MA 110
Linear Algebra and Differential Equations
Lecture 16

Prof. Sudhir R. Ghorpade
Department of Mathematics
IIT Bombay
http://www.math.iitb.ac.in/∼srg/

Spring 2025

Recall that we discussed the following in the last lecture.

- Notion of a unitary matrix. Examples
- $\mathbf{A}$, $\mathbf{B} \in \mathbb{K}^{n\times n}$ are unitarily similar if $\mathbf{B} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}$ for some unitary $\mathbf{U} \in \mathbb{K}^{n\times n}$
- $\mathbf{A} \in \mathbb{K}^{n\times n}$ is unitarily diagonalizable if it is unitarily similar to a diagonal matrix.
- $\mathbf{A} \in \mathbb{K}^{n\times n}$ is unitarily diagonalizable $\iff$ $\mathbb{K}^{n\times 1}$ has an orthonormal basis of eigenvectors of $\mathbf{A}$.
- Eigenvalues and eigenvectors of linear maps.
- Eigenvalues and eigenvectors of a linear map $T : V \to V$ are the same as the eigenvalues and eigenvectors of the matrix $M_E^E(T)$ of T with respect to some ordered basis E of V.
- If V is an n dimensional vector subspace over $\mathbb{K} = \mathbb{C}$ and if $T : V \to V$ is a linear map, then T has an eigenvalue.

Theorem

Let $\mathbb{K} := \mathbb{C}$, and let V be a vector subspace of dimension n. Let $T : V \to V$ be a linear map. Then there is an orthonormal basis E for V such that $\mathbf{M}_E^E(T)$ is upper triangular.

Proof. We use induction on the dimension n of V.

Let $n = 1$. Let $\mathbf{u}$ be a unit vector in V. Then $V = \{\alpha\mathbf{u} : \alpha \in \mathbb{C}\}$. Hence there is $\lambda \in \mathbb{C}$ such that $T(\mathbf{u}) = \lambda\mathbf{u}$. Define $E := (\mathbf{u})$. Then the 1×1 matrix $\mathbf{M}_E^E(T) = [\lambda]$ is clearly upper triangular.

Let $n \geq 2$ and assume that the proposition holds for all vector subspaces of dimension $n - 1$. By the previous lemma, there are a unit vector $\mathbf{u}_1 \in V$ and $\lambda_1 \in \mathbb{C}$ such that $T(\mathbf{u}_1) = \lambda_1\mathbf{u}_1$.

We extend the orthonormal set $\{\mathbf{u}_1\}$ in V to an ordered orthonormal basis $E := (\mathbf{u}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ for V.

Let $W := \text{span}\{\mathbf{x}_2, \ldots, \mathbf{x}_n\}$. Then W is a vector subspace of dimension $n-1$. Let $\mathbf{x} \in W$. Then $T(\mathbf{x}) \in V$ and so we can write $T(\mathbf{x}) = b\,\mathbf{u}_1 + \mathbf{w}$ for unique $b \in \mathbb{C}$ and $\mathbf{w} \in W$.

Define $S(\mathbf{x}) := \mathbf{w}$. It is easy to see that $S : W \to W$ is linear. By the induction hypothesis, there is an ordered orthonormal basis $F := (\mathbf{u}_2, \ldots, \mathbf{u}_n)$ for W such that the $(n-1)\times(n-1)$ matrix $\mathbf{M}_F^F(S)$ is upper triangular.

Let $\mathbf{M}_F^F(S) = [b_{jk}]$, where $b_{jk} \in \mathbb{C}$ for $j, k = 2, \ldots, n$ and $b_{jk} = 0$ if $2 \le k < j \le n$. Then

$$S(\mathbf{u}_k) = \sum_{j=2}^{n} b_{jk}\mathbf{u}_j = \sum_{j=2}^{k} b_{jk}\mathbf{u}_j \quad \text{for } k = 2, \ldots, n.$$

Define $E := (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$. Since $\mathbf{u}_1 \in W^\perp = \{\mathbf{u}_2, \ldots, \mathbf{u}_n\}^\perp$, the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is orthonormal.

Let us find $\mathbf{M}_E^E(T)$. First, $T(\mathbf{u}_1) = \lambda_1 \mathbf{u}_1$.

Next, let $k \in \{2, \ldots, n\}$. Then there are unique $b_{1k} \in \mathbb{C}$ and $\mathbf{w}_k \in W$ such that $T(\mathbf{u}_k) = b_{1k}\mathbf{u}_1 + \mathbf{w}_k$. Hence

$$T(\mathbf{u}_k) = b_{1k}\mathbf{u}_1 + \mathbf{w}_k = b_{1k}\mathbf{u}_1 + S(\mathbf{u}_k) = \sum_{j=1}^{k} b_{jk}\mathbf{u}_j.$$

Define $b_{11} := \lambda_1$ and $b_{j1} := 0$ for $j = 2, \ldots, n$. Then $T(\mathbf{u}_1) = b_{11}\mathbf{u}_1$, and thus

$$T(\mathbf{u}_k) = b_{1k}\mathbf{u}_1 + \cdots + b_{kk}\mathbf{u}_k \quad \text{for } k = 1, \ldots, n.$$

It follows that $M_E^E(T)$ is the $n \times n$ upper triangular matrix $[b_{jk}]$. □

We are now ready to prove that every $\mathbf{A} \in \mathbb{C}^{n \times n}$ is similar to an upper triangular matrix. In fact, we shall prove a stronger result.

Theorem (Schur)

Let $\mathbf{A} \in \mathbb{C}^{n\times n}$. Then $\mathbf{A}$ is unitarily similar to an upper triangular matrix.

Proof. Consider the linear transformation $T_{\mathbf{A}} : \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$ defined by $T_{\mathbf{A}}(\mathbf{x}) := \mathbf{A}\,\mathbf{x}$. By an earlier proposition, there is an ordered orthonormal basis E for $\mathbb{C}^{n\times 1}$ such that $\mathbf{M}_E^E(T_{\mathbf{A}})$ is upper triangular.

Let $\mathbf{B} := \mathbf{M}_E^E(T_{\mathbf{A}})$. Then $\mathbf{B}$ is upper triangular, and by the previous proposition, $\mathbf{A}$ is unitarily similar to $\mathbf{B}$. □

The Schur Theorem is not true for real scalars. For example, $\mathbf{A} := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ is not similar to any upper triangular matrix $\mathbf{B}$; otherwise the diagonal entries of $\mathbf{B}$ would be the eigenvalues of $\mathbf{A}$, but we have seen that $\mathbf{A}$ has no eigenvalues.

Since the proof of the existence of a unitary matrix $\mathbf{U}$ and an upper triangular matrix $\mathbf{B}$ satisfying $\mathbf{B} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}$ is not constructive, Schur's theorem does not help in finding the eigenvalues of $\mathbf{A}$. However, some important information about the eigenvalues of $\mathbf{A}$ can be gathered from the following corollary of Schur's theorem.

Corollary

Let $\mathbf{A} \in \mathbb{C}^{n\times n}$, and let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\mathbf{A}$ counted according to their algebraic multiplicities. Then

(i) $\operatorname{trace} \mathbf{A} = \sum_{j=1}^{n} \lambda_j$,

(ii) $\det \mathbf{A} = \prod_{j=1}^{n} \lambda_j$, and

(iii) $p(\lambda_j) = 0$ for $j = 1, \ldots, n$, whenever $p(t)$ is a polynomial satisfying $p(\mathbf{A}) = \mathbf{O}$.

Proof of Corollary. Consider the characteristic polynomial of $\mathbf{A}$:

$$\begin{aligned}\det(\mathbf{A} - t\mathbf{I}) &= (\lambda_1 - t)\cdots(\lambda_n - t) \\ &= (-1)^n t^n + \left(\sum_{j=1}^{n} \lambda_j\right)(-t)^{n-1} + \cdots + \prod_{j=1}^{n} \lambda_j.\end{aligned}$$

Comparing the coefficient of t^{n-1}, we obtain (i), while comparing the constant coefficient, we obtain (ii).

(iii) Suppose $p(t)$ is a polynomial such that $p(\mathbf{A}) = \mathbf{O}$. Since

$$\mathbf{B}^k = (\mathbf{U}^{-1}\mathbf{A}\mathbf{U})^k = \mathbf{U}^{-1}\mathbf{A}^k\mathbf{U} \quad \text{for all } k \in \mathbb{N},$$

we see that $p(\mathbf{B}) = \mathbf{U}^{-1}p(\mathbf{A})\mathbf{U} = \mathbf{O}$. Now let $j \in \{1, \ldots, n\}$. Then λ_j is the (j,j)th entry of $\mathbf{B}$, and since $\mathbf{B}$ is upper triangular, λ_j^k is the (j,j)th entry of $\mathbf{B}^k$ for all $k \in \mathbb{N}$. It follows that $p(\lambda_j)$ is the (j,j)th entry of $p(\mathbf{B})$. It must be equal to 0 because $p(\mathbf{B}) = \mathbf{O}$. □

Examples

(1) Let $\mathbf{A} \in \mathbb{C}^{3\times 3}$ be such that $\mathbf{A}^2 = 6\mathbf{A}$ and $\operatorname{trace} \mathbf{A} = 12$. Let us determine the eigenvalues of $\mathbf{A}$.

Consider the polynomial $p(t) = t^2 - 6t$. Since $p(\mathbf{A}) = \mathbf{O}$, we see that $p(\lambda) = \lambda^2 - 6\lambda = \lambda(\lambda - 6) = 0$ for every eigenvalue λ of $\mathbf{A}$. Since $\operatorname{trace} \mathbf{A} = 12$, the sum of the eigenvalues of $\mathbf{A}$ (counting algebraic multiplicities) is equal to 12. Hence the eigenvalues of $\mathbf{A}$ are $6, 6, 0$, that is, 6 is an eigenvalue of $\mathbf{A}$ of algebraic multiplicity 2, and 0 is an eigenvalue of $\mathbf{A}$ of algebraic multiplicity 1.

(ii) Let $\mathbf{A} := \begin{bmatrix} \mathbf{e}_2 & \mathbf{e}_3 & \cdots & \mathbf{e}_n & \mathbf{e}_1 \end{bmatrix}$, where $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are the basic column vectors in $\mathbb{C}^{n\times 1}$.
Then $\mathbf{A}\mathbf{e}_1 = \mathbf{e}_2, \ldots, \mathbf{A}\mathbf{e}_{n-1} = \mathbf{e}_n$ and $\mathbf{A}\mathbf{e}_n = \mathbf{e}_1$. Hence
$\mathbf{A} \begin{bmatrix} x_1 & x_2 & \cdots & x_{n-1} & x_n \end{bmatrix}^{\mathsf{T}} = \begin{bmatrix} x_n & x_1 & x_2 & \cdots & x_{n-1} \end{bmatrix}^{\mathsf{T}}$
for $\begin{bmatrix} x_1 & x_2 & \cdots & x_{n-1} & x_n \end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^{n\times 1}$.

The matrix $\mathbf{A}$ represents a **cyclic shift to the right**. It follows that $\mathbf{A}^n = \mathbf{I}$. Let λ be an eigenvalue of $\mathbf{A}$. Then $\lambda^n = 1$.

$$\text{Note that } \mathbf{A} := \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ldots & \ldots & \ldots & \ldots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Let $\omega := e^{2\pi i/n}$. Then the nth roots of 1 are $1, \omega, \omega^2, \ldots, \omega^{n-1}$. Thus $\lambda \in \{1, \omega, \omega^2, \ldots, \omega^{n-1}\}$.

Conversely, we show that each $1, \omega, \omega^2, \ldots, \omega^{n-1}$ is an eigenvalue of $\mathbf{A}$ by finding a corresponding eigenvector.

Let $\mathbf{x} := \begin{bmatrix} x_1 & x_2 & \cdots & x_{n-1} & x_n \end{bmatrix}^\mathsf{T} \in \mathbb{C}^{n\times 1}$. Then $\mathbf{Ax} = \lambda\,\mathbf{x}$ means $x_1 = \lambda x_2, x_2 = \lambda x_3, \ldots, x_{n-1} = \lambda x_n$ and $x_n = \lambda x_1$.

For $j = 0, \ldots, n-1$, define

$$\mathbf{x}_j := \begin{bmatrix} 1 & 1/\omega^j & 1/\omega^{2j} & \cdots & 1/\omega^{(n-1)j} \end{bmatrix}^\mathsf{T}.$$

Now $1/\omega^{(n-1)j} = \omega^j$ since $\omega^n = 1$, and so

$$\mathbf{x}_j := \begin{bmatrix} 1 & 1/\omega^j & 1/\omega^{2j} & \cdots & 1/\omega^{(n-2)j} & \omega^j \end{bmatrix}^\mathsf{T}.$$

Hence $\mathbf{Ax}_j = \begin{bmatrix} \omega^j & 1 & 1/\omega^j & 1/\omega^{2j} & \cdots & 1/\omega^{(n-2)j} \end{bmatrix}^\mathsf{T} = \omega^j \mathbf{x}_j$. Thus $\mathbf{x}_j$ is an eigenvector of $\mathbf{A}$ corresponding to the eigenvalue ω^j for $j = 0, 1, \ldots, n-1$.

We note that since the columns of $\mathbf{A}$ form an orthonormal set in $\mathbb{C}^{n\times 1}$, the matrix $\mathbf{A}$ is unitary.

Unitarily Diagonalizable Matrix

We have seen that if the scalars are complex numbers, then every matrix can be unitarily 'upper triangularized'. Now we take up the question: 'Which matrices can be unitarily diagonalized?'. We saw that a matrix $\mathbf{A} \in \mathbb{K}^{n\times n}$ is unitarily diagonalizable if and only if there is an orthonormal basis for $\mathbb{K}^{n\times 1}$ consisting of eigenvectors of $\mathbf{A}$. Let us investigate this condition further.

Suppose $\mathbf{A} \in \mathbb{K}^{n\times n}$ is unitarily diagonalizable, and let eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ of $\mathbf{A}$ form an orthonormal basis for $\mathbb{K}^{n\times 1}$. Let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues of $\mathbf{A}$, so that $\mathbf{A}\mathbf{u}_j = \lambda_j \mathbf{u}_j$ for $j = 1, \ldots, n$. Let $\mathbf{x} \in \mathbb{K}^{n\times 1}$. Then

$$\mathbf{x} = \sum_{j=1}^{n} \langle \mathbf{u}_j, \mathbf{x} \rangle \mathbf{u}_j \quad \text{and} \quad \mathbf{A}\mathbf{x} = \sum_{j=1}^{n} \langle \mathbf{u}_j, \mathbf{x} \rangle \mathbf{A}\mathbf{u}_j = \sum_{j=1}^{n} \lambda_j \langle \mathbf{u}_j, \mathbf{x} \rangle \mathbf{u}_j.$$

The above representation of $\mathbf{A}\,\mathbf{x}$ can be used for various purposes. For this reason, we would like to find necessary and/or sufficient conditions under which a square matrix can be unitarily diagonalized. We introduce a new class of matrices.

Definition

A matrix $\mathbf{A} \in \mathbb{K}^{n\times n}$ is called **normal** *if it commutes with its adjoint, that is, $\mathbf{A}^*\mathbf{A} = \mathbf{A}\,\mathbf{A}^*$.*

Examples (i) If $\mathbf{A}$ is self-adjoint (that is, $\mathbf{A}^* = \mathbf{A}$), or skew self-adjoint (that is, $\mathbf{A}^* = -\mathbf{A}$), or unitary (that is, $\mathbf{A}^*\mathbf{A} = \mathbf{I} = \mathbf{A}\,\mathbf{A}^*$), then $\mathbf{A}$ is normal.

(ii) The matrix $\mathbf{A} := \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \in \mathbb{K}^{2\times 2}$ is normal, but it is not self-adjoint, or skew self-adjoint, or unitary. However, not every matrix is normal, as the example $\mathbf{A} := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ shows.

Proposition

Let $\mathbf{A} \in \mathbb{K}^{n\times n}$. Then $\mathbf{A}$ is normal if and only if

$$\langle \mathbf{Ax},\ \mathbf{Ay}\rangle = \langle \mathbf{A}^*\mathbf{x},\ \mathbf{A}^*\mathbf{y}\rangle \quad \textit{for all } \mathbf{x}, \mathbf{y} \in \mathbb{K}^{n\times 1}.$$

Proof. For $\mathbf{x}, \mathbf{y}$, $\langle \mathbf{Ax},\ \mathbf{Ay}\rangle = (\mathbf{Ax})^*\mathbf{Ay} = \mathbf{x}^*\mathbf{A}^*\mathbf{Ay}$ and $\langle \mathbf{A}^*\mathbf{x},\ \mathbf{A}^*\mathbf{y}\rangle = (\mathbf{A}^*\mathbf{x})^*\mathbf{A}^*\mathbf{y} = \mathbf{x}^*\mathbf{A}\,\mathbf{A}^*\mathbf{y}$.

If $\mathbf{A}$ is normal, then clearly, $\langle \mathbf{Ax},\ \mathbf{Ay}\rangle = \langle \mathbf{A}^*\mathbf{x},\ \mathbf{A}^*\mathbf{y}\rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{K}^{n\times 1}$. Conversely, suppose $\langle \mathbf{Ax},\ \mathbf{Ay}\rangle = \langle \mathbf{A}^*\mathbf{x},\ \mathbf{A}^*\mathbf{y}\rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{K}^{n\times 1}$. Then letting $\mathbf{x} := \mathbf{e}_j$ and $\mathbf{y} := \mathbf{e}_k$, we see that $\mathbf{e}_j^*\mathbf{A}^*\mathbf{A}\,\mathbf{e}_k = \mathbf{e}_j^*\mathbf{A}\,\mathbf{A}^*\mathbf{e}_k$, that is, the (j, k)th entries of $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\,\mathbf{A}^*$ are the same for all $j, k = 1, \ldots, n$. Hence $\mathbf{A}^*\mathbf{A} = \mathbf{A}\,\mathbf{A}^*$, that is, $\mathbf{A}$ is normal. □

The above condition for normality of a matrix $\mathbf{A}$ says that the lengths of the vectors $\mathbf{Ax}$ and $\mathbf{A}^*\mathbf{x}$ should be the same, and the angle between $\mathbf{Ax}$ and $\mathbf{Ay}$ should be the same as the angle between $\mathbf{A}^*\mathbf{x}$ and $\mathbf{A}^*\mathbf{y}$ for all $\mathbf{x}$, $\mathbf{y} \in \mathbb{K}^{n\times 1}$.