the basis transformation has nice numerical properties

Differences between JCF and SVD:

- JCF has important theoretic applications
- JCF isn't fully diagonal
- SVD has nice numerical properties

LU-decomposition

Definition

The LU-decomposition of a square matrix A is the factorization of A into a lower triangular matrix L and an upper triangular matrix U as follows:

$$A = LU.$$

Why is this useful? Consider the linear system $A\mathbf{x} = \mathbf{b}$

$$LUx = b$$

solve: $Ly = b$ and then $Ux = y$

$Ax = b_1$, $Ax = b_2$, ...

Recall: orthonormal basis

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.

Starting from a set of linearly independent vectors, we can construct another set of vectors which are orthonormal and span the same space using the **Gram-Schmidt Algorithm**.

QR-decomposition

Definition (QR-decomposition)

The QR-decomposition of an $m \times n$ matrix A with linearly independent column vectors is the factorization of A as follows:

$$A = QR,$$

where Q is an $m \times n$ matrix with orthonormal column vectors and R is an $n \times n$ invertible upper triangular matrix.

One obtains the factorization by applying the Gram-Schmidt algorithm to the columns of A. Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be the column vectors of A. Let $\mathbf{q}_1, \ldots, \mathbf{q}_n$ be the orthonormal vectors obtained by applying Gram Schmidt. Then one can write:

$$\begin{aligned}
\mathbf{u}_1 &= \langle \mathbf{u}_1, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_2, \mathbf{q}_2 \rangle \mathbf{q}_2 + \ldots + \langle \mathbf{u}_n, \mathbf{q}_n \rangle \mathbf{q}_n \\
\mathbf{u}_2 &= \langle \mathbf{u}_2, \mathbf{q}_2 \rangle \mathbf{q}_2 + \ldots + \langle \mathbf{u}_n, \mathbf{q}_n \rangle \mathbf{q}_n \\
&\vdots \\
\mathbf{u}_n &= \langle \mathbf{u}_n, \mathbf{q}_n \rangle \mathbf{q}_n
\end{aligned}$$

Thus the orthonormal vectors obtained using Gram-Schmidt form the columns of Q, while R is the terms needed to go between the columns of A and thsoe of Q, i.e.

$$R = \begin{bmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \ldots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \\ 0 & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \ldots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{bmatrix}.$$

Why use QR-decomposition?

$$Ax = b$$

$$QRx = b$$

$$\Rightarrow \underbrace{Qy = b}_{\text{Q doesn't magnify errors}} \quad \& \quad \underbrace{Rx = y}_{\text{nice to solve}}$$

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