Schur Triangularization

Introduction

This chapter delves into the fascinating world of **matrix decompositions**, which are methods for expressing a matrix as a product of other, often simpler, matrices. A familiar example from introductory linear algebra is **diagonalization**. Recall that for a given square matrix $A \in M_n(F)$ (where F is the underlying field), we can sometimes find a diagonal matrix $D \in M_n(F)$ and an invertible matrix $P \in M_n(F)$ such that

$$A = PDP^{-1}$$

This is possible if and only if there exists a basis of F^n consisting of eigenvectors of A. However, not all matrices possess a full set of linearly independent eigenvectors, meaning diagonalization is not universally applicable. This limitation raises important questions:

1. How close to a diagonal matrix can we make an arbitrary matrix A using a **similarity transformation** (i.e., transforming A to $P^{-1}AP$)?

2. What if we restrict the transformation matrix P to be a **unitary matrix** U? How simple can we make the matrix U^*AU (where U^* is the conjugate transpose of U, and $U^* = U^{-1}$)? This is known as a **unitary similarity transformation**.

Unitary matrices are particularly appealing because they represent rigid transformations (like rotations and reflections) that preserve lengths and angles, making them the "nicest" kind of invertible matrices. This chapter focuses on answering the second question. In terms of linear transformations, the first question asks how simple we can make the standard matrix of a linear transformation by choosing an arbitrary basis. The second question asks how simple we can make the standard matrix if we are restricted to using only **orthonormal bases**. Schur Triangularization provides the answer to this second question.

Schur Triangularization Theorem

Theorem 0.1 (Schur Triangularization)**.** Let A be any square matrix with complex entries, i.e., $A \in M_n(\mathbb{C})$. Then there exists a complex unitary matrix $U \in M_n(\mathbb{C})$ and a complex upper triangular matrix $T \in M_n(\mathbb{C})$ such that

$$A = UTU^*$$

Equivalently, $T = U^*AU$.

Remark 0.2. This theorem guarantees that *any* square matrix with complex entries can be transformed into an upper triangular matrix via a unitary similarity transformation. This is a powerful result because, unlike diagonalization, it applies to all square complex matrices, regardless of whether they have a full set of eigenvectors. Real matrices are included since real numbers are a subset of complex numbers, although the resulting U and T matrices might have complex entries even if A is real.

Proof of Schur Triangularization (by Induction)

We prove the theorem by induction on the size n of the matrix A. **Base Case (n=1):** If $n = 1$, $A = [a_{11}]$ is a 1×1 matrix. A 1×1 matrix is trivially upper triangular. We can choose $U = [1]$ (which is unitary) and $T = A = [a_{11}]$. Then $A = UTU^*$ holds since $1 \cdot a_{11} \cdot 1^* = a_{11}$. **Inductive Hypothesis:** Assume that the theorem holds for all $(n-1) \times (n-1)$ complex matrices. That is, for any $B \in M_{n-1}(\mathbb{C})$, there exist a unitary $\tilde{U} \in M_{n-1}(\mathbb{C})$ and an upper triangular $\tilde{T} \in M_{n-1}(\mathbb{C})$ such that $B = \tilde{U}\tilde{T}\tilde{U}^*$. **Inductive Step:** Let $A \in M_n(\mathbb{C})$. Since A is a complex matrix, the Fundamental Theorem of Algebra guarantees that its characteristic polynomial, $\det(A - \lambda I)$, has at least one root in $\mathbb{C}$. This root is an eigenvalue of A. Let λ be such an eigenvalue, and let $\vec{v}$ be a corresponding eigenvector. We can normalize the eigenvector $\vec{v}$ so that it has unit length, i.e., $||\vec{v}|| = \sqrt{\vec{v}^*\vec{v}} = 1$. Now, we can extend the single unit vector $\{\vec{v}\}$ to an orthonormal basis for $\mathbb{C}^n$. This can be done using the Gram-Schmidt process: start with $\{\vec{v}\}$ and add $n-1$ other vectors to form a basis for $\mathbb{C}^n$, then apply Gram-Schmidt to obtain an orthonormal basis $\{\vec{v}, \vec{v}_2, \ldots, \vec{v}_n\}$. Let U be the matrix whose columns are these orthonormal basis vectors, with $\vec{v}$ as the first column. Let V be the matrix whose columns are $\vec{v}_2, \ldots, \vec{v}_n$. Then U can be written in block form as:

$$U = [\vec{v}|V] \in M_n(\mathbb{C})$$

where $V \in M_{n,n-1}(\mathbb{C})$. Since the columns of U form an orthonormal basis, U is a unitary matrix ($U^*U = I$). Now, consider the unitary similarity transformation U^*AU. We compute this product using block matrices: The conjugate transpose of U is $U^* = \left[\begin{array}{c} \vec{v}^* \\ \hline V^* \end{array}\right]$.

$$\begin{aligned} U^*AU &= \left[\begin{array}{c} \vec{v}^* \\ \hline V^* \end{array}\right] A[\vec{v}|V] \\ &= \left[\begin{array}{c} \vec{v}^* \\ \hline V^* \end{array}\right] [A\vec{v}|AV] \end{aligned}$$

Since $\vec{v}$ is an eigenvector of A with eigenvalue λ, we have $A\vec{v} = \lambda\vec{v}$. Substituting this in:

$$\begin{aligned} U^*AU &= \left[\begin{array}{c} \vec{v}^* \\ \hline V^* \end{array}\right] [\lambda\vec{v}|AV] \\ &= \left[\begin{array}{cc} \vec{v}^*(\lambda\vec{v}) & \vec{v}^*AV \\ \hline V^*(\lambda\vec{v}) & V^*AV \end{array}\right] \\ &= \left[\begin{array}{cc} \lambda(\vec{v}^*\vec{v}) & \vec{v}^*AV \\ \hline \lambda(V^*\vec{v}) & V^*AV \end{array}\right] \end{aligned}$$

Since U is unitary, its columns are orthonormal. Thus, $\vec{v}^*\vec{v} = ||\vec{v}||^2 = 1$. Also, the columns of V ($\vec{v}_2, \ldots, \vec{v}_n$) are orthogonal to $\vec{v}$. This means that each row of V^* (which are the conjugate transposes of the columns of V) is orthogonal to $\vec{v}$. Therefore, $V^*\vec{v} = \vec{0}$, where $\vec{0}$ is the zero vector of size $(n-1) \times 1$. Substituting these results back:

$$U^*AU = \left[\begin{array}{cc} \lambda & \vec{v}^*AV \\ \hline \vec{0} & V^*AV \end{array}\right]$$

Let $A_{n-1} = V^*AV$. This is an $(n-1) \times (n-1)$ complex matrix. By the inductive hypothesis, there exists an $(n-1) \times (n-1)$ unitary matrix $\tilde{U}$ and an $(n-1) \times (n-1)$ upper triangular matrix $\tilde{T}$ such that $A_{n-1} = V^*AV = \tilde{U}\tilde{T}\tilde{U}^*$. Now, define a new $n \times n$ block matrix $\tilde{U}_{block}$:

$$\tilde{U}_{block} = \left[\begin{array}{cc} 1 & \vec{0}^* \\ \hline \vec{0} & \tilde{U} \end{array}\right]$$

This matrix is unitary because $\tilde{U}^*_{block}\tilde{U}_{block} = \left[\begin{array}{cc} 1 & \vec{0}^* \\ \hline \vec{0} & \tilde{U}^* \end{array}\right] \left[\begin{array}{cc} 1 & \vec{0}^* \\ \hline \vec{0} & \tilde{U} \end{array}\right] = \left[\begin{array}{cc} 1 & \vec{0}^* \\ \hline \vec{0} & \tilde{U}^*\tilde{U} \end{array}\right] = \left[\begin{array}{cc} 1 & \vec{0}^* \\ \hline \vec{0} & I_{n-1} \end{array}\right] = I_n$. Consider the product $(U\tilde{U}_{block})^*A(U\tilde{U}_{block})$. Let $W = U\tilde{U}_{block}$. Since U and $\tilde{U}_{block}$ are unitary, their product W is also unitary.

$$
\begin{aligned}
W^*AW &= (U\tilde{U}_{block})^*A(U\tilde{U}_{block}) \\
&= \tilde{U}_{block}^*(U^*AU)\tilde{U}_{block} \\
&= \left[\begin{array}{c|c} 1 & \vec{0}^* \\ \hline \vec{0} & \tilde{U}^* \end{array}\right] \left[\begin{array}{c|c} \lambda & \vec{v}^*AV \\ \hline \vec{0} & V^*AV \end{array}\right] \left[\begin{array}{c|c} 1 & \vec{0}^* \\ \hline \vec{0} & \tilde{U} \end{array}\right] \\
&= \left[\begin{array}{c|c} 1 & \vec{0}^* \\ \hline \vec{0} & \tilde{U}^* \end{array}\right] \left[\begin{array}{c|c} \lambda & (\vec{v}^*AV)\tilde{U} \\ \hline \vec{0} & (V^*AV)\tilde{U} \end{array}\right] \\
&= \left[\begin{array}{c|c} \lambda & (\vec{v}^*AV)\tilde{U} \\ \hline \tilde{U}^*(\vec{0}) & \tilde{U}^*(V^*AV)\tilde{U} \end{array}\right] \\
&= \left[\begin{array}{c|c} \lambda & (\vec{v}^*AV)\tilde{U} \\ \hline \vec{0} & \tilde{U}^*(\tilde{U}\tilde{T}\tilde{U}^*)\tilde{U} \end{array}\right] \\
&= \left[\begin{array}{c|c} \lambda & (\vec{v}^*AV)\tilde{U} \\ \hline \vec{0} & (\tilde{U}^*\tilde{U})\tilde{T}(\tilde{U}^*\tilde{U}) \end{array}\right] \\
&= \left[\begin{array}{c|c} \lambda & (\vec{v}^*AV)\tilde{U} \\ \hline \vec{0} & \tilde{T} \end{array}\right]
\end{aligned}
$$

Let $T = \left[\begin{array}{c|c} \lambda & (\vec{v}^*AV)\tilde{U} \\ \hline \vec{0} & \tilde{T} \end{array}\right]$. Since $\tilde{T}$ is upper triangular, the entire matrix T is upper triangular. We have found a unitary matrix $W = U\tilde{U}_{block}$ and an upper triangular matrix T such that $W^*AW = T$. Rearranging gives $A = WTW^*$. This completes the inductive step. Therefore, by induction, Schur Triangularization holds for all $n \geq 1$.

Notes on Schur Triangularization

1. **Eigenvalues on the Diagonal:** A crucial property is that the diagonal entries of the upper triangular matrix T in a Schur decomposition $A = UTU^*$ are precisely the eigenvalues of the original matrix A. To see why, first recall a theorem from introductory linear algebra: the eigenvalues of any triangular matrix (upper or lower) are exactly its diagonal entries. Let $T = [t_{ij}]$. Since T is upper triangular, its characteristic polynomial is $\det(T-\lambda I) = (t_{11}-\lambda)(t_{22}-\lambda)\dots(t_{nn}-\lambda)$. The roots are $t_{11}, t_{22}, \dots, t_{nn}$. Second, we show that A and $T = U^*AU$ have the same eigenvalues because they have the same characteristic polynomial.

$$
\begin{aligned}
\det(T - \lambda I) &= \det(U^*AU - \lambda I) \\
&= \det(U^*AU - \lambda U^*U) \quad (\text{since } U^*U = I) \\
&= \det(U^*(A - \lambda I)U) \\
&= \det(U^*)\det(A - \lambda I)\det(U) \quad (\text{since determinant is multiplicative})
\end{aligned}
$$

Since U is unitary, $U^* = U^{-1}$. We know $\det(U^*) = \det(U^{-1}) = 1/\det(U)$. Therefore,

$$
\det(T - \lambda I) = \frac{1}{\det(U)}\det(A - \lambda I)\det(U) = \det(A - \lambda I)
$$

Since A and T have the same characteristic polynomial, they must have the same eigenvalues. Because the eigenvalues of the upper triangular matrix T are its diagonal entries $t_{11}, \dots, t_{nn}$, these must also be the eigenvalues of A.

2. **Non-Uniqueness:** The Schur decomposition $A = UTU^*$ is *not* unique.

   - The diagonal entries of T (the eigenvalues of A) can appear in any order. This order depends on the choice of eigenvalue and eigenvector at each step of the inductive construction.

- The off-diagonal entries of T are highly non-unique. They depend on the choices made during the Gram-Schmidt extensions at each step.
- The unitary matrix U is also highly non-unique, again depending on the choices made during the construction (which eigenvalue/eigenvector to start with, and how the Gram-Schmidt process is completed).

Different valid Schur triangularizations of the same matrix A can have very different looking U and T matrices (though the diagonal entries of T will always be the same set of eigenvalues, possibly in a different order).

3. **Theoretical Tool vs. Computation:** While the proof provides a constructive method, computing a Schur triangularization step-by-step as in the proof (finding eigenvalues/eigenvectors, performing Gram-Schmidt repeatedly) is computationally expensive and numerically sensitive. In practice, other algorithms (like the QR algorithm) are used to compute Schur decompositions numerically. However, the *existence* guaranteed by the Schur Triangularization theorem is an extremely powerful theoretical tool for proving other results in linear algebra, precisely because it applies to *all* square complex matrices.

Applications of Schur Triangularization

Schur Triangularization allows for elegant proofs of several important properties relating matrix entries to eigenvalues.

Proposition 0.3. Let $A \in M_n(\mathbb{C})$ be a square complex matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ (repeated according to algebraic multiplicity). Then:

1. **Determinant:** $\det(A) = \lambda_1 \lambda_2 \ldots \lambda_n$ (The determinant is the product of the eigenvalues).
2. **Trace:** $\operatorname{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$ (The trace is the sum of the eigenvalues).

Preuve. Let $A = UTU^*$ be a Schur triangularization of A, where U is unitary and T is upper triangular. As established in the notes, the diagonal entries of T are the eigenvalues of A, so $t_{ii} = \lambda_i$ (possibly after reordering).

1. **Determinant:**

$$\begin{aligned}
\det(A) &= \det(UTU^*) \\
&= \det(U)\det(T)\det(U^*) \\
&= \det(U)\det(T)\frac{1}{\det(U)} \quad (\text{since } \det(U^*) = 1/\det(U)) \\
&= \det(T)
\end{aligned}$$

Since T is upper triangular, its determinant is the product of its diagonal entries:

$$\det(T) = t_{11}t_{22}\ldots t_{nn} = \lambda_1\lambda_2\ldots\lambda_n$$

Therefore, $\det(A) = \lambda_1\lambda_2\ldots\lambda_n$.

2. **Trace:** Recall that the trace is invariant under cyclic permutations: $\operatorname{tr}(XYZ) = \operatorname{tr}(ZXY) =$

$\text{tr}(YZX)$.

$$\begin{aligned}\text{tr}(A) &= \text{tr}(UTU^*) \\ &= \text{tr}(U^*UT) \quad \text{(by cyclic commutativity)} \\ &= \text{tr}(IT) \quad (\text{since } U^*U = I) \\ &= \text{tr}(T)\end{aligned}$$

The trace of a matrix is the sum of its diagonal entries:

$$\text{tr}(T) = t_{11} + t_{22} + \cdots + t_{nn} = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

Therefore, $\text{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

□

Conclusion

Schur Triangularization is a fundamental result in linear algebra, asserting that any square complex matrix can be brought to upper triangular form via a unitary similarity transformation. While not always practical for direct computation, its guaranteed existence for all such matrices makes it an invaluable theoretical tool. It provides a bridge between the properties of a general matrix and the simpler, well-understood properties of triangular matrices, enabling elegant proofs for results concerning eigenvalues, determinants, and traces, among others. Many advanced theorems and concepts build upon the foundation laid by Schur Triangularization.