MA 110
Linear Algebra and Differential Equations
Lecture 13

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

Spring 2025

Recall that we discussed the following notions and results.

- Matrix of a linear transformation of vector (sub)spaces
- Eigenvalues, eigenvectors, and eigenspaces of a square matrix with entries in $\mathbb{K}$ (where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$)
- Characteristic polynomial of a square matrix
- Algebraic multiplicity and geometric multiplicity
- Similarity of square matrices. Diagonalizability
- Similarity and change of basis
- $\mathbf{A} \in \mathbb{K}^{n\times n}$ diagonalizable $\iff$ $\mathbb{K}^{n\times 1}$ has a basis of eigenvectors of $\mathbf{A}$
- Linear independence of eigenvectors corresponding to distinct eigenvalues
- $\mathbf{A} \in \mathbb{K}^{n\times n}$ diagonalizable $\iff$ sum of geometric multiplicities of distinct eigenvalues of $\mathbf{A}$ is equal to n
- $\mathbf{A} \in \mathbb{K}^{n\times n}$ diagonalizable $\implies$ alg. mult. $=$ geom. mult. for every eigenvalue of $\mathbf{A}$. Converse is true if $\mathbb{K} = \mathbb{C}$.

Relating geometric and algebraic multiplicities

The inequality shown below may have been used in the proof of a result on diagonalization given earlier.

Proposition

Let $\mathbf{A} \in \mathbb{K}^{n \times n}$ and let λ be an eigenvalue of $\mathbf{A}$. Then the geometric multiplicity of λ is less than or equal to its algebraic multiplicity.

Proof. Let g be the geometric multiplicity of λ. Let $(\mathbf{v}_1, \ldots, \mathbf{v}_g)$ be an ordered basis of the eigenspace of λ; extend it to an ordered basis $(\mathbf{v}_1, \ldots, \mathbf{v}_g, \mathbf{v}_{g+1}, \ldots, \mathbf{v}_n)$ of $\mathbb{K}^{n \times 1}$. Define $\mathbf{P} := \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$. Then $\mathbf{P}$ is invertible since its n columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent. Consider $\mathbf{A}' := \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$. Since $\mathbf{A}\mathbf{v}_j = \lambda \mathbf{v}_j$ and $\mathbf{P}\mathbf{e}_j = \mathbf{v}_j$ for $j = 1, \ldots, g$, we see that the j th column of $\mathbf{A}'$ is given by

$$\mathbf{A}'\mathbf{e}_j = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}\mathbf{e}_j = \mathbf{P}^{-1}\mathbf{A}\mathbf{v}_j = \lambda \mathbf{P}^{-1}\mathbf{v}_j = \lambda \mathbf{e}_j.$$

Hence

$$\mathbf{A}' = \left[\begin{array}{ccc|c} \lambda & \cdots & 0 & \\ \vdots & \ddots & \vdots & \mathbf{C} \\ 0 & \cdots & \lambda & \\ \hline & \mathbf{O} & & \mathbf{D} \end{array}\right]$$

where $\mathbf{C} \in \mathbb{K}^{g\times(n-g)}$, $\mathbf{O} \in \mathbb{K}^{(n-g)\times g}$ and $\mathbf{D} \in \mathbb{K}^{(n-g)\times(n-g)}$. Expanding by the first column, we see that

$$\det(\mathbf{A}' - t\mathbf{I}) = (\lambda - t)^g q(t),$$

where $q(t)$ is a polynomial of degree $n - g$. Thus

$$p_{\mathbf{A}}(t) = p_{\mathbf{A}'}(t) = \det(\mathbf{A}' - t\mathbf{I}) = (\lambda - t)^g q(t).$$

Thus $(\lambda - t)^g$ divides the characteristic polynomial $p_{\mathbf{A}}(t)$ of $\mathbf{A}$. Since the algebraic multiplicity of λ is the largest integer m such that $(\lambda - t)^m$ divides $p_{\mathbf{A}}(t)$, we obtain $g \leq m$. $\square$

We now recall the result on diagonalization given earlier.

Proposition

Let $\mathbf{A} \in \mathbb{K}^{n\times n}$, and let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of $\mathbf{A}$. Let g_j and m_j be the geometric multiplicity and the algebraic multiplicity of λ_j respectively for $j = 1, \ldots, k$.

(i) $\mathbf{A}$ diagonalizable $\Longrightarrow g_j = m_j$ for $j = 1, \ldots, k$.

(ii) $\mathbb{K} = \mathbb{C}$ and $g_j = m_j$ for $j = 1, \ldots, k \Longrightarrow \mathbf{A}$ diagonalizable.

Remark: Let $\mathbf{A}$ be a square matrix and λ an eigenvalue of $\mathbf{A}$. If the geometric multiplicity g of λ is less than the algebraic multiplicity m of λ, then the eigenvalue λ is called **defective**, and $m - g$ is called its **defect**. If a matrix does not have any defective eigenvalue, then the matrix is called **nondefective**.

The above proposition tells us that when $\mathbb{K} = \mathbb{C}$, a square matrix $\mathbf{A}$ is diagonalizable if and only if it is nondefective. We shall later show that if $\mathbb{K} = \mathbb{C}$, then every square matrix can be 'upper triangularized', that is, it is similar to an upper triangular matrix. In fact, we will prove a stronger result.

Existence and Location of Eigenvalues

Let $\mathbf{A} \in \mathbb{K}^{n \times n}$ and $\lambda \in \mathbb{K}$. Since λ is an eigenvalue of $\mathbf{A}$ if and only if λ is root of the characteristic polynomial of $\mathbf{A}$, and since this polynomial is of degree n, the matrix $\mathbf{A}$ can have at most n distinct eigenvalues. Let $k \in \mathbb{N}$ with $1 \leq k \leq n$. It is easy to see that the matrix $\mathbf{A} := \mathrm{diag}(1, 2, \ldots, k, k, \ldots, k)$ has exactly k distinct eigenvalues.

If $\mathbb{K} = \mathbb{R}$, then $\mathbf{A}$ may not have any eigenvalue if n is even, and $\mathbf{A}$ has at least one eigenvalue if n is odd. On the other hand, if $\mathbb{K} = \mathbb{C}$, then $\mathbf{A}$ has exactly n eigenvalues, if we count them according to their algebraic multiplicities.

Often, it is not enough to know that so many eigenvalues of $\mathbf{A}$ exist; one would like to know where they are located. In this connection, we give a 'localization' result.

Gerschgorin disks and Gerschgorin Theorem

Let $\mathbf{A} := [a_{jk}] \in \mathbb{K}^{n\times n}$. For $j \in \{1, \ldots, n\}$, define $r_j := \sum_{k\neq j} |a_{jk}|$, and let $D(a_{jj}, r_j) := \{a \in \mathbb{K} : |a - a_{jj}| \leq r_j\}$, which is a closed disk in $\mathbb{K}$ with centre at the jth diagonal entry of $\mathbf{A}$ and radius equal to the sum of the absolute values of the off-diagonal entries in the jth row of $\mathbf{A}$; it is called the jth **Gerschgorin disk** of the matrix $\mathbf{A}$.

Proposition (Gerschgorin Theorem)

Let $\mathbf{A} \in \mathbb{K}^{n\times n}$. Every eigenvalue of $\mathbf{A}$ belongs in one of the Gerschgorin disks of $\mathbf{A}$.

Proof. Let $\mathbf{A} := [a_{jk}]$. Let $\lambda \in \mathbb{K}$ be an eigenvalue of $\mathbf{A}$, and let $\mathbf{x} := \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^{\mathsf{T}}$ be a corresponding eigenvector. Let $j \in \{1, \ldots, n\}$ be such that $|x_j| = \max\{|x_k| : k = 1, \ldots, n\}$. Then $x_j \neq 0$ since $\mathbf{x} \neq \mathbf{0}$. Multiplying $\mathbf{x}$ by $1/x_j$, we may assume that $x_j = 1$ and $|x_k| \leq 1$ for all $k \neq j$.

Comparing the jth components in the vector equation $\mathbf{Ax} = \lambda\mathbf{x}$, we obtain

$$\sum_{k \neq j} a_{jk}x_k + a_{jj}x_j = \lambda x_j, \text{ that is, } \sum_{k \neq j} a_{jk}x_k + a_{jj} = \lambda.$$

Now the triangle inequality for elements of $\mathbb{K}$ shows that

$$|\lambda - a_{jj}| = \left| \sum_{k \neq j} a_{jk}x_k \right| \leq \sum_{k \neq j} |a_{jk}x_k| \leq \sum_{k \neq j} |a_{jk}| = r_j.$$

Thus λ belongs to the jth Gerschgorin disk of $\mathbf{A}$. $\square$

Example Let $\mathbf{A} := \begin{bmatrix} 10 & -1 & 0 & 1 \\ 0.2 & 8 & 0.3 & 0.1 \\ 1 & -1 & 2 & 1 \\ 1 & 0.5 & -1 & 11 \end{bmatrix}$. The centres of the Gerschgorin disks are 10, 8, 2, 11 with the respective radii 2, 0.6, 3, 2.5. Hence if λ is an eigenvalue of $\mathbf{A}$, then either $|\lambda - 10| \leq 2$ or $|\lambda - 8| \leq 0.6$ or $|\lambda - 2| \leq 3$ or $|\lambda - 11| \leq 2.5$.

Inner Product and Norm

Let $\mathbb{K} := \mathbb{R}$, the set of real numbers, or $\mathbb{K} := \mathbb{C}$, the set of complex numbers. For a scalar $\alpha \in \mathbb{K}$, we denote its conjugate by $\overline{\alpha}$. If $\alpha \in \mathbb{R}$, then of course, $\overline{\alpha} = \alpha$.

Consider column vectors $\mathbf{x} := \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} := \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ in $\mathbb{K}^{n\times 1}$.

The conjugate transpose (or the adjoint) $\mathbf{x}^* := \begin{bmatrix} \overline{x}_1 & \cdots & \overline{x}_n \end{bmatrix}$ of $\mathbf{x}$ is a row vector in $\mathbb{K}^{1\times n}$. The **inner product** of $\mathbf{x}$ with $\mathbf{y}$ is defined by

$$\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^* \mathbf{y} = \overline{x}_1 y_1 + \cdots + \overline{x}_n y_n.$$

Note: If $\mathbb{K} = \mathbb{R}$, then $\langle \mathbf{x}, \mathbf{y} \rangle$ is just the scalar product of $\boldsymbol{x} := (x_1, \ldots, x_n)$ and $\boldsymbol{y} := (y_1, \ldots, y_n)$ in $\mathbb{R}^n$.

The inner product function $\langle \cdot\, , \cdot \rangle : \mathbb{K}^{n\times 1} \times \mathbb{K}^{n\times 1} \to \mathbb{K}$ has the following crucial properties. For $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{K}^{n\times 1}$ and $\alpha, \beta \in \mathbb{K}$,

1. $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \iff \mathbf{x} = \mathbf{0}$ (positive definite),
2. $\langle \mathbf{x}, \alpha\mathbf{y} + \beta\mathbf{z} \rangle = \alpha\langle \mathbf{x}, \mathbf{y} \rangle + \beta\langle \mathbf{x}, \mathbf{z} \rangle$ (linear in 2nd variable),
3. $\langle \mathbf{y}, \mathbf{x} \rangle = \overline{\langle \mathbf{x}, \mathbf{y} \rangle}$ (conjugate symmetric).

From the above three crucial properties, conjugate linearity in the 1st variable follows: $\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \overline{\alpha}\langle \mathbf{x}, \mathbf{z} \rangle + \overline{\beta}\langle \mathbf{y}, \mathbf{z} \rangle$.

Let $\mathbf{x} := \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^{\mathsf{T}} \in \mathbb{K}^{n\times 1}$. We define the **norm** of $\mathbf{x}$ by

$$\|\mathbf{x}\| := \langle \mathbf{x}, \mathbf{x} \rangle^{1/2} = (|x_1|^2 + \cdots + |x_n|^2)^{1/2}.$$

For $n = 1$, the norm of $x \in \mathbb{K}$ is the absolute value $|x|$ of x.

Clearly, $\max\{|x_1|, \ldots, |x_n|\} \leq \|\mathbf{x}\| \leq |x_1| + \cdots + |x_n|$.

If $\mathbf{x} \in \mathbb{K}^{n\times 1}$ and $\|\mathbf{x}\| = 1$, then we say that $\mathbf{x}$ is a **unit vector**.

Theorem

Let $\mathbf{x}, \mathbf{y} \in \mathbb{K}^{n\times 1}$. Then
(i) (Schwarz Inequality) $|\langle \mathbf{x}, \mathbf{y}\rangle| \le \|\mathbf{x}\|\|\mathbf{y}\|$.

(ii) (Triangle Inequality) $\|\mathbf{x}+\mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$.

Proof. Suppose $\mathbf{x} := \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^{\mathsf{T}}$ and $\mathbf{y} := \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^{\mathsf{T}}$.

(i) If $\|\mathbf{x}\| = 0$ or $\|\mathbf{y}\| = 0$, then $\mathbf{x} = \mathbf{0}$ or $\mathbf{y} = \mathbf{0}$. Hence we are done. Now let $\|\mathbf{x}\| \neq 0$ and $\|\mathbf{y}\| \neq 0$. Then

$$\frac{|\overline{x}_j|}{\|\mathbf{x}\|}\frac{|y_j|}{\|\mathbf{y}\|} \le \frac{1}{2}\Big(\frac{|x_j|^2}{\|\mathbf{x}\|^2} + \frac{|y_j|^2}{\|\mathbf{y}\|^2}\Big) \quad \text{for } j = 1, \ldots, n,$$

since $|\overline{\alpha}\beta| = |\alpha|\,|\beta| \le (|\alpha|^2 + |\beta|^2)/2$ for all $\alpha, \beta \in \mathbb{K}$. Hence

$$|\langle \mathbf{x}, \mathbf{y}\rangle| \le \sum_{j=1}^{n} |\overline{x}_j|\,|y_j| \le \frac{\|\mathbf{x}\|\|\mathbf{y}\|}{2}(1+1) = \|\mathbf{x}\|\|\mathbf{y}\|.$$

(ii) Since $\langle \mathbf{x},\, \mathbf{y} \rangle + \langle \mathbf{y},\, \mathbf{x} \rangle = 2\,\Re\,\langle \mathbf{x},\, \mathbf{y} \rangle$, we see that

$$\begin{aligned} \|\mathbf{x}+\mathbf{y}\|^2 &= \langle \mathbf{x}+\mathbf{y},\, \mathbf{x}+\mathbf{y} \rangle = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\,\Re\,\langle \mathbf{x},\, \mathbf{y} \rangle \\ &\le \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\,|\langle \mathbf{x},\, \mathbf{y} \rangle| \\ &\le \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\,\|\mathbf{x}\|\|\mathbf{y}\| \quad \text{(by the Schwarz inequality)} \\ &= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2. \end{aligned}$$

Thus $\|\boldsymbol{x}+\boldsymbol{y}\| \le \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$. $\square$

We observe that the norm function $\|\cdot\| : \mathbb{K}^{n\times 1} \to \mathbb{K}$ satisfies the following three crucial properties:

(i) $\|\mathbf{x}\| \ge 0$ for all $\mathbf{x} \in \mathbb{K}^{n\times 1}$ and $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$,

(ii) $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for all $\alpha \in \mathbb{K}$ and $\mathbf{x} \in \mathbb{K}^{n\times 1}$,

(iii) $\|\mathbf{x}+\mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{K}^{n\times 1}$.

The properties of the norm function allow us to define the distance between two vectors in $\mathbb{K}^{n\times 1}$. Let $\mathbf{x}, \mathbf{y} \in \mathbb{K}^{n\times 1}$. Then the **distance** between $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|.$$

The distance function $d : \mathbb{K}^{n\times 1} \times \mathbb{K}^{n\times 1} \to \mathbb{K}$ has the following analogous properties.

(i) $d(\mathbf{x}, \mathbf{y}) \geq 0$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{K}^{n\times 1}$, $d(\mathbf{x}, \mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}$,

(ii) $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{K}^{n\times 1}$,

(iii) $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{K}^{n\times 1}$.

The inner product defined earlier allows us to say when two column vectors are perpendicular to each other.