Topics:

a) Determinant of a matrix, column operations, Laplace’s expansion.

b) Characteristic polynomial, Cayley-Hamilton theorem, eigenvalues/eigenvectors.

c) Diagonalizable matrices, Jordan form, minimal polynomial.

Determinant: To motivate the concept of determinant of a square matrix $A$, consider the case where $A$ is a $2 \times 2$ matrix with columns $a_1, a_2 \in \mathbb{R}^2$. If we view $\mathbb{R}^2$ as embedded in the three-dimensional space $\mathbb{R}^3$, the outer product

$$a_1 \times a_2 = \det(a_1, a_2)e_3$$

(1)

where $e_3$ is the unit vector along the third axis and

$$\det(a_1, a_2) = |a_1||a_2|\sin(\theta)$$

(2)

measures the area of the parallelogram spanned by $a_1$ and $a_2$ as shown in Fig.1 below.

Figure 1: Interpretation of the determinant in two dimensions as the oriented area of the parallelogram spanned by $a_1$ and $a_2$.

In expression (2),

$$|a_i| = (a_i^T a_i)^{1/2}$$

denotes the length (Euclidean norm) of vector $a_i$ for $i = 1, 2$ and $\theta$ is the oriented angle going from vector $a_1$ to vector $a_2$, so that $\det(a_1, a_2)$ is an oriented area in the sense that

$$\det(a_2, a_1) = -\det(a_1, a_2),$$
since when \( \mathbf{a}_1 \) and \( \mathbf{a}_2 \) are interchanged, \( \theta \) becomes \( -\theta \), and \( \sin(-\theta) = -\sin \theta \).

From the above definition, we see that \( \det(\mathbf{a}_1, \mathbf{a}_2) = 0 \) whenever \( \sin \theta = 0 \), i.e. for \( \theta = 0, \pi \). Thus, the determinant of vectors \( \mathbf{a}_1 \) and \( \mathbf{a}_2 \) is zero whenever they are colinear.

For the \( 2 \times 2 \) case, if
\[
A = \begin{bmatrix} a_1 & a_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},
\]
an analytical expression equivalent to (2) is given by
\[
\det A = \det(\mathbf{a}_1, \mathbf{a}_2) = a_{11}a_{22} - a_{21}a_{12}.
\]

Consider now an \( n \times n \) matrix \( A \) with columns \( \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \). Based on the above motivation, we say that \( \det(\mathbf{a}_1, \ldots, \mathbf{a}_n) \) is a measure of the oriented volume of the parallelepiped of \( \mathbb{R}^n \) spanned by vectors \( \mathbf{a}_1, \ldots, \mathbf{a}_n \). Thus, \( \det(\mathbf{a}_1, \ldots, \mathbf{a}_n) \) is a map from \( (\mathbb{R}^n)^n \) to \( \mathbb{R} \):
\[
(\mathbf{a}_1, \ldots, \mathbf{a}_n) \in (\mathbb{R}^n)^n \rightarrow \det(\mathbf{a}_1, \ldots, \mathbf{a}_n) \in \mathbb{R}
\]
which has the following properties:

(i) It is linear with respect to each vector \( \mathbf{a}_i \), \( 1 \leq i \leq n \) taken separately. Thus, if \( \mathbf{x} \) and \( \mathbf{y} \) are two arbitrary vectors of \( \mathbb{R}^n \) and if \( \mathbf{u} \) and \( \mathbf{v} \) are arbitrary real numbers
\[
\det(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, u\mathbf{x} + v\mathbf{y}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n) = u \det(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{x}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n) + v \det(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{y}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n).
\]

(ii) If vectors \( \mathbf{a}_i \) and \( \mathbf{a}_j \) with \( i < j \) are interchanged, we have
\[
\det(\mathbf{a}_1, \ldots, \mathbf{a}_i, \mathbf{a}_j, \ldots, \mathbf{a}_n) = -\det(\mathbf{a}_1, \ldots, \mathbf{a}_j, \mathbf{a}_i, \mathbf{a}_n)
\]

(iii) If there exists a nontrivial linear dependence relation
\[
\sum_{i=1}^{n} u_i \mathbf{x}_i = 0 \quad \text{with} \quad \mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \neq 0 \tag{5}
\]
between vectors \( \mathbf{a}_1, \ldots, \mathbf{a}_n \), then
\[
\det(\mathbf{a}_1, \ldots, \mathbf{a}_n) = 0.
\]

This is due to the fact that when (5) is satisfied, the vectors \( \mathbf{a}_1, \ldots, \mathbf{a}_n \) belong to a lower dimensional hyperplane of \( \mathbb{R}^n \) so that the volume of the parallelepiped spanned by \( \mathbf{a}_1, \ldots, \mathbf{a}_n \) is zero.

The above three axioms can be used to derive the following analytical expression for the determinant of \( A = (a_{ij}) \), \( 1 \leq i, j \leq n \) in terms of its entries:
\[
\det A = \sum_{\pi} (-1)^{t(\pi)} a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)}, \tag{6}
\]
where in (6) the sum is over all permutations \( \pi \) of the index set \( \{1, 2, \ldots, n\} \), and \( t(\pi) \) is the number of transpositions occurring in the permutation \( \sigma \). For example, if we consider the permutation
\[
\{1, 2, 3, 4, 5\} \xrightarrow{\pi} \{2, 4, 3, 1, 5\}
\]
\( t(\pi) \) can be computed by observing that
-2 occurs before 1
-4 occurs before 1 and 3
-3 occurs before 1
so that \( t(\pi) = 4 \).

The expression (6) shows that \( \det A \) is obtained by performing all the products of \( n \) entries of \( A \) such that one element of each row and one of each column appears in the product. In the \( 3 \times 3 \) case, with
\[
A = \begin{bmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
a_{31} & a_{32} & a_{33}
\end{bmatrix},
\]
this leads to the usual expression
\[
\det A = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.
\]

**Elementary column operations**: The three axioms of determinants can be used to characterize the effect of elementary column operations on matrix determinants.

(i) Multiplication of a column by \( c \neq 0 \). Let
\[
A = \begin{bmatrix}
a_1 & \ldots & a_i & \ldots & a_n
\end{bmatrix} \longrightarrow A_s = \begin{bmatrix}
a_1 & \ldots & ca_i & \ldots & a_n
\end{bmatrix}.
\]
Then, according to (3)
\[
\det A_s = c \det A.
\]

(ii) Exchange of two columns. If
\[
A = \begin{bmatrix}
a_1 & \ldots & a_i & \ldots & a_j & \ldots & a_n
\end{bmatrix} \longrightarrow A_p = \begin{bmatrix}
a_1 & \ldots & a_j & \ldots & a_i & \ldots & a_n
\end{bmatrix},
\]
then \( \det A_p = -\det A \).

(iii) Adding to a column a multiple of another column. Let
\[
A = \begin{bmatrix}
a_1 & \ldots & a_i & \ldots & a_j & \ldots & a_n
\end{bmatrix} \longrightarrow A_c = \begin{bmatrix}
a_1 & \ldots & a_i + va_j & \ldots & a_j & \ldots & a_n
\end{bmatrix},
\]
with \( v \in \mathbb{R} \). The multilinearity property (3) implies
\[
\det A_c = \det A + v \det B
\]
where the matrix
\[
B = \begin{bmatrix}
a_1 & \ldots & a_j & \ldots & a_j & \ldots & a_n
\end{bmatrix}
\]
has two identical columns, so that according to the third axiom of determinants \( \det B = 0 \). Hence we conclude that \( \det A_c = \det A \), so that an elementary linear combination of columns does not affect the determinant.
Since we can always use elementary column operations to reduce an arbitrary square matrix $A$ to a lower triangular matrix $L$ whose determinant is the product of its diagonal elements, the following strategy can be employed to evaluate determinants.

**Step 1:** Use elementary column (resp. row) operations to reduce $A$ to a lower (resp. upper) triangular matrix $L$ (resp. $U$), while keeping track of the effect of the elementary operations on the determinant of $A$.

**Step 2:** Evaluate the determinant of $L$ (resp. $U$).

**Example:** We have

$$A = \begin{bmatrix}
2 & -1 & 0 & 0 \\
-1 & 2 & -1 & 0 \\
0 & -1 & 2 & -1 \\
0 & 0 & -1 & 2
\end{bmatrix} \quad \rightarrow \quad \begin{bmatrix}
2 & 0 & 0 & 0 \\
-1 & 3/2 & -1 & 0 \\
0 & -1 & 2 & -1 \\
0 & 0 & -1 & 2
\end{bmatrix} \quad \rightarrow \quad \begin{bmatrix}
2 & 0 & 0 & 0 \\
0 & 0 & -1 & 2 \\
0 & 0 & -1 & 5/4
\end{bmatrix} = L,$$

where elementary column operations are used to bring $A$ to the lower triangular form $L$. Specifically, on the first line, we multiply column 1 by $1/2$ and add it to column 2. Then on the second line we multiply column 2 by $2/3$ and add it to column 3. Finally, we multiply column 3 by $3/4$ and add it to column 4. Since $L$ is lower triangular

$$\det A = \det L = 2 \times \frac{3}{2} \times \frac{4}{3} \times \frac{5}{4} = 5.$$  

**Laplace’s expansion:** Another useful result is that the determinant of an $n \times n$ matrix $A$ can be expanded in terms of the entries of row $i$ as

$$\det A = \sum_{j=1}^{n} a_{ij} C_{ij} \quad (7)$$

where the cofactor $C_{ij}$ of the $(i,j)$-th element $a_{ij}$ of $A$ is given by

$$C_{ij} = (-1)^{i+j} \det A_{ij}, \quad (8)$$

where $A_{ij}$ is the $(n-1) \times (n-1)$ matrix obtained by deleting the $i$-th row and $j$-th column of $A$. In (7) the choice of row $i$ is arbitrary. $\det A$ admits also a similar expression in terms of the entries of column $j$. The above formula is particularly convenient if $A$ contains rows or columns with many zero entries.
Example: Consider the \( n \times n \) tridiagonal matrix.

\[
A_n = \begin{bmatrix}
2 & -1 & 0 \\
-1 & 2 & -1 \\
& \ddots & \ddots & \ddots \\
& & 0 & -1 & 2 & -1 \\
& & & & -1 & 2 \\
& & & & & -1 & 2
\end{bmatrix}.
\]

If \( D_n = \det A_n \), by expanding \( D_n \) with respect to the first row of \( A_n \) we find

\[
D_n = 2D_{n-1} + \det \begin{bmatrix}
-1 & -1 & 0 \\
0 & 2 & -1 & 0 \\
& \ddots & \ddots & \ddots \\
& & & -1 & 2 & -1 \\
& & & & 0 & \ddots \\
& & & & & 0 & -1 & 2
\end{bmatrix}
\]

\[
= 2D_{n-1} - D_{n-2},
\]

with \( D_1 = 2, D_2 = 3 \). This yields \( D_n = n + 1 \) and for \( n = 4 \) we obtain \( D_4 = 5 \), which is exactly the result obtained in the first example on page 4.

Properties of determinants:

(i) If \( A \) and \( B \) are two square matrices of equal size, we have

\[
\det AB = \det BA = \det A \det B.
\]

(ii) Applying the above identity for \( B = A^{-1} \), we find

\[
\det A^{-1} = 1/ \det A.
\]

(iii) \( \det A^T = \det A \).

(iv) Laplace's expansion of the determinant of \( A \) can be written in matrix form as

\[
A \hat{A} = (\det A) I_n,
\]

where \( \hat{A} \) is the adjugate matrix of \( A \). \( \hat{A} = (a_{ij}, 1 \leq i, j \leq n) \) is the transpose of the matrix formed by the cofactors of \( A \), i.e., \( \hat{a}_{ij} = C_{ji} \) for all \( i \) and \( j \).

Characteristic polynomial: Let \( A \) be an \( n \times n \) matrix. Then \( a(s) = \det(sI - A) \) is the characteristic polynomial of \( A \). Using Laplace's formula to expand

\[
a(s) = \det(sI - A) = \det \begin{bmatrix}
s - a_{11} & -a_{12} & \cdots & -a_{1n} \\
-a_{21} & s - a_{22} & \cdots & -a_{2n} \\
& \ddots & \ddots & \ddots \\
& & -a_{n1} & \cdots & s - a_{nn}
\end{bmatrix},
\]

5
we see that \( a(s) \) is a polynomial of degree \( n \) where the coefficient of \( s^n \) equals 1. Thus
\[
a(s) = s^n + a_1 s^{n-1} + \cdots + a_n = \prod_{i=1}^{k} (s - \lambda_i)^{n_i},
\]
where the multiplicities \( n_i \) of roots \( \lambda_i \) with \( 1 \leq i \leq k \) satisfy
\[
\sum_{i=1}^{k} n_i = n.
\]

**Cayley-Hamilton theorem**: An important property of matrix \( A \) is that it annihilates its characteristic polynomial, i.e.,
\[
a(A) = A^n + a_1 A^{n-1} + a_2 A^{n-2} + \cdots + a_n I_n = 0. \tag{9}
\]
This important result can be established by using the identity
\[
(sI - A)\overline{(sI - A)} = a(s)I_n \tag{10}
\]
and noting that the adjugate matrix \( \overline{sI - A} \) is a matrix polynomial of degree \( n - 1 \), so it can be written as
\[
\overline{sI - A} = R_1 s^{n-1} + R_2 s^{n-2} + \cdots + R_n. \tag{11}
\]
Substituting (11) in (10) and identifying successive coefficients of \( s^i \) with \( 0 \leq i \leq n \) in decreasing order on both sides of (10)
\[
\begin{align*}
    s^n & : R_1 = I \\
    s^{n-1} & : R_2 - AR_1 = a_1 I \\
    \cdots & \\
    s^0 & : -AR_n = a_n I,
\end{align*}
\]
and progressively eliminating \( R_1, R_2, \ldots, R_n \) from the above relations yields (9).

**Eigenvalues and eigenvectors**: \( \lambda \in \mathbb{C} \) is an eigenvalue of \( A \) and \( \mathbf{x} \in \mathbb{C}^n \) is a right eigenvector associated with it if
\[
A \mathbf{x} = \lambda \mathbf{x}
\]
with \( \mathbf{x} \neq 0 \).

Since \( (\lambda I - A) \mathbf{x} = 0 \), the matrix \( \lambda I - A \) is singular so that \( a(\lambda) = \det(\lambda I - A) = 0 \), i.e., \( \lambda \) is one of the roots \( \lambda_1, \ldots, \lambda_k \) of \( a(s) \). The eigenvalues of \( A \) can be complex, but since the coefficients \( a_{ij} \) of \( A \) are real, if \( \lambda_i \) is an eigenvalue of \( A \) with multiplicity \( n_i \), \( \lambda_i^* \) is also an eigenvalue with the same multiplicity. The eigenvalues of \( A \) are therefore symmetric with respect to the real axis, as depicted in Fig.2 below.

The right eigenvectors of \( A \) associated with distinct eigenvalues have the following property.
Lemma 1: If \( \{ \mathbf{x}_i, 1 \leq i \leq k \} \) are right eigenvectors of \( A \) associated with eigenvalues \( \{ \lambda_i, 1 \leq i \leq k \} \) such that \( \lambda_i \neq \lambda_j \) for \( i \neq j \), they are linearly independent.

Proof: Suppose there exists a nontrivial linear dependence relation between the vectors \( \mathbf{x}_i \), so that

\[
\sum_{i=1}^{k} u_i \mathbf{x}_i = 0 ,
\]

where at least one of the coefficients \( u_i \) is different from zero, say \( u_1 \neq 0 \). Then, multiplying (12) on the left by \( \prod_{i=2}^{k} (A - \lambda_i I) \) and observing that \( A - \lambda_i I \) and \( A - \lambda_j I \) commute, we find

\[
\prod_{i=2}^{k} (A - \lambda_i I) \left( \sum_{j=1}^{k} u_j \mathbf{x}_j \right) = u_1 \prod_{i=2}^{k} (\lambda_1 - \lambda_i) \mathbf{x}_1 = 0 ,
\]

which is a contradiction since \( u_1 \neq 0 \), \( \prod_{i=2}^{k} (\lambda_1 - \lambda_i) \neq 0 \) and \( \mathbf{x}_1 \neq 0 \). Thus, the vectors \( \{ \mathbf{x}_i, 1 \leq i \leq k \} \) must be linearly independent. \( \square \)

The eigenstructure of \( A \) is particularly simple when it has \( n \) distinct eigenvalues \( \lambda_1, \ldots, \lambda_n \). In this case, as shown above, the corresponding right eigenvectors \( \{ \mathbf{x}_i, 1 \leq i \leq n \} \) are linearly independent and form a basis of \( \mathbb{R}^n \). The relations \( A \mathbf{x}_i = \lambda_i \mathbf{x}_i, 1 \leq i \leq n \) can be written in matrix form as

\[
AX = \Lambda X
\]

with

\[
X = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}
\]

and

\[
\Lambda = \text{diag} (\lambda_1, \lambda_2, \ldots, \lambda_n) .
\]

Since the columns of \( X \) form a basis of \( \mathbb{R}^n \), \( X \) is invertible so that

\[
A = X \Lambda X^{-1} .
\]

This shows that \( A \) is related to the diagonal matrix \( \Lambda \) through a similarity transformation. In this context, \( A \) is said to be diagonalizable. In this context, it is useful to observe that
Lemma 2: If $A$ and $B = TAT^{-1}$ are related through an invertible similarity transformation $T$, they have the same characteristic polynomial.

Proof:

$$\det(sI - B) = \det \left( T(sI - A)T^{-1} \right)$$

$$= \det T \det(sI - A) \det T^{-1} = \det(sI - A),$$

where the last equality was obtained by using $\det T^{-1} = 1/\det T$. \hfill \Box

When $A$ does not have distinct eigenvalues, it still may be possible to diagonalize it. This depends on whether for each eigenvalue $\lambda_i$ with multiplicity $n_i$, we can find $n_i$ independent eigenvectors $x_{i\ell}$, $1 \leq \ell \leq n_i$ associated to $\lambda_i$. When this is the case, the identity (13) remains valid with

$$X = \begin{bmatrix} X_1 & X_2 & \cdots & X_k \end{bmatrix},$$

$$\Lambda = \text{diag} \left( D_1, D_2, \ldots, D_k \right),$$

where

$$X_i = \begin{bmatrix} x_{i1} & \cdots & x_{i\ell} & \cdots & x_{im_i} \end{bmatrix} \quad \text{and} \quad D_i = \lambda_i I_{n_i}.$$ 

The columns of $X$ are still linearly independent. To see this, assume that there exists a linear dependence relation

$$\sum_{i=1}^{k} \left( \sum_{\ell=1}^{n_i} u_{i\ell} x_{i\ell} \right) = 0$$

(15)

between the columns of $X$. Let

$$x_i \triangleq \sum_{\ell=1}^{n_i} u_{i\ell} x_{i\ell}.$$ 

Depending on whether the coefficients $u_{i\ell}$, $1 \leq \ell \leq n$ are all zero or not, $x_i$ is either the zero vector or an eigenvector of $A$ associated with eigenvalue $\lambda_i$ (it is a linear combination of such eigenvectors). If $x_i \neq 0$ for at least one $i$, the relation (15) indicates that there exists a linear dependence relation between several eigenvectors of $A$ associated to distinct eigenvalues $\lambda_i$, $1 \leq i \leq k$. According to Lemma 1, this is impossible, so that we must have $x_i = 0$ for all $i$. But for each $i$, the eigenvectors $x_{i\ell}$ with $1 \leq \ell \leq n_i$ are linearly independent, so we must have $u_{i\ell} = 0$ for all $i$ and $\ell$. Thus the columns of $X$ are linearly independent, so that $X$ is invertible and $A$ admits the representation (14).

Consider now the case where $A$ has some eigenvalues $\lambda_i$ for which the number of independent eigenvectors is less than their multiplicity $n_i$ in $o(s)$. In this case, $A$ cannot be diagonalized, but we can represent it in terms of its generalized eigenvectors.

Definition: $x \neq 0$ is a generalized eigenvector of grade $r$ of $A$ if

(i) $(\lambda I - A)^{\ell} x \neq 0$ for $\ell < r$

(ii) $(\lambda I - A)^r x = 0$.
Let \( G(\lambda_i) = \{ \mathbf{x} : (\lambda I - A)^r \mathbf{x} = 0 \text{ for some } r \} \) be the generalized eigenspace of \( A \) associated to eigenvalue \( \lambda_i \). Then Lemma 1 can be extended as follows.

**Lemma 3:** If \( \{ \mathbf{x}_i, 1 \leq i \leq k \} \) are generalized eigenvectors of \( A \) corresponding to eigenvalues \( \{\lambda_i, 1 \leq i \leq k \} \) with \( \lambda_i \neq \lambda_j \) for \( i \neq j \), they are linearly independent.

**Proof:** For each \( i \), since \( \mathbf{x}_i \) is a generalized eigenvector of \( A \) corresponding to \( \lambda_i \), we have \( (\lambda_i I - A)^{r_i} \mathbf{x}_i = 0 \) for some \( r_i \). Then, suppose there exists a nontrivial linear dependence relation

\[
\sum_{i=1}^{k} u_i \mathbf{x}_i = 0
\]

between the \( \mathbf{x}_i \)'s, where at last one of the coefficients \( u_i \) is different from zero, say \( u_1 \neq 0 \). Consider now the polynomials

\[
p_1(s) = (s - \lambda_1)^{r_1}, \quad p_2(s) = \prod_{i=2}^{k} (s - \lambda_i)^{r_i}.
\]

By multiplying (16) on the left by \( p_2(A) \), we find

\[
u_1 p_2(A) \mathbf{x}_1 = 0
\]

where \( u_1 \neq 0 \), so that \( p_2(A) \mathbf{x}_1 = 0 \). On the other hand, we also know that \( p_1(A) \mathbf{x}_1 = 0 \). Since \( p_1(s) \) and \( p_2(s) \) have no common roots, they are coprime, so that there exist polynomials \( m_1(s) \) and \( m_2(s) \) such that

\[
m_1(s)p_1(s) + m_2(s)p_2(s) = 1.
\]

This implies that

\[
\mathbf{x}_1 = m_1(A)p_1(A) \mathbf{x}_1 + m_2(A)p_2(A) \mathbf{x}_1 = 0,
\]

which is a contradiction since the vector \( \mathbf{x}_1 \) must be nonzero in order to be a generalized eigenvector of \( A \). \( \square \)

Then, we have:

**Lemma 4:** Any vector \( \mathbf{x} \) of \( \mathbb{R}^n \) can be expressed as a linear combination of vectors in \( G(\lambda_i) \) for \( 1 \leq i \leq k \), i.e.,

\[
\mathbb{R}^n = G(\lambda_1) \oplus G(\lambda_2) \oplus \cdots \oplus G(\lambda_k). \tag{17}
\]

This means that a basis of \( \mathbb{R}^n \) can be obtained by combining bases of generalized eigenspaces \( G(\lambda_i) \).

**Proof:** According to Lemma 3, vectors belonging to different eigenspaces \( G(\lambda_i) \) are linearly independent. To show that they span \( \mathbb{R}^n \), consider the characteristic polynomial

\[
det(sI - A) = a(s) = \prod_{i=1}^{k} (s - \lambda_i)^{r_i}
\]


and let
\[ p_i(s) = \prod_{j \neq i} (s - \lambda_j)^{n_j} \]
for \(1 \leq i \leq k\). The polynomials \(p_i(s)\) are coprime, so that there exist polynomials \(m_i(s)\) such that
\[ 1 = \sum_{i=1}^{k} m_i(s)p_i(s). \]
This implies
\[ I = \sum_{i=1}^{k} m_i(A)p_i(A), \]
so that for an arbitrary vector \(x \in \mathbb{R}^n\) we have
\[ x = \sum_{i=1}^{k} x_i \quad (18) \]
with
\[ x_i \triangleq m_i(A)p_i(A)x. \]
For each \(i\), \(x_i\) belongs to the generalized eigenspace \(G(\lambda_i)\) since
\[ (\lambda_i I - A)^{n_i}x_i = m_i(A)a(A)x = 0, \]
where the last equality uses Cayley-Hamilton’s identity \(a(A) = 0\).

The relation (18) shows that an arbitrary vector \(x\) of \(\mathbb{R}^n\) can be expressed as the linear combination of vectors in \(G(\lambda_i)\) with \(1 \leq i \leq k\). \(\square\)

Then, if \(x\) is a generalized eigenvector of grade \(r\) of \(A\) associated to eigenvalue \(\lambda\), we can construct the chain
\[
\begin{align*}
x_r &= x \\
x_{r-1} &= (A - \lambda I)x_r \\
\vdots &= \vdots \\
x_1 &= (A - \lambda I)x_2 \\
0 &= (A - \lambda I)x_1,
\end{align*}
\]
where each vector \(x_j\) belongs to \(G(\lambda)\). The only eigenvector in this chain is \(x_1\). All other vectors \(x_2, \ldots, x_r\) are generalized eigenvectors of grade \(2, \ldots, r\). The effect of \(A\) on this chain is given by
\[
A \begin{bmatrix} x_1 & x_2 & \cdots & x_r \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_r \end{bmatrix} J
\]
with
\[
J \triangleq \begin{bmatrix}
\lambda & 1 & 0 \\
\lambda & \cdots & \lambda \\
0 & \cdots & 1 \\
\end{bmatrix}.
\]
By constructing a basis for each generalized eigenspace \( G(\lambda_i) \) in terms of such chains, and letting

\[
X = [ \; x_1 \; x_2 \; \ldots \; x_n \; ]
\]

be the basis of \( \mathbb{R}^n \) obtained by combining all such bases of \( G(\lambda_i) \) we find that \( A \) can be expressed as

\[
A = XJX^{-1}
\]

with

\[
J = \begin{bmatrix}
J_1 & 0 & \cdots & 0 \\
0 & J_2 & & 0 \\
\vdots & & \ddots & \vdots \\
0 & 0 & \cdots & J_s
\end{bmatrix}
\]

\[
J_j = \begin{bmatrix}
\lambda_j & 1 & 0 \\
0 & \lambda_j & 1 \\
\vdots & \vdots & \ddots \\
0 & \cdots & 0 & \lambda_j
\end{bmatrix},
\]

where each block \( J_j \) has size \( r_j \times r_j \). The matrix \( J \) is called the **Jordan form** of \( A \) and \( J_j \) is a Jordan block of size \( r_j \) corresponding to eigenvalue \( \lambda_j \). There may be several Jordan blocks with the same eigenvalue. If \( n \) is the dimension of \( A \), we have \( r_1 + r_2 + \ldots + r_s = n \).

**Example:** If

\[
J = \begin{bmatrix}
21 & 0 & 0 & 0 \\
0 & 21 & 0 & 0 \\
0 & 0 & 21 & 0 \\
0 & 0 & 0 & 21
\end{bmatrix},
\]

\[
J \text{ has 3 blocks of size 1, one block of size 2 and one block of size 3 associated to eigenvalue } \lambda = 2. \text{ The characteristic polynomial of } J \text{ is } a(s) = (s-2)^3, \text{ but there are only 5 eigenvectors associated to } \lambda = 2 \text{ (one for each Jordan block). These are given by}
\]

\[
e_1 = \begin{bmatrix}
1 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}, \quad e_2 = \begin{bmatrix}
0 \\
1 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}, \quad e_3 = \begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

\[
e_4 = \begin{bmatrix}
0 \\
0 \\
0 \\
1 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}, \quad \text{and } e_6 = \begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
1 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}.
\]
These eigenvectors have a 1 in the row corresponding to the beginning of each Jordan block (remember that each Jordan block has only one eigenvector).

**Minimal polynomial:** Let $A$ be an arbitrary matrix. Consider the set of all polynomials $p(s)$ such that $p(A) = 0$. The characteristic polynomial $a(s) = \det(sI - A)$ belongs to this set, since by the Cayley-Hamilton theorem $a(A) = 0$. However $a(s)$ need not be the polynomial of smallest degree which is annulled by $A$. The polynomial $m(s)$ of smallest degree such that $m(A) = 0$ is called the minimal polynomial of $A$.

An important feature of $m(s)$ is that it must divide $a(s)$, i.e., there exists a polynomial $q(s)$ such that $a(s) = m(s)q(s)$. To see why this is the case, assume that $m(s)$ does not divide $a(s)$. Then, by Euclidean division, we can find polynomials $q(s)$ and $r(s)$ such that

$$a(s) = q(s)m(s) + r(s)$$

with $\deg r(s) < \deg m(s)$. But $a(A) = m(A) = 0$, so

$$r(A) = a(A) - q(A)m(A) = 0,$$

i.e., we have constructed a polynomial $r(s)$ of smaller degree than $m(s)$ such that $r(A) = 0$, a contradiction since $m(s)$ is the minimal polynomial. Thus $m(s)$ is a divisor of $a(s)$.

To find the minimal polynomial, observe that if $J$ is the Jordan form of $A$ and $p(s)$ is an arbitrary polynomial, then

$$p(A) = Xp(J)X^{-1}$$

with

$$p(J) = \begin{bmatrix}
  p(J_1) & & & \\
  & \ddots & & \\
  & & p(J_j) & \\
  & & & 0 \\
  0 & & & p(J_k)
\end{bmatrix}.$$  

The minimal polynomial $m(s)$ must be the polynomial $p(s)$ of least degree such that $p(J_j) = 0$ for all Jordan blocks $J_j$. This implies that $m(s)$ is the least common multiple of the minimal polynomials $m_j(s)$ of the Jordan block $J_j$. To find $m_j(s)$, note $a_j(s) = \det(sI - J_j) = (s - \lambda_j)^{r_j}$ where $r_j$ is the size of $J_j$. Furthermore

$$J_j - \lambda_j I = N_j = \begin{bmatrix}
  0 & 1 & & \\
  & \ddots & \ddots & \\
  & & \ddots & 0 \\
  0 & & & 1 \\
  & & & 0
\end{bmatrix}$$

is a nilpotent matrix of grade $r_j$ since

$$N_j^r_j = \begin{bmatrix}
  0 & 1 & & \\
  & \ddots & \ddots & \\
  & & \ddots & 1 \\
  & & & 0
\end{bmatrix} \neq 0$$

12
has ones along its \( \ell \)-th superdiagonal for \( \ell < r_j \) and \( N_j^\ell = 0 \) for \( \ell \geq r_j \). This implies that the minimal polynomial of \( J_j \) is equal to its characteristic polynomial, i.e., \( m_j(s) = a_j(s) = (s - \lambda_j)^{r_j} \).

If \( \{\lambda_1, \ldots, \lambda_k\} \) is the set of distinct eigenvalues of \( A \), the minimal polynomial (the least common multiple of the polynomials \( m_j(s) \)) is therefore given by

\[
m(s) = \prod_{i=1}^{k} (s - \lambda_i)^{r_i^{\text{max}}},
\]

where \( r_i^{\text{max}} \) is the size of the largest Jordan block associated to eigenvalue \( \lambda_i \).

**Example:** If we consider the matrix \( J \) given in (26), the size of the largest Jordan block associated to \( \lambda = 2 \) is 3, so that \( m(s) = (s - 2)^3 \).

**Comment:** At this point, it is worth noting that although the concepts of Jordan form and minimal polynomial can be useful for analytical derivations, they are somewhat unreliable from a numerical viewpoint, since small perturbations in the entries of a matrix have the effect of making all its eigenvalues distinct, thus making the computation of its Jordan form rather difficult.