UNIVERSITY OF CALIFORNIA, DAVIS
Department of Electrical and Computer Engineering

EEC250 Linear Systems and Signals Fall 2009

Lecture 4

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 2×2 matrix with columns $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^2$. If we view $\mathbb{R}^2$ as embedded in the three-dimensional space $\mathbb{R}^3$, the outer product

$$\mathbf{a}_1 \times \mathbf{a}_2 = \det(\mathbf{a}_1, \mathbf{a}_2)\mathbf{e}_3 \tag{1}$$

where $\mathbf{e}_3$ is the unit vector along the third axis and

$$\det(\mathbf{a}_1, \mathbf{a}_2) = |\mathbf{a}_1||\mathbf{a}_2| \sin(\theta) \tag{2}$$

measures the area of the parallelogram spanned by $\mathbf{a}_1$ and $\mathbf{a}_2$ as shown in Fig.1 below.

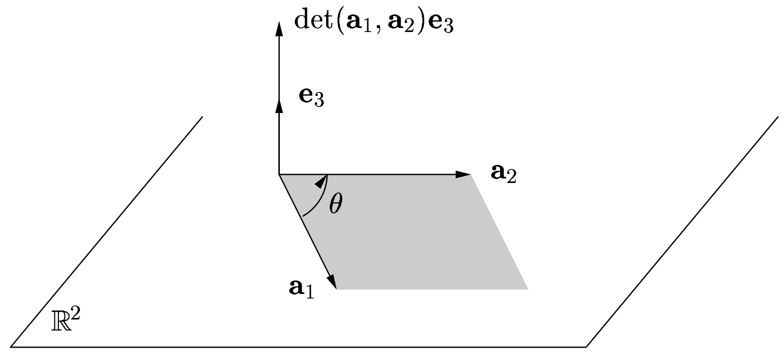

Figure 1: Interpretation of the determinant in two dimensions as the oriented area of the parallelogam spanned by $\mathbf{a}_1$ and $\mathbf{a}_2$.

In expression (2),

$$|\mathbf{a}_i| = (\mathbf{a}_i^T \mathbf{a}_i)^{1/2}$$

denotes the length (Euclidean norm) of vector $\mathbf{a}_i$ for $i = 1, 2$ and θ is the oriented angle going from vector $\mathbf{a}_1$ to vector $\mathbf{a}_2$, so that $\det(\mathbf{a}_1, \mathbf{a}_2)$ is an *oriented area* in the sense that

$$\det(\mathbf{a}_2, \mathbf{a}_1) = -\det(\mathbf{a}_1, \mathbf{a}_2) ,$$

since when $\mathbf{a}_1$ and $\mathbf{a}_2$ are interchanged, θ 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 2×2 case, if

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

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 $\mathbb{R}$:

$$(\mathbf{a}_1, \ldots, \mathbf{a}_n) \in (\mathbb{R}^n)^n \to \det(\mathbf{a}_1, \ldots, \mathbf{a}_n) \in \mathbb{R}$$

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 u and v are arbitrary real numbers

$$\begin{aligned} \det(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, u\mathbf{x} + v\mathbf{y}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n) &= u \det(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{x}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n) \\ &+ v \det(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{y}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n). \end{aligned} \quad (3)$$

(ii) If vectors $\mathbf{a}_i$ and $\mathbf{a}_j$ with $i < j$ are interchanged, we have

$$\det(\mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_n) = -\det(\mathbf{a}_1, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_i, \mathbf{a}_n) \quad (4)$$

(iii) If there exists a nontrivial linear dependence relation

$$\sum_{i=1}^{n} u_i \mathbf{x}_i = \mathbf{0} \quad \text{with} \quad \mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \neq \mathbf{0} \quad (5)$$

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:

$$\det A = \sum_{\pi} (-1)^{t(\pi)} a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)}, \quad (6)$$

where in (6) the sum is over all permutations π of the index set $\{1, 2, \dots, n\}$, and $t(\pi)$ is the number of transpositions occuring in the permutation σ. 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 3×3 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

$$\begin{aligned} \det A &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} . \end{aligned}$$

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

$$A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_i & \dots & \mathbf{a}_n \end{bmatrix} \longrightarrow A_s = \begin{bmatrix} \mathbf{a}_1 & \dots & c\mathbf{a}_i & \dots & \mathbf{a}_n \end{bmatrix} .$$

Then, according to (3)

$$\det A_s = c \det A .$$

(ii) Exchange of two columns. If

$$A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_i & \dots & \mathbf{a}_j & \dots & \mathbf{a}_n \end{bmatrix} \longrightarrow A_p = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_j & \dots & \mathbf{a}_i & \dots \mathbf{a}_n \end{bmatrix},$$

then $\det A_p = -\det A$.

(iii) Adding to a column a multiple of another column. Let

$$A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_i & \dots & \mathbf{a}_j & \dots & \mathbf{a}_n \end{bmatrix} \longrightarrow A_c = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_i + v\mathbf{a}_j & \dots & \mathbf{a}_j & \dots \mathbf{a}_n \end{bmatrix},$$

with $v \in \mathbb{R}$. The multilinearity property (3) implies

$$\det A_c = \det A + v \det B$$

where the matrix

$$B = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_j & \dots & \mathbf{a}_j & \dots & \mathbf{a}_n \end{bmatrix}$$

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

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

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

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

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

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

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.

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

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

$$\begin{aligned} D_n &= 2D_{n-1} + \det \begin{bmatrix} -1 & -1 & & & & \\ 0 & 2 & -1 & & 0 & \\ \vdots & -1 & 2 & -1 & & \\ \vdots & & \ddots & \ddots & \ddots & \\ \vdots & 0 & & \ddots & \ddots & \\ 0 & & & & -1 & 2 \end{bmatrix} \\ &= 2D_{n-1} - D_{n-2} , \end{aligned}$$

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

$$A\tilde{A} = (\det A) I_n ,$$

where $\tilde{A}$ is the adjugate matrix of A. $\tilde{A} = (\tilde{a}_{ij},\ 1 \le i, j \le n)$ is the transpose of the matrix formed by the cofactors of A, i.e., $\tilde{a}_{ij} = C_{ji}$ 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

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

we see that $a(s)$ is a polynomial of degree n where the coefficient of s^n equals 1. Thus

$$a(s) = s^n + a1s^{n-1} + \ldots + a_n = \prod_{i=1}^{k}(s - \lambda_i)^{n_i} ,$$

where the multiplicities n_i of roots λ_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 charcteristic polynomial, i.e.,

$$a(A) = A^n + a_1 A^{n-1} + a_2 A^{n-2} + \ldots + a_n I_n = 0 . \tag{9}$$

This important result can be established by using the identity

$$(sI - A)(\widetilde{sI - A}) = a(s) I_n \tag{10}$$

and noting that the adjugate matrix $\widetilde{sI - A}$ is a matrix polynomial of degree $n - 1$, so it can be written as

$$\widetilde{sI - A} = R1s^{n-1} + R_2 s^{n-2} + \ldots + R_n. \tag{11}$$

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)

$$\begin{aligned} s^n &: \quad R_1 = I \\ s^{n-1} &: \quad R_2 - AR_1 = a_1 I \\ \ldots & \\ s^0 &: \quad -AR_n = a_n I , \end{aligned}$$

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 $\mathbf{x} \neq \mathbf{0}$.

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., λ 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 λ_i is an eigenvalue of A with multiplicity n_i, λ_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.

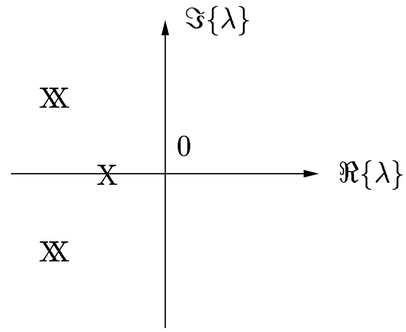

Figure 2: Real axis symmetry of the eigenvalues of a real matrix.

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

$$\sum_{i=1}^{k} u_i \mathbf{x}_i = \mathbf{0}\,, \tag{12}$$

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

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

which is a contradiction since $u_1 \neq 0$, $\prod_{i=2}^{k}(\lambda_1 - \lambda_i) \neq 0$ and $\mathbf{x}_1 \neq \mathbf{0}$. 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

$$AX = X\Lambda \tag{13}$$

with

$$X \triangleq \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \end{bmatrix}$$

and

$$\Lambda = \mathrm{diag}\ (\lambda_1, \lambda_2, \ldots, \lambda_n)\,.$$

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

$$A = X\Lambda X^{-1}\,. \tag{14}$$

This shows that A is related to the diagonal matrix Λ through a similarity transformation. In this case, 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:

$$\begin{aligned}\det(sI - B) &= \det\left(T(sI - A)T^{-1}\right) \\ &= \det T \det(sI - A) \det T^{-1} = \det(sI - A) ,\end{aligned}$$

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

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

$$\begin{aligned}X &= \begin{bmatrix} X_1 & X_2 & \dots X_k \end{bmatrix} \\ \Lambda &= \text{diag}\,(D_1, D_2, \dots, D_k) ,\end{aligned}$$

where

$$X_i = \begin{bmatrix} \mathbf{x}_{i1} & \dots \mathbf{x}_{i\ell} & \dots & \mathbf{x}_{in_i} \end{bmatrix} \quad \text{and} \quad D_i = \lambda_i I_{n_i} .$$

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

$$\sum_{i=1}^{k} \Big(\sum_{\ell=1}^{n_i} u_{i\ell} \mathbf{x}_{i\ell} \Big) = \mathbf{0} \tag{15}$$

between the columns of X. Let

$$\mathbf{x}_i \triangleq \sum_{\ell=1}^{n_i} u_{i\ell} \mathbf{x}_{i\ell} .$$

Depending on whether the coefficients $u_{i\ell}$, $1 \leq \ell \leq n$ are all zero or not, $\mathbf{x}_i$ is either the zero vector or an eigenvector of A associated with eigenvalue λ_i (it is a linear combination of such eigenvectors). If $\mathbf{x}_i \neq \mathbf{0}$ for at least one i, the relation (15) indicates thar there exists a linear dependence relation between several eigenvectors of A associated to distinct eigenvalues λ_i, $1 \leq i \leq k$. According to Lemma 1, this is impossible, so that we must have $\mathbf{x}_i = 0$ for all i. But for each i, the eigenvectors $\mathbf{x}_{i\ell}$ with $1 \leq \ell \leq n_i$ are linearly independent, so we must have $u_{i\ell} = 0$ for all i and ℓ. 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 λ_i for which the number of independent eigenvectors is less than their multiplicity n_i in $a(s)$. In this case, A cannot be diagonalized, but we can represent it in terms of its generalized eigenvectors.

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

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

(ii) $(\lambda I - A)^{r} \mathbf{x} = \mathbf{0}$.

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 λ_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 λ_i, we have $(\lambda_i I - A)^{r_i} \mathbf{x} = \mathbf{0}$ for some r_i. Then, suppose there exists a nontrivial linear dependence relation

$$\sum_{i=1}^{k} u_i \mathbf{x}_i = \mathbf{0} \tag{16}$$

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

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

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.,

$$\mathbb{R}^n = G(\lambda_1) \oplus G(\lambda_2) \ldots \oplus G(\lambda_k). \tag{17}$$

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

$$\det(sI - A) = a(s) = \prod_{i=1}^{k} (s - \lambda_i)^{n_i}$$

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 $\mathbf{x} \in \mathbb{R}^n$ we have

$$\mathbf{x} = \sum_{i=1}^{k} \mathbf{x}_i \tag{18}$$

with

$$\mathbf{x}_i \triangleq m_i(A) p_i(A) \mathbf{x} \, .$$

For each i, $\mathbf{x}_i$ belongs to the generalized eigenspace $G(\lambda_i)$ since

$$(\lambda_i I - A)^{n_i} \mathbf{x}_i = m_i(A) a(A) \mathbf{x} = \mathbf{0} \, ,$$

where the last equality uses Cayley-Hamilton's identity $a(A) = 0$.

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

Then, if $\mathbf{x}$ is a generalized eigenvector of grade r of A associated to eigenvalue λ, we can construct the chain

$$\begin{array}{rcl} \mathbf{x}_r & = & \mathbf{x} \\ \mathbf{x}_{r-1} & = & (A - \lambda I)\mathbf{x}_r \\ \dots & & \dots \\ \mathbf{x}_1 & = & (A - \lambda I)\mathbf{x}_2 \\ \mathbf{0} & = & (A - \lambda I)\mathbf{x}_1 \, , \end{array}$$

where each vector $\mathbf{x}_j$ belongs to $G(\lambda)$. The only eigenvector in this chain is $\mathbf{x}_1$. All other vectors $\mathbf{x}_2, \dots, \mathbf{x}_r$ are generalized eigenvectors of grade $2, \dots, r$. The effect of A on this chain is given by

$$A \left[\begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_r \end{array} \right] = \left[\begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_r \end{array} \right] J$$

with

$$J \triangleq \begin{bmatrix} \lambda & 1 & & 0 \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{bmatrix} .$$

By constructing a basis for each generalized eigenspace $G(\lambda_i)$ in terms of such chains, and letting

$$X = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \end{bmatrix}$$

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

$$A = XJX^{-1} \tag{19}$$

with

$$J = \begin{bmatrix} J_1 & & & & 0 \\ & \ddots & & & \\ & & J_j & & \\ & & & \ddots & \\ 0 & & & & J_s \end{bmatrix} \quad J_j = \begin{bmatrix} \lambda_j & 1 & & 0 & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & \lambda_j & 1 \\ 0 & & & & \lambda_j \end{bmatrix},$$

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 λ_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

$$J = \left[\begin{array}{c|c|c|cc|ccc} 2 & 0 & & & & & & \\ \hline 0 & 2 & 0 & & & & 0 & \\ \hline & & 2 & & & & & \\ \hline & & & 2 & 1 & & & \\ & & & 0 & 2 & & & \\ \hline & & & & & 2 & 1 & 0 \\ & 0 & & & & 0 & 2 & 1 \\ & & & & & 0 & 0 & 2 \end{array}\right], \tag{20}$$

J has 3 blocks of size 1, one block of size 2 and one block of size 3 associated to eigenvalue $\lambda = 2$. The characteristic polynomial of J is $a(s) = (s-2)^8$, but there are only 5 eigenvectors associated to $\lambda = 2$ (one for each Jordan block). These are given by

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{e}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{e}_6 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

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

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

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

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

is a nil potent matrix of grade r_j since

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

has ones along its ℓ-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

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

where $r_i^{\max} =$ is the size of the largest Jordan block associated to eigenvalue λ_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.