ENGG 5781: Matrix Analysis and Computations **2022-23 First Term**

Lecture 3: Eigenvalues and Eigenvectors

Instructor: Wing-Kin Ma

1 Eigenvalue Problem

The eigenvalue problem is as follows. Given $\mathbf{A} \in \mathbb{C}^{n\times n}$, find a vector $\mathbf{v} \in \mathbb{C}^n$, $\mathbf{v} \neq \mathbf{0}$, such that

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}, \tag{1}$$

for some $\lambda \in \mathbb{C}$. If we can find a 2-tuple $(\mathbf{v}, \lambda)$ such that (1) holds, we say that $(\mathbf{v}, \lambda)$ is an eigen-pair of $\mathbf{A}$, λ is an eigenvalue of $\mathbf{A}$, and $\mathbf{v}$ is an eigenvector of $\mathbf{A}$ associated with λ. From (1), we observe two facts. First, (1) is equivalent to finding a λ such that

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0. \tag{2}$$

Second, (1) is equivalent to

$$\mathbf{v} \in \mathcal{N}(\mathbf{A} - \lambda \mathbf{I}), \qquad \mathcal{N}(\mathbf{A} - \lambda \mathbf{I}) \neq \{\mathbf{0}\}. \tag{3}$$

Let us consider the solution to (2). Denote $p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$. From the definition of determinant, one can verify that $p(\lambda)$ is a polynomial of degree n. Since a polynomial of degree n has n roots, we may write

$$p(\lambda) = \prod_{i=1}^{n} (\lambda_i - \lambda),$$

where $\lambda_1, \ldots, \lambda_n$ are the roots of $p(\lambda)$. The above equation shows that (2) holds if and only if $\lambda = \lambda_i$ for any $i = 1, \ldots, n$. Thus, we conclude that the eigenvalue problem in (1) always has a solution, and that the roots $\lambda_1, \ldots, \lambda_n$ of $p(\lambda)$ are the solutions to (1) . For convenience, let us denote

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i = 1, \ldots, n, \tag{4}$$

where $\mathbf{v}_i$ denotes an eigenvector associated with λ_i.

Next, we take a look at the solution to (3) w.r.t. $\mathbf{v}$, given an eigenvalue λ. Since $\mathcal{N}(\mathbf{A} - \lambda \mathbf{I})$ is a subspace and $\mathcal{N}(\mathbf{A} - \lambda \mathbf{I}) \neq \{\mathbf{0}\}$, we can represent it by $\mathcal{N}(\mathbf{A} - \lambda \mathbf{I}) = \text{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_r\}$ for some basis $\{\mathbf{b}_1, \ldots, \mathbf{b}_r\} \subset \mathbb{C}^n$ and for some $r \geq 1$. In particular, the dimension of $\mathcal{N}(\mathbf{A} - \lambda \mathbf{I})$ is r. From the above representation, we notice that there may be multiple eigenvectors associated with the same λ. We should be careful when we describe the multiplicity of eigenvectors: A scaled version of an eigenvector $\mathbf{v}$, i.e., $\alpha\mathbf{v}$ for some $\alpha \in \mathbb{C}$, $\alpha \neq 0$, is also an eigenvector, but such a case is trivial. From such a viewpoint, the eigenvector $\mathbf{v}$ associated with λ is unique subject to a complex scaling if $\dim \mathcal{N}(\mathbf{A} - \lambda \mathbf{I}) = 1$. Moreover, for $\dim \mathcal{N}(\mathbf{A} - \lambda \mathbf{I}) > 1$, there are infinitely many eigenvectors associated with λ even if we do not count the complex scaling cases; however, we can find a number of $r = \dim \mathcal{N}(\mathbf{A} - \lambda \mathbf{I})$ linearly independent eigenvectors associated with λ. Also, $\dim \mathcal{N}(\mathbf{A} - \lambda \mathbf{I})$ is the maximal number of linearly independent eigenvectors we can obtain for λ.

2 Multiplicity of Eigenvalues and Eigenvectors

We are concerned with the multiplicity of an eigenvalue and the multiplicity of its associated eigenvectors. For ease of exposition of ideas, let us assume w.l.o.g. that the eigenvalues $\lambda_1, \ldots, \lambda_n$ are ordered such that $\{\lambda_1, \ldots \lambda_k\}$, $k \leq n$, is the set of all distinct eigenvalues of $\mathbf{A}$; i.e., $\lambda_i \neq \lambda_j$ for all $i, j \in \{1, \ldots, k\}$ with $i \neq j$, and $\lambda_i \in \{\lambda_1, \ldots \lambda_k\}$ for all $i \in \{1, \ldots, n\}$. Then, consider the following definitions:

- The *algebraic multiplicity* of an eigenvalue λ_i, $i \in \{1, \ldots, k\}$, is defined as the number of times that λ_i appears as a root of $p(\lambda)$. We will denote the algebraic multiplicity of λ_i as μ_i.
- The *geometric multiplicity* of λ_i, $i \in \{1, \ldots, k\}$, is defined as the maximal number of linearly independent eigenvectors associated with λ_i. We will denote the geometric multiplicity of λ_i as γ_i, and note that $\gamma_i = \dim \mathcal{N}(\mathbf{A} - \lambda_i \mathbf{I})$.

Intuitively, it seems that if an eigenvalue is repeated r times, then we should also have r (linearly independent) eigenvectors associated with it. We will show that

Property 3.1 *We have $\mu_i \geq \gamma_i$ for $i = 1, \ldots, k$.*

However, there exist instances for which $\mu_i > \gamma_i$. An example is as follows:

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

It can be verified that the roots of $p(\lambda)$ are $\lambda_1 = \lambda_2 = \lambda_3 = 0$. Thus, we have $\mu_1 = 3$, $k = 1$. However, one can also verify that

$$\mathcal{N}(\mathbf{A} - \lambda_1 \mathbf{I}) = \mathcal{N}(\mathbf{A}) = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\},$$

and consequently, $\gamma_1 = \dim \mathcal{N}(\mathbf{A} - \lambda_1 \mathbf{I}) = 2$.

Proof of Property 3.1: For convenience, let $\bar{\lambda} \in \{\lambda_1, \ldots, \lambda_k\}$ be any eigenvalue of $\mathbf{A}$, and denote $r = \dim \mathcal{N}(\mathbf{A} - \bar{\lambda}\mathbf{I})$. We aim to show that the characteristic polynomial $\det(\mathbf{A} - \lambda \mathbf{I})$ has at least r repeated roots for $\lambda = \bar{\lambda}$. From basic subspace concepts (cf. Lecture 1), we can find a collection of orthonormal vectors $\mathbf{q}_1, \ldots, \mathbf{q}_r \in \mathcal{N}(\mathbf{A} - \lambda \mathbf{I})$ and a collection of vectors $\mathbf{q}_{r+1}, \ldots, \mathbf{q}_n \in \mathbb{C}^n$ such that $\mathbf{Q} = [\ \mathbf{q}_1, \ldots, \mathbf{q}_n\ ]$ is unitary. Let $\mathbf{Q}_1 = [\ \mathbf{q}_1, \ldots, \mathbf{q}_r\ ]$, $\mathbf{Q}_2 = [\ \mathbf{q}_{r+1}, \ldots, \mathbf{q}_n\ ]$, and note $\mathbf{Q} = [\ \mathbf{Q}_1\ \mathbf{Q}_2\ ]$. We have

$$\mathbf{Q}^H \mathbf{A} \mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1^H \\ \mathbf{Q}_2^H \end{bmatrix} [\ \mathbf{A}\mathbf{Q}_1\ \mathbf{A}\mathbf{Q}_2\ ] = \begin{bmatrix} \mathbf{Q}_1^H \mathbf{A} \mathbf{Q}_1 & \mathbf{Q}_1^H \mathbf{A} \mathbf{Q}_2 \\ \mathbf{Q}_2^H \mathbf{A} \mathbf{Q}_1 & \mathbf{Q}_2^H \mathbf{A} \mathbf{Q}_2 \end{bmatrix}.$$

Since $\mathbf{A}\mathbf{q}_i = \bar{\lambda}\mathbf{q}_i$ for $i = 1, \ldots, r$, we get $\mathbf{A}\mathbf{Q}_1 = \bar{\lambda}\mathbf{Q}_1$. By also noting that $\mathbf{Q}_1^H \mathbf{Q}_1 = \mathbf{I}$ and $\mathbf{Q}_2^H \mathbf{Q}_1 = \mathbf{0}$, the above matrix equation can be simplified to

$$\mathbf{Q}^H \mathbf{A} \mathbf{Q} = \begin{bmatrix} \bar{\lambda}\mathbf{I} & \mathbf{Q}_1^H \mathbf{A} \mathbf{Q}_2 \\ \mathbf{0} & \mathbf{Q}_2^H \mathbf{A} \mathbf{Q}_2 \end{bmatrix}.$$

Consequently, we have the following equivalence for the characteristic polynomial of $\mathbf{A}$:

$$
\begin{aligned}
\det(\mathbf{A} - \lambda\mathbf{I}) &= \det(\mathbf{Q}^H(\mathbf{A} - \lambda\mathbf{I})\mathbf{Q}) = \det(\mathbf{Q}^H\mathbf{A}\mathbf{Q} - \lambda\mathbf{I}) \\
&= \det(\bar{\lambda}\mathbf{I} - \lambda\mathbf{I}) \det(\mathbf{Q}_2^H\mathbf{A}\mathbf{Q}_2 - \lambda\mathbf{I}) \\
&= \left(\bar{\lambda} - \lambda\right)^r \det(\mathbf{Q}_2^H\mathbf{A}\mathbf{Q}_2 - \lambda\mathbf{I})
\end{aligned}
$$

where the second equality is due to the determinant result for block upper triangular matrices; here note that $\det(\mathbf{Q}_2^H\mathbf{A}\mathbf{Q}_2 - \lambda\mathbf{I})$ is a polynomial of degree of $n - r$. From the above equation we see that $\det(\mathbf{A} - \lambda\mathbf{I})$ has at least r repeated roots for $\lambda = \bar{\lambda}$. The proof is complete. ■

3 Similarity, Diagonalizability, and Eigendecomposition

3.1 Similarity and Diagonalizability

To set the stage for describing eigendecomposition, we start with introducing the concepts of similarity and diagonalizability. A matrix $\mathbf{B} \in \mathbb{C}^{n\times n}$ is said to be *similar* to another matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ if there exists a nonsingular $\mathbf{S} \in \mathbb{C}^{n\times n}$ such that

$$\mathbf{B} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}.$$

Similar matrices are similar in the sense that their characteristic polynomials are the same. Specifically, if $\mathbf{A}$ is similar to $\mathbf{B}$ then we have

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det(\mathbf{S}^{-1}(\mathbf{A} - \lambda\mathbf{I})\mathbf{S}) = \det(\mathbf{B} - \lambda\mathbf{I}).$$

It is easy to verify that similar matrices have the following properties:

1. If $\mathbf{B}$ is similar to $\mathbf{A}$, $\mathbf{A}$ is also similar to $\mathbf{B}$.
2. If $\mathbf{A}, \mathbf{B}$ are similar, they have the same set of eigenvalues.
3. If $\mathbf{A}, \mathbf{B}$ are similar, they have $\det(\mathbf{A}) = \det(\mathbf{B})$.

In matrix analysis we are curious about whether a matrix can be similar to a diagonal matrix—obviously because diagonal matrices are easy to deal with. A matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is said to be *diagonalizable* if it is similar to a diagonal matrix; i.e., there exists a nonsingular $\mathbf{S} \in \mathbb{C}^{n\times n}$ and a diagonal $\mathbf{D} \in \mathbb{C}^{n\times n}$ such that

$$\mathbf{D} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S},$$

or equivalently,

$$\mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}.$$

Now, observe that the above equation can be equivalently rewritten as $\mathbf{A}\mathbf{S} = \mathbf{S}\mathbf{D}$, or, in column-by-column form

$$\mathbf{A}\mathbf{s}_i = d_i\mathbf{s}_i, \quad i = 1, \ldots, n,$$

where d_i denotes the (i, i)th entry of $\mathbf{D}$. Hence, every $(\mathbf{s}_i, d_i)$ must be an eigen-pair of $\mathbf{A}$. Also, since $\mathbf{A}$ and $\mathbf{D}$ are similar they have the same set of eigenvalues; more precisely, $d_1, \ldots, d_n$ must equal $\lambda_1, \ldots, \lambda_n$ or any of its permutation. Let us assume w.l.o.g. that $d_i = \lambda_i$ for $i = 1, \ldots, n$. Then every $\mathbf{s}_i$ is an eigenvector associated with λ_i. Furthermore, it is important to note that $\mathbf{s}_1, \ldots, \mathbf{s}_n$ must be linearly independent.

3.2 Eigendecomposition

We are now ready to consider eigendecomposition. A matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ is said to admit an eigendecomposition if there exists a nonsingular $\mathbf{V} \in \mathbb{C}^{n \times n}$ such that

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1},$$

where $\mathbf{\Lambda} = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\mathbf{A}$. Essentially, *eigendecomposition is the same as the diagonalization discussed above.*

A natural question that arises from the above definition is whether eigendecomposition exists. From the discussion in the last subsection, we see that the question is the same as asking whether we can find a collection of eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, with each $\mathbf{v}_i$ being associated with λ_i, such that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent. Let us consider the following property:

Property 3.2 *Let $\mathbf{A} \in \mathbb{C}^{n \times n}$, and suppose that the eigenvalues $\lambda_1, \ldots, \lambda_n$ are ordered such that $\{\lambda_1, \ldots \lambda_k\}$, $k \leq n$, is the set of all distinct eigenvalues of $\mathbf{A}$. Also, let $\mathbf{v}_i$ be* any *eigenvector associated with λ_i. Then $\mathbf{v}_1, \ldots, \mathbf{v}_k$ must be linearly independent.*

We relegate the proof of the above property to the next subsection. Property 3.2 gives rise to the following implications:

1. If the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\mathbf{A}$ are all distinct, then $\mathbf{A}$ admits an eigendecomposition.
2. By considering also Property 3.1, $\mathbf{A}$ admits an eigendecomposition if and only if $\mu_i = \gamma_i$ for all $i = 1, \ldots, k$.
3. $\mathbf{A}$ does not admit an eigendecomposition if $\mu_i > \gamma_i$ for some $i \in \{1, \ldots, k\}$. Such instances exist as discussed in the last section.

Before we close this section, we should mention that eigendecomposition is guaranteed to exist in some matrix subclasses. For example, a circulant matrix always admits an eigendecomposition. Also, we will show later that it is easy to find an arbitrarily close approximation of a given matrix $\mathbf{A}$ such that the approximate matrix has its eigenvalues all being distinct and thus admits an eigendecomposition.

3.3 Proof of Property 3.2

We prove Property 3.2 by contradiction. Suppose that $\lambda_i \neq \lambda_j$ for all $i, j \in \{1, \ldots, k\}$ with $i \neq j$, but we can find a collection of linearly dependent $\mathbf{v}_1, \ldots, \mathbf{v}_k$, i.e.,

$$\sum_{i=1}^{k} \alpha_i \mathbf{v}_i = \mathbf{0}, \tag{5}$$

for some $\boldsymbol{\alpha} \neq \mathbf{0}$. Let us assume w.l.o.g. that $\alpha_1 \neq 0$, $k \geq 2$. From (5), we obtain two equations

$$\mathbf{0} = \mathbf{A}\left(\sum_{i=1}^{k} \alpha_i \mathbf{v}_i\right) = \sum_{i=1}^{k} \alpha_i \lambda_i \mathbf{v}_i, \tag{6}$$

$$\mathbf{0} = \lambda_k\left(\sum_{i=1}^{k} \alpha_i \mathbf{v}_i\right) = \sum_{i=1}^{k} \alpha_i \lambda_k \mathbf{v}_i. \tag{7}$$

Subtracting (6) from (7) results in

$$\sum_{i=1}^{k-1} \alpha_i(\lambda_k - \lambda_i)\mathbf{v}_i = \mathbf{0}. \tag{8}$$

By repeatedly applying the trick in (5)-(8), one can show that

$$\left[\alpha_1 \prod_{i=2}^{k} (\lambda_i - \lambda_1)\right] \mathbf{v}_1 = \mathbf{0}.$$

Since $\alpha_1 \neq 0$ and $\lambda_i \neq \lambda_j$ for all $i, j \in \{1, \ldots, k\}$, $i \neq j$, the above equation holds only when $\mathbf{v}_1 = \mathbf{0}$. This contradicts the fact that $\mathbf{v}_1$ is an eigenvector, and the proof is complete.

4 Eigendecomposition for Hermitian and Real Symmetric Matrices

A square matrix $\mathbf{A}$ is said to be *Hermitian* if $\mathbf{A} = \mathbf{A}^H$, and *symmetric* if $\mathbf{A} = \mathbf{A}^T$. We denote $\mathbb{H}^n$ to be the set of all $n \times n$ complex Hermitian matrices and $\mathbb{S}^n$ to be the set of all $n \times n$ real symmetric matrices. As we will see, Hermitian and real symmetric matrices always admit eigendecompositions. To give some insights, consider the following property.

Property 3.3 *Let $\mathbf{A} \in \mathbb{H}^n$. We have the following results:*

1. *The eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\mathbf{A}$ are real.*
2. *Suppose that λ_i's are ordered such that $\{\lambda_1, \ldots, \lambda_k\}$, $k \leq n$, is the set of all distinct eigenvalues of $\mathbf{A}$. Also, let $\mathbf{v}_i$ be* any *eigenvector associated with λ_i. Then $\mathbf{v}_1, \ldots, \mathbf{v}_k$ must be orthonormal.*

Proof: For any eigen-pair $(\mathbf{v}, \lambda)$ of $\mathbf{A}$, we have

$$\mathbf{v}^H \mathbf{A} \mathbf{v} = \mathbf{v}^H (\lambda \mathbf{v}) = \lambda \|\mathbf{v}\|_2^2.$$

If $\mathbf{A}$ is Hermitian we also have

$$\mathbf{v}^H \mathbf{A} \mathbf{v} = (\mathbf{A}\mathbf{v})^H \mathbf{v} = (\lambda \mathbf{v})^H \mathbf{v} = \lambda^* \|\mathbf{v}\|_2^2.$$

The above two equations implies that $\lambda = \lambda^*$, or that λ is real. Moreover, consider the following equations for any $i, j \in \{1, \ldots, k\}$, $i \neq j$:

$$\begin{aligned} \mathbf{v}_j^H \mathbf{A} \mathbf{v}_i &= \mathbf{v}_j^H (\mathbf{A} \mathbf{v}_i) = \lambda_i \mathbf{v}_j^H \mathbf{v}_i, \\ \mathbf{v}_j^H \mathbf{A} \mathbf{v}_i &= (\mathbf{A} \mathbf{v}_j)^H \mathbf{v}_j = \lambda_j \mathbf{v}_j^H \mathbf{v}_i, \end{aligned}$$

where in the second equation we have used the fact that $\mathbf{A}$ is Hermitian and its corresponding eigenvalues are real. The above equations imply $(\lambda_i - \lambda_j)\mathbf{v}_j^H \mathbf{v}_i = 0$. Since $\lambda_i \neq \lambda_j$, it must hold that $\mathbf{v}_j^H \mathbf{v}_i = 0$. Hence, we have shown that any collection of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ must be orthonormal. ■

Note that as a direct corollary of the first result of Property 3.3, any eigenvector of a real symmetric matrix can be taken as real. Next, we present the main result.

Theorem 3.1 *Every* $\mathbf{A} \in \mathbb{H}^n$ *admits an eigendecomposition*

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H,$$

where $\mathbf{V} \in \mathbb{C}^{n\times n}$ *is unitary, and* $\mathbf{\Lambda} = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ *with* $\lambda_i \in \mathbb{R}$ *for all* i. *Also, if* $\mathbf{A} \in \mathbb{S}^n$, $\mathbf{V}$ *can be taken as real orthogonal.*

The proof of Theorem 3.1 requires another theorem, namely, the *Schur deocmposition theorem.* We will consider the Schur decomposition in the next section.

5 The Schur Decomposition

If we cannot always diagonalize a matrix, which is true as we discussed, our next question would be to ask if a matrix can always be similar to another matrix that is closer to a diagonal matrix. The Schur decomposition seeks to look at *triangularizability*—given $\mathbf{A} \in \mathbb{C}^{n\times n}$, can we find a nonsingular $\mathbf{S} \in \mathbb{C}^{n\times n}$ such that

$$\mathbf{A} = \mathbf{S}\mathbf{T}\mathbf{S}^{-1}$$

for some upper triangular $\mathbf{T} \in \mathbb{C}^{n\times n}$? Having such a result is still good: When $\mathbf{T}$ is upper triangular, its characteristic polynomial is simply

$$\det(\mathbf{T} - \lambda\mathbf{I}) = \prod_{i=1}^{n}(t_{ii} - \lambda).$$

As a result, we can have $t_{ii} = \lambda_i$ for all i.

As a beautiful result in matrix analysis, we can always triangularize a matrix. The result is summarized as follows.

Theorem 3.2 *Let* $\mathbf{A} \in \mathbb{C}^{n\times n}$, *and let* $\lambda_1, \ldots, \lambda_n$ *be its eigenvalues. The matrix* $\mathbf{A}$ *admits a decomposition*

$$\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H, \tag{9}$$

for some unitary $\mathbf{U} \in \mathbb{C}^{n\times n}$ *and for some upper triangular* $\mathbf{T} \in \mathbb{C}^{n\times n}$ *with* $t_{ii} = \lambda_i$ *for all* i. *If* $\mathbf{A}$ *is real and* $\lambda_1, \ldots, \lambda_n$ *are all real,* $\mathbf{U}$ *and* $\mathbf{T}$ *can be taken as real.*

The decomposition in (9) will be called the *Schur decomposition* in the sequel. The Schur decomposition not only shows that any square matrix is similar to an upper triangular matrix, it also reveals that the "triangularizer" $\mathbf{S}$ can be unitary.

5.1 Proof of the Schur Decomposition

Consider the following lemma.

Lemma 3.1 *Let* $\mathbf{X} \in \mathbb{C}^{n\times n}$ *and suppose that* $\mathbf{X}$ *takes a block upper triangular form*

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{0} & \mathbf{X}_{22} \end{bmatrix},$$

where $\mathbf{X}_{11} \in \mathbb{C}^{k\times k}$, $\mathbf{X}_{12} \in \mathbb{C}^{k\times(n-k)}$, $\mathbf{X}_{22} \in \mathbb{C}^{(n-k)\times(n-k)}$, *with* $0 \leq k < n$. *There exists an unitary* $\mathbf{U} \in \mathbb{C}^{n\times n}$ *such that*

$$\mathbf{U}^H\mathbf{X}\mathbf{U} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{Y}_{12} \\ \mathbf{0} & \mathbf{Y}_{22} \end{bmatrix}, \quad \mathbf{Y}_{22} = \begin{bmatrix} \bar{\lambda} & \times \\ \mathbf{0} & \times \end{bmatrix},$$

for some $\mathbf{Y}_{12} \in \mathbb{C}^{k\times(n-k)}$, $\mathbf{Y}_{22} \in \mathbb{C}^{(n-k)\times(n-k)}$, $\bar{\lambda} \in \mathbb{C}$.

Proof of Lemma 3.1: The proof may be seen as a variation of the proof of Property 3.1 in Section 2. Let $\bar{\lambda}$ be any eigenvalue of $\mathbf{X}_{22}$, and $\mathbf{v} \in \mathbb{C}^{n-k}$ be an eigenvector of $\mathbf{X}_{22}$ associated with $\bar{\lambda}$. Following the same proof as in Property 3.1, there exists a collection of vectors $\mathbf{q}_2, \ldots, \mathbf{q}_{n-k} \in \mathbb{C}^{n-k}$ such that $\mathbf{Q} = [\ \mathbf{v},\ \mathbf{q}_2, \ldots, \mathbf{q}_{n-k}\ ] \in \mathbb{C}^{(n-k)\times(n-k)}$ is unitary, and it can be shown that $\mathbf{Q}^H\mathbf{X}_{22}\mathbf{Q}$ takes the form

$$\mathbf{Q}^H\mathbf{X}_{22}\mathbf{Q} = \begin{bmatrix} \bar{\lambda} & \times \\ \mathbf{0} & \times \end{bmatrix}.$$

Now, let

$$\mathbf{U} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}.$$

We have

$$\begin{aligned} \mathbf{U}^H\mathbf{X}\mathbf{U} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^H \end{bmatrix} \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{0} & \mathbf{X}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^H \end{bmatrix} \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12}\mathbf{Q} \\ \mathbf{0} & \mathbf{X}_{22}\mathbf{Q} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12}\mathbf{Q} \\ \mathbf{0} & \mathbf{Q}^H\mathbf{X}_{22}\mathbf{Q} \end{bmatrix}. \end{aligned}$$

The result is obtained, as desired. ■

The idea of proving the Schur decomposition in Theorem 3.2 is to apply Lemma 3.1 recursively. To put into context, let $\mathbf{A}_0 = \mathbf{A}$ and consider the following iterations:

$$\begin{aligned} \mathbf{A}_1 &= \mathbf{U}_1^H\mathbf{A}_0\mathbf{U}_1 \\ \mathbf{A}_2 &= \mathbf{U}_2^H\mathbf{A}_1\mathbf{U}_2 \\ &\vdots \\ \mathbf{A}_{n-1} &= \mathbf{U}_{n-1}^H\mathbf{A}_{n-2}\mathbf{U}_{n-1} \end{aligned}$$

where every $\mathbf{U}_i$ is unitary and obtained by applying Lemma 3.1 with $\mathbf{X} = \mathbf{A}_{i-1}$ and $k = i-1$. From Lemma 3.1 we observe that $\mathbf{A}_i$ takes the form

$$\mathbf{A}_i = \begin{bmatrix} \mathbf{T}_{ii} & \times \\ \mathbf{0} & \times \end{bmatrix},$$

for some upper triangular $\mathbf{T}_{ii} \in \mathbb{C}^{i\times i}$. Hence, it follows that $\mathbf{A}_{n-1}$ is upper triangular. Let

$$\mathbf{U} = \mathbf{U}_1\mathbf{U}_2\cdots\mathbf{U}_{n-1}.$$

It can be verified that $\mathbf{U}$ is unitary. By noting that $\mathbf{A}_{n-1} = \mathbf{U}^H\mathbf{A}\mathbf{U}$, we obtain the Schur decomposition formula $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H$ where $\mathbf{T} = \mathbf{A}_{n-1}$.

We should also mention how the result $t_{ii} = \lambda_i$ for all i is concluded. By the similarity of $\mathbf{A}$ and $\mathbf{T}$ and by $\det(\mathbf{T} - \lambda\mathbf{I}) = \prod_{i=1}^n (t_{ii} - \lambda)$, the n roots $t_{11}, \ldots, t_{nn}$ of $\det(\mathbf{T} - \lambda\mathbf{I})$ must be $\lambda_1, \ldots, \lambda_n$ or its permutation. Hence, we can assume w.l.o.g. that $t_{ii} = \lambda_i$ for $i = 1, \ldots, n$.

Furthermore, it can be verified that if $\mathbf{A}$ is real and its eigenvalues $\lambda_1, \ldots, \lambda_n$ are also real, then we can choose $\mathbf{U}$ as real orthogonal and $\mathbf{T}$ as real in the Schur decomposition. This is left as a self-practice problem for you.

5.2 Implications of the Schur Decomposition

Here we discuss some implications of the Schur decomposition.

- *Computations of the Schur decomposition:* The proof of the Schur decomposition is constructive, and it also tells how we may write an algorithm to compute the Schur factors $\mathbf{U}$ and $\mathbf{T}$. From the proof in the last subsection, we see that we need two sub-algorithms to construct $\mathbf{U}$ and $\mathbf{T}$, namely, i) an algorithm for computing an eigenvector of a given matrix, and ii) an algorithm that finds a unitary matrix $\mathbf{Q}$ such that its first column vector is fixed as a given vector $\mathbf{v}$. Task i) may be done by the power method, while Task ii) can be accomplished by the so-called QR decomposition—which we will study later.

  It should be further mentioned that the procedure mentioned above is arguably not the best approach for computing the Schur factors, although it is top-down, insightful and easy to understand. There exist computationally more efficient methods for computing the Schur factors, and they were established based on some rather specific and sophisticated ideas which are beyond the scope of this course. The only keyword I can give you is QR decomposition.

- *Proof of Theorem 3.1 and beyond:* With the Schur decomposition, we can easily show that any Hermitian matrix admits an eigendecomposition. Let $\mathbf{A}$ be Hermitian and let $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H$ be its Schur decomposition. We see that

  $$\mathbf{0} = \mathbf{A} - \mathbf{A}^H = \mathbf{U}\mathbf{T}\mathbf{U}^H - \mathbf{U}\mathbf{T}^H\mathbf{U}^H = \mathbf{U}(\mathbf{T} - \mathbf{T}^H)\mathbf{U}^H.$$

  This implies that $\mathbf{T} - \mathbf{T}^H = \mathbf{0}$, or $\mathbf{T}$ must be diagonal. Consequently, the Schur dcomposition reduces to the eigendecomposition with unitary $\mathbf{V}$. The implication $\mathbf{T} - \mathbf{T}^H = \mathbf{0}$ also implies that $t_{ii} = t_{ii}^*$, or the eigenvalues λ_i's are real. The same result also applies to a real symmetric $\mathbf{A}$.[1] Using the same idea, we can also show that if $\mathbf{A}$ satisfies $\mathbf{A} = -\mathbf{A}^H$, i.e., it is the so-called skew-Hermitian matrix, then $\mathbf{A}$ admits an eigendecomposition with unitary $\mathbf{V}$ and with purely imaginary λ_i's.

- *Existence of eigendecomposition:* We have mentioned that even though $\mathbf{A}$ does not admit an eigendecomposition, it is not hard to find an approximation of $\mathbf{A}$ such that the approximate matrix admits an eigendecomposition. This is described as follows.

  Proposition 3.1 *Let $\mathbf{A} \in \mathbb{C}^{n\times n}$. For every $\varepsilon > 0$, there exists a matrix $\tilde{\mathbf{A}} \in \mathbb{C}^{n\times n}$ such that the n eigenvalues of $\tilde{\mathbf{A}}$ are distinct and*

  $$\|\mathbf{A} - \tilde{\mathbf{A}}\|_F \leq \varepsilon.$$

[1] As a minor point to note, for the real symmetric case we first need the first result of Property 3.3 to confirm that λ_i's are real.

The proof is simple and is follows: Let $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H$ be the Schur decomposition of $\mathbf{A}$, and let $\tilde{\mathbf{A}} = \mathbf{U}(\mathbf{T} + \mathbf{D})\mathbf{U}^H$ where $\mathbf{D} = \mathrm{Diag}(d_1, \ldots, d_n)$ and $d_1, \ldots, d_n$ are chosen such that $|d_i|^2 \leq \varepsilon/n$ for all i and such that $t_{11} + d_1, \ldots, t_{nn} + d_n$ are distinct. It is easy to verify that $\|\mathbf{A} - \tilde{\mathbf{A}}\|_F^2 = \|\mathbf{U}\mathbf{D}\mathbf{U}^H\|_F^2 = \|\mathbf{D}\|_F^2 \leq \varepsilon^2$.

The above proposition suggests that for any square $\mathbf{A}$, we can always find a matrix $\tilde{\mathbf{A}}$ that is arbitrarily close to $\mathbf{A}$ and admits an eigendecomposition.

- *The Jordan canonical form:* Recall that we have been asking the question of whether a matrix can be similar to a diagonal matrix, and if not possible, a matrix that is closer to the diagonal matrix. The Jordan canonical form says the following: any square $\mathbf{A}$ can be decomposed as $\mathbf{A} = \mathbf{S}\mathbf{J}\mathbf{S}^{-1}$ where $\mathbf{S}$ is nonsingular and $\mathbf{J}$ takes some kind of tri-diagonal structures. The Jordon canonical form is beyond the scope of this course, but we should note that it is an enhancement of the Schur decomposition. In a nutshell, the proof of the Jordon canonical form first applies the Schur decomposition. Then, as the non-trivial part of the proof, it shows that the Schur factor $\mathbf{T}$ is similar to another matrix $\mathbf{J}$ that takes the tri-diagonal structure.