Similarity Transformations

As mentioned in class, if \( A \in \mathbb{R}^{n \times n} \) and \( B \in \mathbb{R}^{n \times n} \) are real square matrices related by
\[
B = P^{-1}AP,
\]
where \( P \in \mathbb{R}^{n \times n} \) is a nonsingular matrix, then \( A \) and \( B \) are said to be similar. Since
\[
\det(Is - A) = \det(P^{-1}) \det(Is - A) \det(P) = \det(Is - B),
\]
(1)
it follows that similar matrices have the same characteristic equation. If \( A \) is a real symmetric matrix, then its eigenvalues are all real and the eigenvectors of \( A \) can be taken to be orthogonal. If we choose \( P \) to be a matrix whose columns are the orthonormal eigenvectors of \( A \) (which is clearly nonsingular since the eigenvectors are linearly independent), then the matrix \( B = P^{-1}AP \) is both real and diagonal. Hence we have

**Theorem 1 (Similarity Transformation for Real Symmetric Matrix)** Any real symmetric matrix \( A \in \mathbb{R}^{n \times n} \) can be diagonalized by a similarity transformation, i.e., \( A = P^{-1} \Lambda P \) for some invertible \( P \in \mathbb{R}^{n \times n} \) and \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n) \) where \( \lambda_i \)'s are the real eigenvalues of \( A \).

Such a simplification (reduction) however is not generally true for non-symmetric real matrices. Two situations may occur instead: 1. \( A \) has complex eigenvalues (even if it can be diagonalized); 2. \( A \) simply cannot be diagonalized at all. In view of this, a relevant question is: To what extent can an arbitrary real square matrix be simplified by a real or complex similarity transformation?

First, we consider the case that we only allow \( P \) to be a real matrix in \( \mathbb{R}^{n \times n} \). The answer to this is provided in the following theorem, whose proof is quite beyond the scope of this course.

**Theorem 2 (Real Similarity Transformation for Real Matrix)** If \( A \in \mathbb{R}^{n \times n} \) is a real square matrix, there exists a real nonsingular matrix \( P \in \mathbb{R}^{n \times n} \) such that \( B = P^{-1}AP \) has the form
\[
B = \begin{bmatrix}
B_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & B_2 & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & B_k & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & B_{k+1} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & B_{k+2} & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & B_l
\end{bmatrix}
\]
(2)
where \( B_1, \ldots, B_k \) are in the form: for \( 1 \leq i \leq k \),
\[
B_i = \begin{bmatrix}
S_i & I_{2 \times 2} & 0 & \cdots & 0 \\
0 & S_i & I_{2 \times 2} & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & S_i & I_{2 \times 2} \\
0 & 0 & \cdots & 0 & S_i
\end{bmatrix}, \quad S_i = \begin{bmatrix}
\sigma_i & \omega_i \\
-\omega_i & \sigma_i
\end{bmatrix}
\]
(3)
and \( B_{k+1}, \ldots, B_l \) are in the form: for \( k + 1 \leq j \leq l \),

\[
B_j = \begin{bmatrix}
s_j & 1 & 0 & \cdots & 0 \\
0 & s_j & 1 & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & s_j & 1 \\
0 & 0 & \cdots & 0 & s_j
\end{bmatrix}, \quad s_j \in \mathbb{R}.
\]

Second, we allow \( P \) to be a (possibly) complex matrix in \( \mathbb{C}^{n \times n} \). Note that since now the space of all possible similarity transformations is enlarged, the resulting reduced form of \( A \) can be simpler than that given in Theorem 2.

**Theorem 3 ((Complex) Similarity Transformation for Real Matrix)** If \( A \in \mathbb{R}^{n \times n} \) is a real square matrix, there exists a possibly complex nonsingular matrix \( P \in \mathbb{C}^{n \times n} \) such that \( B = P^{-1}AP \) has the form

\[
B = \begin{bmatrix}
B_1 & 0 & \cdots & 0 & 0 \\
0 & B_2 & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & B_{m-1} & 0 \\
0 & 0 & \cdots & 0 & B_m
\end{bmatrix}
\]

where \( B_1, \ldots, B_m \) are in the form: for \( 1 \leq i \leq m \),

\[
B_i = \begin{bmatrix}
s_i & 1 & 0 & \cdots & 0 \\
0 & s_i & 1 & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & s_i & 1 \\
0 & 0 & \cdots & 0 & s_i
\end{bmatrix}, \quad s_i \in \mathbb{C}.
\]

In Theorem 3, the matrix \( B \) is called the *Jordan form* of matrix \( A \), and \( s_i \)'s are simply all the (possibly complex) eigenvalues of \( A \). Note the same statement holds even for the case when the matrix \( A \) itself is complex. The proof of Theorem 3 is relatively easy and can be found in most classic books on linear algebra. One useful application of the Jordan form is to find (or define) the so called *minimal polynomial* of \( A \), i.e., the polynomial \( \Psi(s) \) of the lowest degree such that \( \Psi(A) = 0_{n \times n} \). Another application of the Jordan form is to prove the so called *Cayley-Hamilton* theorem.

**Theorem 4 (Cayley-Hamilton)** For any square matrix \( A \in \mathbb{C}^{n \times n} \), let \( \chi(s) = \det(sI - A) \) be the characteristic polynomial of \( A \). Then we always have \( \chi(A) = 0_{n \times n} \).

Using the Jordan form of \( A \) from the previous theorem, the proof is straightforward.