MA 110:
Lecture 12

Saurav Bhaumik
Department of Mathematics
IIT Bombay

Spring 2025

Diagonalizability Revisited

Recall the definition.

Definition

A matrix $\mathbf{A} \in \mathbb{K}^{n \times n}$ *is called* **diagonalizable** *(over* $\mathbb{K}$*) if* $\mathbf{A}$ *is similar to a diagonal matrix (over* $\mathbb{K}$*).*

We stated the following characterization of diagonalizability.

Proposition

A matrix $\mathbf{A} \in \mathbb{K}^{n \times n}$ is diagonalizable if and only if there is a basis for $\mathbb{K}^{n \times 1}$ consisting of eigenvectors of $\mathbf{A}$. In fact,

$$\begin{aligned} & \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}, \text{ where } \mathbf{P}, \mathbf{D} \in \mathbb{K}^{n \times n} \text{ are of the form} \\ & \mathbf{P} = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix} \text{ and } \mathbf{D} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \\ \iff \quad & \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \text{ is a basis for } \mathbb{K}^{n \times 1} \text{ and} \\ & \mathbf{A}\mathbf{x}_k = \lambda_k \mathbf{x}_k \text{ for } k = 1, \ldots, n. \end{aligned}$$

Proof of a characterization of diagonalizability

Proof. The result is a consequence of the earlier characterization of similarity. It can also be seen as follows.

$$
\begin{aligned}
& \mathbf{A} \text{ is diagonalizable} \\
\iff\ & \exists \text{ invertible matrix } \mathbf{P} \text{ and diagonal matrix } \mathbf{D} \text{ in } \mathbb{K}^{n\times n} \\
& \text{such that } \mathbf{AP} = \mathbf{PD} \\
\iff\ & \exists \text{ a basis } \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \text{ for } \mathbb{K}^{n\times 1} \text{ and } \lambda_1, \dots, \lambda_n \in \mathbb{K} \text{ such} \\
& \text{that } \mathbf{A}\begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix} \operatorname{diag}(\lambda_1, \dots, \lambda_n) \\
\iff\ & \exists \text{ a basis } \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \text{ of } \mathbb{K}^{n\times 1} \text{ and } \lambda_1, \dots, \lambda_n \in \mathbb{K} \text{ such} \\
& \text{that } \mathbf{A}\mathbf{x}_k = \lambda_k \mathbf{x}_k \text{ for } k = 1, \dots, n. \qquad \square
\end{aligned}
$$

Application: If a matrix $\mathbf{A}$ is diagonalizable and we find invertible $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, then any power of $\mathbf{A}$ can be found easily. This is seen as follows:

$$\mathbf{A}^m = (\mathbf{PDP}^{-1}) \cdots (\mathbf{PDP}^{-1}) = \mathbf{PD}^m\mathbf{P}^{-1} = \mathbf{P} \operatorname{diag}(\lambda_1^m, \dots, \lambda_n^m)\mathbf{P}^{-1}.$$

Example: Consider $\mathbf{A} := \begin{bmatrix} 4 & -3 \\ 2 & -1 \end{bmatrix}$. Then

$$p_{\mathbf{A}}(t) = \det \begin{bmatrix} t-4 & 3 \\ -2 & t+1 \end{bmatrix} = (t-4)(t+1)+6 = (t-2)(t-1).$$

Thus 2 and 1 are the eigenvalues of $\mathbf{A}$ and it is easy to see that $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ are corresponding eigenvectors. So $\mathbf{P} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}$ satisfies $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathrm{diag}(2,1)$, which can be written as

$$\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix}.$$

Hence

$$\begin{aligned} \mathbf{A}^m &= \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2^m & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 2^m 3 & 1 \\ 2^m 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 2^m 3 - 2 & -2^m 3 + 3 \\ 2^m 2 - 2 & -2^m 2 + 3 \end{bmatrix} \quad \text{for } m \in \mathbb{N}. \end{aligned}$$

Eigenvectors corresponding to distinct eigenvalues

Our next result is about the linear independence of eigenvectors corresponding to distinct eigenvalues of a matrix.

Lemma

Let $\mathbf{A} \in \mathbb{K}^{n\times n}$, and let $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues of $\mathbf{A}$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_k$ belong to the eigenspaces of $\mathbf{A}$ corresponding to $\lambda_1, \ldots, \lambda_k$ respectively. Then

$$\mathbf{x}_1 + \cdots + \mathbf{x}_k = \mathbf{0} \iff \mathbf{x}_1 = \cdots = \mathbf{x}_k = \mathbf{0}.$$

In particular, if $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are eigenvectors of $\mathbf{A}$ corresponding to $\lambda_1, \ldots, \lambda_k$ respectively, then the set $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ is linearly independent.

Proof. We use induction on the number k of distinct eigenvalues of $\mathbf{A}$. Clearly, the result holds for $k = 1$.

Let $k \geq 2$ and assume that the result holds for $k - 1$.

Suppose $\mathbf{x} := \mathbf{x}_1 + \cdots + \mathbf{x}_{k-1} + \mathbf{x}_k = \mathbf{0}$. Then $\mathbf{A}\mathbf{x} = \mathbf{0}$, that is, $\lambda_1\mathbf{x}_1 + \cdots + \lambda_{k-1}\mathbf{x}_{k-1} + \lambda_k\mathbf{x}_k = \mathbf{0}$. Also, multiplying the first equation by λ_k, we obtain $\lambda_k\mathbf{x}_1 + \cdots + \lambda_k\mathbf{x}_{k-1} + \lambda_k\mathbf{x}_k = \mathbf{0}$. Subtraction gives $(\lambda_1 - \lambda_k)\mathbf{x}_1 + \cdots + (\lambda_{k-1} - \lambda_k)\mathbf{x}_{k-1} = \mathbf{0}$.

By the induction hypothesis, $(\lambda_1 - \lambda_k)\mathbf{x}_1 = \cdots = (\lambda_{k-1} - \lambda_k)\mathbf{x}_{k-1} = \mathbf{0}$. Since $\lambda_1 \neq \lambda_k, \ldots, \lambda_{k-1} \neq \lambda_k$, we obtain $\mathbf{x}_1 = \cdots = \mathbf{x}_{k-1} = \mathbf{0}$, and so $\mathbf{x}_k = \mathbf{0}$ as well.

Now let $\mathbf{x}_1, \ldots, \mathbf{x}_k$ be eigenvectors. If $\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k = \mathbf{0}$, then $\alpha_1\mathbf{x}_1 = \cdots = \alpha_k\mathbf{x}_k = \mathbf{0}$. But $\mathbf{x}_1 \neq \mathbf{0}, \ldots, \mathbf{x}_k \neq \mathbf{0}$, so that $\alpha_1 = \cdots = \alpha_k = 0$. Thus $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ is linearly independent.

Theorem

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 be the geometric multiplicity of λ_j for $j = 1, \ldots, k$. Then $g_1 + \cdots + g_k \leq n$. Further, $\mathbf{A}$ is diagonalizable if and only if $g_1 + \cdots + g_k = n$.

Proof. Let V_j denote the eigenspace $\mathcal{N}(\mathbf{A} - \lambda_j \mathbf{I})$ of $\mathbb{K}^{n\times 1}$, and let E_j be a basis for V_j consisting of g_j eigenvectors of $\mathbf{A}$ corresponding to λ_j for $j = 1, \ldots, k$.

We claim that the set $E := E_1 \cup \cdots \cup E_k$ is linearly independent. Let $\mathbf{x}$ be a linear combination of elements of E. Collate the elements of E_j for each $j = 1, \ldots, k$ and write $\mathbf{x} = \mathbf{x}_1 + \cdots + \mathbf{x}_k$, where $\mathbf{x}_j \in V_j$ for $j = 1, \ldots, k$. Suppose $\mathbf{x} = \mathbf{0}$. Then $\mathbf{x}_j = \mathbf{0}$ for $j = 1, \ldots, k$ by the previous lemma. For $j \in \{1, \ldots, k\}$, $\mathbf{x}_j$ is a linear combination of elements of the set E_j, and since the set E_j is linearly independent, every coefficient in this linear combination must be 0. Since this holds for each $j = 1, \ldots, k$, we see that every coefficient in the linear combination $\mathbf{x}$ of elements of E must be 0. Hence O.K.

The number of elements in the linearly independent set E is $g_1 + \cdots + g_k$. Since E is a subset of the n dimensional vector space $\mathbb{K}^{n\times 1}$, it follows that $g_1 + \cdots + g_k \leq n$.

Now suppose $g_1 + \cdots + g_k = n$. Then E is a linearly independent subset of $\mathbb{K}^{n\times 1}$ having n elements. Thus E is a basis for $\mathbb{K}^{n\times 1}$ consisting of eigenvectors of $\mathbf{A}$. Hence $\mathbf{A}$ is diagonalizable.

Conversely, suppose $\mathbf{A}$ is diagonalizable. Then there is a basis for $\mathbb{K}^{n\times 1}$ consisting of n eigenvectors of $\mathbf{A}$. For $j = 1, \ldots, k$, let h_j elements of this basis belong to V_j. Since these elements form a linearly independent subset of V_j, we see that $h_j \leq g_j$ for $j = 1, \ldots, k$. Hence $n = h_1 + \cdots + h_k \leq g_1 + \cdots + g_k \leq n$. This shows that $g_1 + \cdots + g_k = n$. $\square$

Corollary

If $\mathbf{A} \in \mathbb{K}^{n\times n}$ has n distinct eigenvalues, then $\mathbf{A}$ is diagonalizable.

Proof. Clearly, $n = 1 + \cdots + 1 \leq g_1 + \cdots + g_n \leq n$, and so $g_1 + \cdots + g_n = n$. Hence the above theorem applies. $\square$

The case of $\mathbb{K} = \mathbb{C}$

In case $\mathbb{K} = \mathbb{C}$, then by the Fundamental Theorem of Algebra, every polynomial of degree n with coefficients in $\mathbb{C}$ has exactly n roots in $\mathbb{C}$, counting multiplicities. In particular, the characteristic polynomial $p_{\mathbf{A}}(t)$ of any $\mathbf{A} \in \mathbb{C}^{n \times n}$ has exactly n roots in $\mathbb{C}$, counting multiplicities. More precisely, we can factor

$$p_{\mathbf{A}}(t) = (\lambda_1 - t)^{m_1} \cdots (\lambda_k - t)^{m_k},$$

where $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ are distinct and $m_1, \ldots, m_k \in \mathbb{N}$ satisfy $m_1 + \cdots + m_k = n$.

As an immediate consequence, we obtain the following result.

Theorem

Let $\mathbf{A} \in \mathbb{C}^{n \times n}$. Then there are distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ of $\mathbf{A}$ having algebraic multiplicities $m_1, \ldots, m_k$ such that $m_1 + \cdots + m_k = n$.

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) If $\mathbf{A}$ be diagonalizable, then $g_j = m_j$ for $j = 1, \ldots, k$.

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

Proof. (i) Since $\mathbf{A}$ is diagonalizable, $g_1 + \cdots + g_k = n$. Hence

$$0 \leq (m_1 - g_1) + \cdots + (m_k - g_k) \leq n - n = 0.$$

But $m_j - g_j \geq 0$, and so $g_j = m_j$ for $j = 1, \ldots, k$.

(ii) Since $\mathbb{K} = \mathbb{C}$, $m_1 + \cdots + m_k = n$. Also, since $g_j = m_j$ for $j = 1, \ldots, k$, we see that $g_1 + \cdots + g_k = m_1 + \cdots + m_k = n$. Hence $\mathbf{A}$ is diagonalizable. $\square$

Remark

Part (ii) of the above proposition does not hold if $\mathbb{K} = \mathbb{R}$.

For example, let $\mathbf{A} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$. Then $p_{\mathbf{A}}(t) = (1-t)(1+t^2)$, and the geometric multiplicity as well as the algebraic multiplicity of the only (real) eigenvalue 1 of $\mathbf{A}$ is equal to 1. Thus the 3×3 matrix $\mathbf{A}$ is not diagonalizable (over $\mathbb{R}$) since the sum of the geometric multiplicities of its eigenvalues is less than 3.

On the other hand, if $\mathbb{K} = \mathbb{C}$, then $p_{\mathbf{A}}(t) = (1-t)(t-i)(t+i)$, and for each of the eigenvalues $1, i, -i$ of $\mathbf{A}$, the geometric multiplicity as well as the algebraic multiplicity is equal to 1, and so $\mathbf{A}$ is diagonalizable (over $\mathbb{C}$).

Let $\mathbf{A}$ be a square matrix, and λ be 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 shall prove an even 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.

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 **Gershgorin disk** of the matrix $\mathbf{A}$.

Proposition (Gershgorin circle 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 $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, 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

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

Thus λ belongs to the jth Gershgorin 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 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, $\max\{|x_1|, \ldots, |x_m|\} \leq \|\mathbf{x}\| \leq |x_1| + \cdots + |x_m|$.

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

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

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

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

(i) $\|\mathbf{x}\| \geq 0$ 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) $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{K}^{n\times 1}$.