Linear Algebra
Lecture 15

Saurav Bhaumik
Department of Mathematics
IIT Bombay
Spring 2025

Theorem (Schur)

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

Proof. We use induction on n. If $n = 1$, then there is nothing to prove. Let $n \geq 2$ and assume that the result holds for all $(n-1)\times(n-1)$ matrices with complex scalars.

We shall show that there is an $n\times n$ unitary matrix $\mathbf{U}$ and also an $n\times n$ upper triangular matrix $\mathbf{B}$ such that $\mathbf{A} = \mathbf{U}\mathbf{B}\mathbf{U}^*$.

By the Fundamental Theorem of Algebra, the characteristic polynomial of $\mathbf{A}$ has a root in $\mathbb{C}$. Hence there is $\lambda_1 \in \mathbb{C}$ and there is nonzero $\mathbf{x}_1 \in \mathbb{C}^{n\times 1}$ such that $\mathbf{A}\,\mathbf{x}_1 = \lambda_1 \mathbf{x}_1$. We may assume WLOG that $\|\mathbf{x}_1\| = 1$.

We extend the orthonormal set $\{\mathbf{x}_1\}$ in $\mathbb{C}^{n\times 1}$ to an ordered orthonormal basis $E := (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ for $\mathbb{C}^{n\times 1}$.

Define $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix}$. Then $\mathbf{X}$ is a unitary matrix, so that $\mathbf{X}^{-1} = \mathbf{X}^*$.

Consider the linear transformation $T_{\mathbf{A}} : \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$ defined by $T_{\mathbf{A}}(\mathbf{x}) := \mathbf{A}\,\mathbf{x}$. Let $\mathbf{C}$ denote the matrix $\mathbf{M}_E^E(T)$ of this linear map with respect to the basis E. Then $\mathbf{A}\,\mathbf{P} = \mathbf{P}\,\mathbf{C}$, that is, $\mathbf{A} = \mathbf{P}\,\mathbf{C}\,\mathbf{P}^*$. Also, since $\mathbf{A}\,\mathbf{x}_1 = \lambda_1 \mathbf{x}_1$, we obtain

$$\mathbf{C} = \left[\begin{array}{c|ccc} \lambda_1 & \alpha_2 & \cdots & \alpha_n \\ \hline 0 & & & \\ \vdots & & \mathbf{A}_1 & \\ 0 & & & \end{array}\right],$$

where $\alpha_2, \ldots, \alpha_n \in \mathbb{C}$ and $\mathbf{A}_1 \in \mathbb{C}^{(n-1)\times(n-1)}$. By the induction hypothesis, $\mathbf{A}_1 = \mathbf{P}_1 \mathbf{B}_1 \mathbf{P}_1^*$, where $\mathbf{P}_1$ is an $(n-1)\times(n-1)$ unitary matrix and $\mathbf{B}_1 = \mathbf{P}_1^* \mathbf{A}_1 \mathbf{P}_1$ is an $(n-1)\times(n-1)$ upper triangular matrix.
We now 'border' the unitary matrix $\mathbf{P}_1$ as follows.

Define
$$\mathbf{U}_1 := \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{P}_1 & \\ 0 & & & \end{array}\right].$$

Clearly, $\mathbf{U}_1$ is unitary. Now define $\mathbf{B} := \mathbf{U}_1^*\mathbf{C}\mathbf{U}_1$. Then

$$\mathbf{B} = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{P}_1 & \\ 0 & & & \end{array}\right]^* \left[\begin{array}{c|ccc} \lambda_1 & \alpha_2 & \cdots & \alpha_n \\ \hline 0 & & & \\ \vdots & & \mathbf{A}_1 & \\ 0 & & & \end{array}\right] \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{P}_1 & \\ 0 & & & \end{array}\right]$$

that is,

$$\left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{P}_1^* & \\ 0 & & & \end{array}\right] \left[\begin{array}{c|ccc} \lambda_1 & \beta_2 & \cdots & \beta_n \\ \hline 0 & & & \\ \vdots & & \mathbf{A}_1\mathbf{P}_1 & \\ 0 & & & \end{array}\right] = \left[\begin{array}{c|ccc} \lambda_1 & \beta_2 & \cdots & \beta_n \\ \hline 0 & & & \\ \vdots & & \mathbf{P}_1^*\mathbf{A}_1\mathbf{P}_1 & \\ 0 & & & \end{array}\right],$$

where $\beta_2, \ldots, \beta_n \in \mathbb{C}$. Now the matrix

$$\mathbf{B} = \mathbf{U}_1^* \mathbf{C} \mathbf{U}_1 = \left[\begin{array}{c|ccc} \lambda_1 & \beta_2 & \cdots & \beta_n \\ \hline 0 & & & \\ \vdots & & \mathbf{P}_1^* \mathbf{A}_1 \mathbf{P}_1 & \\ 0 & & & \end{array}\right] = \left[\begin{array}{c|ccc} \lambda_1 & \beta_2 & \cdots & \beta_n \\ \hline 0 & & & \\ \vdots & & \mathbf{B}_1 & \\ 0 & & & \end{array}\right]$$

is upper triangular since $\mathbf{B}_1$ is upper triangular. Also, since $\mathbf{C} = \mathbf{U}_1 \mathbf{B}\, \mathbf{U}_1^*$, we see that

$$\mathbf{A} = \mathbf{P}\,\mathbf{C}\,\mathbf{P}^* = \mathbf{P}(\mathbf{U}_1 \mathbf{B}\, \mathbf{U}_1^*)\mathbf{P}^* = (\mathbf{P}\mathbf{U}_1)\mathbf{B}(\mathbf{P}\,\mathbf{U}_1)^*.$$

Because $\mathbf{P}$ and $\mathbf{U}_1$ are unitary, so is $\mathbf{U} := \mathbf{P}\mathbf{U}_1$, and we obtain $\mathbf{A} = \mathbf{U}\mathbf{B}\mathbf{U}^*$, as desired. □

It may seem that we have cracked the eigenvalue problem for $n \times n$ matrices if we admit complex scalars: Find an upper triangular matrix $\mathbf{B} := \mathbf{U}^{-1}\mathbf{A}\mathbf{U}$ which is (unitarily) similar to $\mathbf{A}$. Then the diagonal entries of $\mathbf{B}$ are the eigenvalues of $\mathbf{A}$. Also, we can find all eigenvectors of $\mathbf{B}$ corresponding to an eigenvalue λ by back substitution, and observe that $\mathbf{y}$ is an eigenvector of $\mathbf{B}$ if and only if $\mathbf{U}\mathbf{y}$ is an eigenvector of $\mathbf{A}$.

The above argument sounds convincing, but it is of no practical use since no algorithm is known for finding an upper triangular matrix which is similar to a given matrix. Note that our proof of the Schur theorem is based on induction.

However, there are well-known constructive methods (like the QR-algorithm) which 'upper triangularize' a given matrix approximately, that is, makes the subdiagonal entries arbitrarily small. These can be studied in a course on Numerical Analysis.

The Schur Theorem does not hold for real scalars. For example, the matrix $\mathbf{A} := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ 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 $\mathbf{A}$ does not have any real eigenvalue.

The Schur theorem is of great theoretical importance.

Corollary

Let $\mathbf{A} \in \mathbb{C}^{n \times n}$, and let $\lambda_1, \dots, \lambda_n$ 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 $j = 1, \dots, n$, whenever $p(t)$ is a polynomial satisfying $p(\mathbf{A}) = \mathbf{O}$.

Proof. By Schur's theorem, there is an upper triangular matrix $\mathbf{B}$, and also a unitary matrix $\mathbf{U}$ such that $\mathbf{A} = \mathbf{U}\mathbf{B}\mathbf{U}^{-1}$.

Since the characteristic polynomial of $\mathbf{B}$ is the same as the characteristic polynomial of $\mathbf{A}$, the eigenvalues of $\mathbf{B}$ are the same as those of $\mathbf{A}$, counted according to their algebraic multiplicities. Hence

(i)
$\operatorname{trace} \mathbf{A} = \operatorname{trace} \mathbf{U}\mathbf{B}\mathbf{U}^{-1} = \operatorname{trace} \mathbf{U}^{-1}\mathbf{U}\mathbf{B} = \operatorname{trace} \mathbf{B} = \sum_{j=1}^{n} \lambda_j$.

(ii)
$\det \mathbf{A} = \det \mathbf{U}\mathbf{B}\mathbf{U}^{-1} = \det \mathbf{U} \det \mathbf{B} \det \mathbf{U}^{-1} = \det \mathbf{B} = \prod_{j=1}^{n} \lambda_j$.

(iii) Suppose $p(t)$ is a polynomial such that $p(\mathbf{A}) = \mathbf{O}$. Since $\mathbf{B}^k = (\mathbf{U}^{-1}\mathbf{A}\mathbf{U})^k = \mathbf{U}^{-1}\mathbf{A}^k\mathbf{U}$ for all $k \in \mathbb{N}$, we see that $p(\mathbf{B}) = \mathbf{U}^{-1}p(\mathbf{A})\mathbf{U} = \mathbf{O}$.

Let $j \in \{1, \dots, n\}$. 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 $p(\mathbf{B}) = \mathbf{O}$. $\square$

The 3 results given in the above corollary are useful in determining eigenvalues of $\mathbf{A}$ in some cases.

After a detour of inner products and orthonormal sets, we come back to the matrix eigenvalue problem. We shall show that if the scalars are complex numbers, then every square matrix be 'upper triangularized'. For this purpose, it is convenient to talk about eigenvalues and eigenvectors of a linear transformation.

Let V be a vector subspace of dimension n (possibly contained in a higher dimensional space of column vectors). Let $T : V \to V$ be a linear transformation. We say that $\lambda \in \mathbb{K}$ is an **eigenvalue** of T if there is nonzero $\mathbf{x} \in V$ such that $T(\mathbf{x}) = \lambda\,\mathbf{x}$, and then such a nonzero vector is called an **eigenvector** of T corresponding to λ.

Recall that if $E := (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is an ordered basis for V, then the linear transformation T can be represented by the $n \times n$ matrix $\mathbf{A} := \mathbf{M}_E^E(T)$ whose kth column is $\begin{bmatrix} a_{1k} & \cdots & a_{nk} \end{bmatrix}^{\mathsf{T}}$, where $T(\mathbf{x}_k) = a_{1k}\mathbf{x}_1 + \cdots + a_{nk}\mathbf{x}_n$.

Lemma

Let V be an n dimensional vector subspace, and let E be an ordered basis for V. Suppose $T : V \to V$ is a linear map. Then $\lambda \in \mathbb{K}$ is an eigenvalue of T if and only if λ is an eigenvalue of $M_E^E(T)$.
In particular, if $\mathbb{K} := \mathbb{C}$, then T has an eigenvalue.

Proof. Let $E := (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $\mathbf{A} := M_E^E(T) = [a_{jk}]$.
Consider $\lambda \in \mathbb{K}$. Let $\mathbf{x} := \alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n \in V$. Then

$$T(\mathbf{x}) = \lambda\,\mathbf{x} \iff \mathbf{A}\begin{bmatrix}\alpha_1 & \cdots & \alpha_n\end{bmatrix}^{\mathsf{T}} = \lambda\begin{bmatrix}\alpha_1 & \cdots & \alpha_n\end{bmatrix}^{\mathsf{T}}.$$

Also, $\mathbf{x} \neq \mathbf{0} \iff \begin{bmatrix}\alpha_1 & \cdots & \alpha_n\end{bmatrix}^{\mathsf{T}} \neq \mathbf{0}$. Hence $\mathbf{x}$ is an eigenvector of $T \iff \begin{bmatrix}\alpha_1 & \cdots & \alpha_n\end{bmatrix}^{\mathsf{T}}$ is an eigenvector of $\mathbf{A}$.
Let $\mathbb{K} = \mathbb{C}$. Since $\mathbf{A}$ has an eigenvalue, so does T $\square$

Proposition

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. Let $F = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be any othonormal basis of V, and let $\mathbf{B} = \mathbf{M}_F^F(T)$. This means, $T[\mathbf{x}_1, \ldots, \mathbf{x}_n] = [\mathbf{x}_1, \ldots, \mathbf{x}_n]\mathbf{B}$. By Schur's theorem, there is a unitary matrix $\mathbf{U}$ such that $\mathbf{A} = \mathbf{U}^{-1}\mathbf{B}\mathbf{U}$ is upper triangular. Now $T[\mathbf{x}_1, \ldots, \mathbf{x}_1]\mathbf{U} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]\mathbf{B}\mathbf{U} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]\mathbf{U}\mathbf{A}$. Write $[\mathbf{x}_1, \ldots, \mathbf{x}_n]\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$. As $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is an orthonormal basis, $[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ is unitary. As $\mathbf{U}$ is also unitary, so is $[\mathbf{u}_1, \ldots, \mathbf{u}_n]$, which means $E := (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ is an orthonormal basis. Again, the above implies, $T[\mathbf{u}_1, \ldots, \mathbf{u}_n] = [\mathbf{u}_1, \ldots, \mathbf{u}_n]\mathbf{A}$, which means $\mathbf{M}_E^E(T) = \mathbf{A}$, which is upper triangular. □

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**. Observe that $\mathbf{A}^n = \mathbf{I}$ and also that $\mathbf{A}$ is a unitary matrix given by

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

Let λ be an eigenvalue of $\mathbf{A}$. Since $\mathbf{A}^n = \mathbf{I}$, we see that $\lambda^n = 1$. 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 $\mathbf{A}\mathbf{x} = \lambda\,\mathbf{x}$ 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 $\mathbf{A}\mathbf{x}_j = \begin{bmatrix} \omega^j & 1 & 1/\omega^j & 1/\omega^{2j} & \cdots & 1/\omega^{(n-2)j} \end{bmatrix}^{\mathsf{T}} = \omega^j \mathbf{x}_j$. Thus $\mathbf{x}_j$ is an eigenvector of $\mathbf{A}$ corresponding to the eigenvalue ω^j for $j = 0, 1, \ldots, n-1$.

Remark: $\mathbf{A}$ is an example of a circulant matrix. In general a circulant matrix is a square matrix each of whose rows are obtained by a cyclic shift to the right of the preceding rows.