LINEAR ALGEBRA

This article deals with linear vector spaces, transformations, quadratic forms, and structural relationships between algebraic systems. Matrix theory concepts necessary to compute functions of matrices involved in system theory are developed.

VECTOR SPACES

Definition

Many different topics, such as matrices, orthogonal polynomials, Fourier series, and integrodifferential equations, can be united as a study of linear vector spaces, because they all satisfy the following definition: A *linear vector space* **v** (over a scalar field R or C) is a set of objects $\boldsymbol{x}, \boldsymbol{y}, \ldots$ called *vectors,* together with the two operations of *addition* and *scalar multiplication* with the following properties: If vectors $\mathbf{0}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \ldots \in \mathbf{v}$ and α and β are complex numbers, then

1. $\boldsymbol{x} + \boldsymbol{y} \in \mathbf{v}$
2. $\alpha\boldsymbol{x} \in \boldsymbol{v}$
3. $\alpha(\boldsymbol{x} + \boldsymbol{y}) = \alpha\boldsymbol{x} + \alpha\boldsymbol{y}$
4. $(\alpha\beta)\boldsymbol{x} = \alpha(\beta\boldsymbol{x})$
5. $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$
6. $1\mathbf{x} = \mathbf{x}$
7. $\mathbf{0} + \boldsymbol{x} = \boldsymbol{x}$
8. $\boldsymbol{x} + (-\boldsymbol{x}) = \boldsymbol{0}$

A *metric vector space* is associated with a real valued non-negative function such that:

1. $g(\mathbf{x},\mathbf{y}) = g(\mathbf{y},\mathbf{x})$
2. $g(\mathbf{x},\mathbf{y}) = 0$, if $\mathbf{x} = \mathbf{y}$
3. $g(\mathbf{x},\mathbf{y}) \leq g(\mathbf{x},\mathbf{z}) + g(\mathbf{z},\mathbf{y})$

A *metric vector space* is called complete if every "Cauchy sequence" $\mathbf{x}_n$ in the *metric space* converges to some $\mathbf{x} \in \mathbf{v}$. A metric space is normed if for all vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{v}$ and a scalar α a norm $\| \ \|$ is defined with the following properties:

1. $\|\mathbf{x}\| > 0$, $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = 0$
2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + |\|\mathbf{y}\|$

A normed metric space which is complete is called *Banach* space.

Some important metric notions such as length, direction, and energy can be expressed if the vector space is endowed with the additional *inner product* $(\boldsymbol{x}, \boldsymbol{y})$ of $\boldsymbol{x}$ and $\boldsymbol{y}$ which satisfies

1. $(\boldsymbol{x}, \boldsymbol{y}) = \overline{(\boldsymbol{y}, \boldsymbol{x})}$
2. $(\boldsymbol{x}, \boldsymbol{x}) \geq \mathbf{0}$
3. $(\alpha\boldsymbol{x}, \boldsymbol{y}) = \overline{\alpha}(\boldsymbol{x}, \boldsymbol{y})$
4. $(\mathbf{x}, \mathbf{x}) = \mathbf{0} \rightarrow \mathbf{x} = \mathbf{0}$
5. $(\mathbf{x}, \mathbf{y} + \mathbf{z}) = (\mathbf{x}, \mathbf{y}) + (\mathbf{x}, \mathbf{z})$

where the overbar stands for complex conjugation.

A *Banach space* with the inner product defined as above is known as *Hilbert space.*

Different Types of Linear Spaces

***n*-Dimensional Euclidean Space.** Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are a n-tuples complex numbers

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Their inner product is

$$(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} \overline{x}_i y_i$$

The *length* of $\boldsymbol{x}$ is

$$(\boldsymbol{x}, \boldsymbol{x})^{1/2} = \sum_{i=1}^{n} (\overline{x}_i x_i)^{1/2} = \sum_{i=1}^{n} |x_i| = \|\boldsymbol{x}\| \geq 0$$

The *Cauchy–Schwartz inequality* states

$$|(\boldsymbol{x}, \boldsymbol{y})| \leq \|\boldsymbol{x}\| \, \|\boldsymbol{y}\|$$

Vectors $\boldsymbol{x}_i$ form an *orthonormal set* if

$$(\boldsymbol{x}_i, \boldsymbol{x}_j) = \delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \qquad (i, j = 1, 2, \ldots)$$

The $L^2(a, b)$ Space Known as Hilbert Space. The sum of two functions (vectors) $\boldsymbol{f}(t)$ and $\boldsymbol{g}(t)$ is $\boldsymbol{f}(t) + \boldsymbol{g}(t)$, and their inner product is

$$(\boldsymbol{f}, \boldsymbol{g}) = \int_a^b \overline{\boldsymbol{f}}(t)\boldsymbol{g}(t)\,dt$$

Polynomial Space. Let

$$\boldsymbol{f}(t) = \sum_{i=1}^{n} a_i t^i$$

Addition is defined as usual, and for a weighting function $w(t) > 0$ and interval (a, b), the inner product is

$$(\boldsymbol{f}, \boldsymbol{g}) = \int_a^b \overline{\boldsymbol{f}}(t)w(t)\boldsymbol{g}(t)\,dt$$

Generalized Fourier Space. Define an orthonormal set of vectors $\boldsymbol{e}_k$ ($k = 0, \pm1, \pm2, \ldots$). If $\boldsymbol{x}$ is any arbitrary vector in $\mathbf{v}$, then its *Fourier expansion* is

$$\boldsymbol{x} = \sum_{k=-\infty}^{\infty} c_k \boldsymbol{e}_k$$

with *Fourier coefficients*

$$(\boldsymbol{e}_k, \boldsymbol{x}) = c_k \qquad (k = 0, \pm1, \pm2, \ldots)$$

In particular, the *Fourier vectors* are given by

$$\boldsymbol{e}_k = \boldsymbol{e}_k(t) = e^{jk\omega t}$$

on the interval $[a, b] \equiv [-T, T]$, $\omega = \pi/T$.

Parseval's identity is

$$(\boldsymbol{x}, \boldsymbol{x}) = \sum_{k=-\infty}^{\infty} |(\mathbf{e}_k, \mathbf{x})|^2 = \sum_{k=-\infty}^{\infty} |c_k|^2$$

If we choose n-dimensional orthogonal space, the *least-squares error approximation* of $\boldsymbol{x}$ is defined as

$$\boldsymbol{x}^* = \sum_{k=-n}^{n} c_k \boldsymbol{e}_k, \qquad (\boldsymbol{e}_k, \boldsymbol{e}_j) = \delta_{kj} \quad (k, j = 0, \pm 1, \pm 2, \ldots)$$

Bessel's inequality is

$$(\boldsymbol{x}, \boldsymbol{x}) = \sum_{k=-\infty}^{\infty} |c_k|^2 \geq \sum_{k=-n}^{n} |c_k|^2$$

An orthogonal set is *complete* if ther is $\boldsymbol{x}$ for which

$$\|\boldsymbol{x} - \boldsymbol{x}^*\| < \epsilon, \qquad \epsilon \geq 0$$

The *Riemann-Lebesgue lemma* states that

$$|(\boldsymbol{e}_k, \boldsymbol{x})| = |c_k| \to 0$$

as $k \to \infty$.

Gram–Schmidt Orthogonalization

Given a vector set $\boldsymbol{e}_i$ ($i = 1, 2, \ldots, n$) and constants α_i ($i = 1, 2, \ldots, n$) not all zero, such that

$$\sum_{i=1}^{n} \alpha_i \boldsymbol{e}_i = \mathbf{0}$$

the vectors are said to be *linearly dependent.* Otherwise the set is composed of *linearly independent* vectors.

Given a linearly independent set of vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$, suppose we are required to determine a new orthonormal set $\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_n$.

Take $\boldsymbol{f}_1 = \boldsymbol{e}_1/\|\boldsymbol{e}_1\|$. Assume $\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_k$ ($k < n$) have been computed; then

$$\boldsymbol{f}'_{k+1} = \boldsymbol{e}_{k+1} - \sum_{j=1}^{k} (\boldsymbol{f}_j, \boldsymbol{e}_{k+1}) \boldsymbol{f}_j, \quad \boldsymbol{f}_{k+1} = \frac{\boldsymbol{f}'_{k+1}}{\|\boldsymbol{f}'_{k+1}\|} \qquad k = 2, 3, \ldots, n$$

A vector space $\mathbf{v}$ is *n-dimensional* if it contains only n linearly independent vectors. Every set of $n + 1$ vectors is linearly dependent. The set of linearly independent vectors *spans* a space $\mathbf{v}$ if every vector $\boldsymbol{x} \in \mathbf{v}$ can be expressed as

$$\boldsymbol{x} = \sum_{i=1}^{n} \alpha_i \boldsymbol{e}_i$$

The $\boldsymbol{e}_i$, $i = 1, \ldots, n$, are called a *basis* of $\mathbf{v}$.

Linear Operators

A *linear operator* $\mathbf{T}$ on a vector space $\mathbf{v}$ defines a rule that computes $\mathbf{T}x$ for $\boldsymbol{x} \in \mathbf{v}$ such that

$$\mathbf{T}(\alpha_1 \boldsymbol{x} + \alpha_2 \boldsymbol{y}) = \alpha_1 \mathbf{T}\boldsymbol{x} + \alpha_2 \mathbf{T}\boldsymbol{y}$$

If any vector in $\mathbf{v}$ can be expressed as sum of vectors from *subspaces* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ ($\boldsymbol{x} = \sum_{i=1}^{n} \boldsymbol{x}_i$, $\boldsymbol{x} \in \mathbf{v}$, $\boldsymbol{x}_i \in \mathbf{v}_i$), then $\mathbf{v}$ itself is called the *sum of the subspaces:*

$$\mathbf{v} = \sum_{i=1}^{n} \mathbf{v}_i$$

The *direct sum* of vector spaces such that $\mathbf{v}_i \cap \mathbf{v}_j = 0$, $i \neq j$ is denoted

$$\mathbf{v} = \mathbf{v}_1 \dot{+} \mathbf{v}_2 \dot{+} \cdots = \dot{\sum_{i=1}^{k}} \mathbf{v}_i$$

The set $\{\mathbf{v}_i\}$ is called a *direct decomposition* of $\mathbf{v}$.

The *projection theorem* states that

$$\mathbf{v} = \mathbf{w} \dot{+} \mathbf{w}^{\perp}, \quad \boldsymbol{x} \in \mathbf{w}, \text{ and } \boldsymbol{y} \in \mathbf{w}^{\perp} \quad \text{implies} \quad (\boldsymbol{x}, \boldsymbol{y}) = 0$$

$$\dim(\mathbf{v}_1 + \mathbf{v}_2) = \dim \mathbf{v}_1 + \dim \mathbf{v}_2 - \dim(\mathbf{v}_1 \cap \mathbf{v}_2)$$

$$\dim \mathbf{v} = \dim(\mathbf{w} \dot{+} \mathbf{w}^{\perp}) = \dim \mathbf{w} + \dim \mathbf{w}^{\perp}$$

Two spaces $\mathbf{v}$ and $\mathbf{w}$ are *dual* to each other if the basis vectors of $\mathbf{v}$ are $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$, the basis vectors of $\mathbf{w}$ are $\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_n$, and

$$(\boldsymbol{e}_i, \boldsymbol{f}_j) = \delta_{ij} \qquad (i, j = 1, 2, \ldots)$$

The set $\boldsymbol{e}_i$ ($i = 1, \ldots, n$) from a *basis* for the space $\mathbf{v}$ if every vector $\boldsymbol{x} \in \mathbf{v}$ can be expressed as a linear combination of these vectors. The *dimension* of the space is the maximal number of linearly independent vectors in the space. In an n-dimensional linear vector space any set of n linearly independent vectors qualifies as a basis for the vector space.

EUCLIDIAN SPACE AND MATRIX REPRESENTATION

Consider two separate spaces $\mathbf{E}_m$ and $\mathbf{E}_n$ with bases $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_m$ and $\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_n$, respectively, given by

$$(\boldsymbol{e}_k)_l = (l\text{th component of } \boldsymbol{e}_k) = \delta_{kl} \qquad (k, l = 1, 2, \ldots, m)$$
$$(\boldsymbol{f}_j)_i = (i\text{th component of } \boldsymbol{f}_j) = \delta_{ij} \qquad (i, j = 1, 2, \ldots, n)$$

Let the operator $\mathbf{A}$ be a *transport,* or *linear mapping,* from $\mathbf{E}_m$ to $\mathbf{E}_n$ (see Fig. 1):

$$y = \mathbf{A}\boldsymbol{x}, \qquad \boldsymbol{x} \in \mathbf{E}_m, \quad \boldsymbol{y} \in \mathbf{E}_n$$

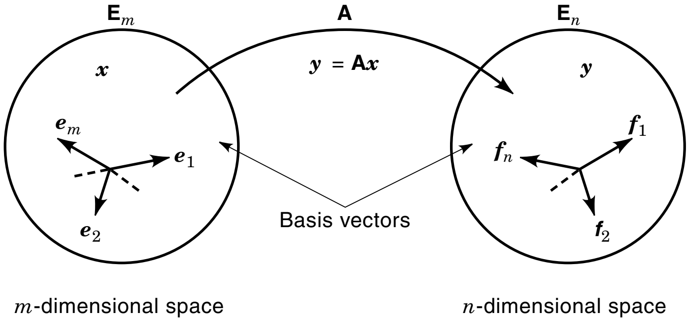

Figure 1. Matrix representation.

Let

$$\mathbf{A}\boldsymbol{e}_i = \boldsymbol{g}_i \qquad (i = 1, 2, \ldots, m)$$

Since $\boldsymbol{g}_i \in \mathbf{E}_n$, it can be represented as a linear combination of the $\boldsymbol{f}_j$:

$$\boldsymbol{g}_i = \sum_{j=1}^{n} a_{ij}\boldsymbol{f}_j$$

The operator $\mathbf{A}$ determines the numbers a_{ij}. We have

$$x = \sum_{i=1}^{m} x_i\boldsymbol{e}_i, \qquad y = \sum_{j=1}^{n} y_j\boldsymbol{f}_j \tag{1}$$

where x_i and y_j are the ith and jth components of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, and

$$\begin{aligned} y = \mathbf{A}x &= \sum_{i=1}^{m} x_i\mathbf{A}e_i = \sum_{i=1}^{m} x_i\boldsymbol{g}_i \\ &= \sum_{i=1}^{m} x_i \sum_{j=1}^{n} a_{ij}\boldsymbol{f}_j = \sum_{j=1}^{n}\left(\sum_{i=1}^{m} a_{ij}x_i\right)\boldsymbol{f}_j \end{aligned} \tag{2}$$

From (1) and (2),

$$y_j = \sum_{i=1}^{m} a_{ij}x_i \qquad (j = 1, 2, \ldots, n) \tag{3}$$

The action of the operator $\mathbf{A}$ can be fully computed from the numbers a_{ij} ($i = 1, \ldots, m; j = 1, \ldots, n$). These numbers, when arranged as a table of n rows and m columns, constitute an $n \times m$ matrix $\underline{\mathbf{A}}$. Thus Eq. (3) can be written

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \tag{4}$$

or

$$\underbrace{\underline{\boldsymbol{y}} = \underline{\mathbf{A}}\underline{\boldsymbol{x}}}_{\text{matrix relation}} \quad \leftrightarrow \quad \underbrace{\boldsymbol{y} = \mathbf{A}\boldsymbol{x}}_{\text{space relation}} \tag{5}$$

Thus, $\underline{\boldsymbol{x}}$, $\underline{\boldsymbol{y}}$, $\underline{\mathbf{A}}$ are the matrix representations of the vectors $\boldsymbol{x}$, $\boldsymbol{y}$ and the operator $\mathbf{A}$ with respect to the basis $\{\boldsymbol{e}_i\}_1^m$ in $\mathbf{E}_m$ and the basis $\{\boldsymbol{f}_j\}_1^n$ in $\mathbf{E}_n$. Basis vectors are analogous to coordinates. However, vectors and operators exist indenendently of the basis assigned to them.

A vector space whose vectors belong to some larger space is called a *subspace.* This concept is very useful in developing the *canonical form* of a matrix. Let $\mathbf{A}$ be a mapping of $\mathbf{E}_n$ onto itself. A subspace $\mathbf{E}_{n_i}$ of $\mathbf{E}_n$ is *invariant* with respect to $\mathbf{A}$ if $\mathbf{A}x \in \mathbf{E}_{n_i}$ implies $\boldsymbol{x} \in \mathbf{E}_{n_i}$. The structure of a *invariant mapping* (matrix) can be very usefully exploited by means of its invariant subspaces:

$$\mathbf{E}_n = \sum_{i=1}^{k} \mathbf{E}_{n_i} \qquad \left(i = 1, \ldots, k; \quad \sum_{j=1}^{k} n_j = n\right)$$

The basis of $\mathbf{E}_{n_i}$ consists of $\boldsymbol{e}_{ij}$ ($i = 1, \ldots, k; j = 1, \ldots, n_j$).

In the basis chosen, $\mathbf{A}$ can be represented by a *quasidiagonal* form

$$\underline{\mathbf{A}} = \begin{bmatrix} \underline{\mathbf{A}}_1 & & & \\ & \underline{\mathbf{A}}_2 & & \\ & & \ddots & \\ & & & \underline{\mathbf{A}}_n \end{bmatrix}$$

All other entries besides boxes along the main diagonal are zeros.

MATRIX ALGEBRA

A scalar is a special case of a matrix with one row and one column. Following is a review of matrix theory fundamentals:

1. *Column matrix* (or vector):

$$\underline{\boldsymbol{x}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad (n \times 1 \text{ matrix})$$

2. *Row matrix* (or vector):

$$\underline{\mathbf{x}}^{\mathrm{T}} = [x_1, \ldots, x_n] \qquad (1 \times n \text{ matrix})$$

3. *Matrix* of order $n \times m$:

$$\underline{\mathbf{A}} = (a_{ij}) \qquad (i = 1, \ldots, n; \quad j = 1, \ldots, m)$$

4. *Addition of matrices:*

$$\underline{\mathbf{A}} + \underline{\mathbf{B}} = (a_{ij} + b_{ij}) \qquad (i = 1, \ldots, n; \ j = 1, \ldots, m)$$

5. *Multiplication:* If $\underline{\mathbf{A}}$ is $n \times p$ and $\underline{\mathbf{B}}$ is $p \times m$,

$$\underline{\mathbf{A}}\,\underline{\mathbf{B}} = \sum_{k=1}^{p} a_{ik}b_{kj} \qquad (i = 1, \ldots, n; \quad j = 1, \ldots, m)$$

6. *Adjoint:* Let $\underline{\mathbf{A}}^{\mathrm{T}}$ be the transpose of $\underline{\mathbf{A}}$, $\underline{\mathbf{A}}^{\mathrm{T}} = (a_{ji})$ ($\underline{\mathbf{A}}$ with rows and columns exchanged). Then the *adjoint matrix* of $\underline{\mathbf{A}}$ is

$$\overline{\underline{\mathbf{A}}^{\mathrm{T}}} = \underline{\mathbf{A}}^*$$

$\underline{\mathbf{A}}$ is a *unitary matrix* if

$$\underline{\mathbf{A}}^{-1} = \underline{\mathbf{A}}^*$$

a *symmetric matrix* if

$$\underline{\mathbf{A}}^{\mathrm{T}} = \underline{\mathbf{A}}$$

and a *Hermitian matrix* (useful in physics) if

$$\overline{\underline{\mathbf{A}}^{\mathrm{T}}} = \underline{\mathbf{A}} = \underline{\mathbf{A}}^*$$

The *commutator* of $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$ is

$$\underline{\mathbf{AB}} - \underline{\mathbf{BA}} = [\underline{\mathbf{A}}, \underline{\mathbf{B}}]$$

$[\underline{\mathbf{A}}, \underline{\mathbf{B}}] = 0$ implies $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$ are Hermitian.

7. *Inverse matrix:* We shall assume knowledge of the determinant of a matrix and its elementary properties. Let $\Delta(\underline{\mathbf{A}})$ be the determinant of $\underline{\mathbf{A}}$, where $\underline{\mathbf{A}}$ is an $n \times n$ square matrix. Let A^{ij} be the *ij cofactor* of a_{ij}, that is, the determinant of the matrix $\underline{\mathbf{A}}$ after striking out the ith row and the jth column, multiplied by $(-1)^{i+j}$. Then

$$\sum_{j=1}^{n} a_{ij} A^{ij} = \Delta(\underline{\mathbf{A}})$$

(the *Laplace expansion*), and

$$\sum_{j=1}^{n} a_{ij} A^{kj} = \Delta(\underline{\mathbf{A}})\delta_{ik}$$

The *inverse matrix* $\underline{\mathbf{A}}^{-1}$ is given by

$$(\underline{\mathbf{A}}^{-1})_{ij} = [\Delta(\underline{\mathbf{A}})]^{-1} A^{ji}, \qquad \underline{\mathbf{A}}^{-1} = [\Delta(\underline{\mathbf{A}})]^{-1}(\mathrm{Adj}\boldsymbol{A})$$

and we have

$$\underline{\mathbf{AA}}^{-1} = \underline{\mathbf{A}}^{-1}\underline{\mathbf{A}} = \underline{\mathbf{I}} \qquad \text{(identity)}$$

$$(\underline{\mathbf{AA}}^{-1})_{ij} = \sum_{k=1}^{n} a_{ik}\,(\underline{\mathbf{A}}^{-1})_{kj} = [\Delta(\underline{\mathbf{A}})]^{-1} \sum_{k=1}^{n} a_{ik} A^{kj} = \delta_{ij}$$

8. The *determinant of a product of matrices* is $\Delta(\underline{\mathbf{AB}}) = \Delta(\underline{\mathbf{A}})\Delta(\underline{\mathbf{B}})$.
9. A *singular matrix* $\underline{\mathbf{A}}$ is one such that

$$\Delta(\underline{\mathbf{A}}) = 0$$

10. Sometimes it is useful to represent a $n \times m$ matrix as a collection of n row vectors or m column vectors:

$$\underline{\mathbf{A}} = \begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_n^T \end{bmatrix} \quad \text{or} \quad \underline{\mathbf{A}} = [\underline{\boldsymbol{a}}_1 \quad \dots \quad \underline{\boldsymbol{a}}_m]$$

b_i^T is a $1 \times m$ row vector, and $\underline{\boldsymbol{a}}_j$ is an $n \times 1$ column vector ($i = 1, \ldots, n; j = 1, \ldots, m$).

11. *Minors:* Choose any k rows and any k columns from the matrix $\underline{\mathbf{A}}$ and form a matrix. The determinant of this matrix, with rows and columns in their natural order, is called a *minor* of $\underline{\mathbf{A}}$ of the kth order.
12. *Projection matrices:* If $\underline{\mathbf{P}}$ is Hermitian and $\underline{\mathbf{P}}^n = \underline{\mathbf{P}}$ ($n = 2, \ldots$), then $\underline{\mathbf{P}}$ is called a *projection matrix.* Any arbitrary vector $\underline{\boldsymbol{x}}$ can be decomposed into $\underline{\boldsymbol{y}}$ and $\underline{\boldsymbol{z}}$ such that

$$\underline{\mathbf{Py}} = \underline{\mathbf{y}}, \qquad \underline{\mathbf{Pz}} = \underline{\mathbf{0}}, \qquad \underline{\mathbf{x}} = \underline{\mathbf{y}} + \underline{\mathbf{z}}, \qquad \underline{\mathbf{z}} = (\underline{\mathbf{I}} - \underline{\mathbf{P}})\underline{\mathbf{x}}$$

The vector space $\mathbf{v}$ can be decomposed into $\mathbf{w}$ and $\mathbf{w}^{\perp}$, where

$$(\underline{\mathbf{P}\boldsymbol{x}} = \underline{\boldsymbol{y}}) \in \mathbf{w} \text{ and } (\underline{\mathbf{I}} - \underline{\mathbf{P}})\underline{\boldsymbol{x}} = \underline{\boldsymbol{z}} \in \mathbf{w}^{\perp}, \qquad \mathbf{w} + \mathbf{w}^{\perp} = \mathbf{v}$$

RANK OF A MATRIX

The rank is a very useful concept in the solution of simultaneous equations. It can be defined in many different (but equivalent) ways. In particular, it is the largest order of a nonvanishing minor, and it is the maximum number of linearly independent rows (or of linearly independent columns). Thus, given an $n \times m$ matrix $\underline{\mathbf{A}}$, the rank $r \le n, m$.

Kernel and Range

Let $\underline{\mathbf{A}}$ be a transformation on $\mathbf{E}_n$ from $\mathbf{E}_m$. The *kernel* of $\underline{\mathbf{A}}$ is the totality of $\underline{\boldsymbol{x}} \in \mathbf{E}_m$ for which $\underline{\mathbf{A}x} = \underline{\mathbf{0}}$. The *range* of $\underline{\mathbf{A}}$ is the totality of vectors $\underline{\mathbf{A}x} \in \mathbf{E}_n$. These are denoted as Ker $\underline{\mathbf{A}}$ and rng $\underline{\mathbf{A}}$. Let dim stand for dimension. Then dim Ker $\underline{\mathbf{A}}$ is also known as the *nullity* of $\underline{\mathbf{A}}$. Furthermore, dim rng $\underline{\mathbf{A}}$ is the rank of $\underline{\mathbf{A}}$.

Sylvester's law of nullity states that

$$\dim \mathrm{Ker}\,\underline{\mathbf{A}} + \dim \mathrm{rng}\,\underline{\mathbf{A}} = \dim \mathbf{E}_n$$

Systems of Linear Algebraic Equations

Let $\underline{\mathbf{A}}$ be an $n \times m$ matrix, $\underline{\mathbf{x}}$ be an $m \times 1$ matrix, and $\underline{\boldsymbol{b}}$ be an $n \times 1$ matrix forming a system of equations

$$\underline{\mathbf{A}\boldsymbol{x}} = \underline{\boldsymbol{b}}$$

Let $\underline{\mathbf{B}} = [\underline{\mathbf{A}}, \underline{\boldsymbol{b}}]$ be the $n \times (m + 1)$ augmented matrix.

This system has a solution only if $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$ have same rank r. Only r of the m components of $\underline{\boldsymbol{x}}$ can be uniquely determined, and $m - r$ can be assigned at will.

Nonsingular Matrices

For an $n \times n$ matrix $\underline{\mathbf{A}}$ to be nonsingular (invertible), its determinant $\Delta(\mathbf{A})$ must be nonzero, which implies that its rows (or columns) are linearly independent. This means the rank of $\underline{\mathbf{A}}$ is n. For an invertible $\underline{\mathbf{A}}$, $\underline{\mathbf{A}x} = \underline{\mathbf{0}}$ implies no linearly independent solutions besides $\underline{\boldsymbol{x}} = \underline{\mathbf{0}}$.

EIGENVALUES AND EIGENVECTORS OF MATRICES

We shall consider only $n \times n$ square matrices.

Eigenvalue and Eigenvector

Suppose $\underline{\mathbf{A}x} = \lambda\underline{\boldsymbol{x}}$. Then the scalar λ is known as an *eigenvalue* and $\underline{\boldsymbol{x}}$ as an *eigenvector* of $\underline{\mathbf{A}}$. We have

$$(\lambda\underline{\mathbf{I}} - \underline{\mathbf{A}})\underline{\boldsymbol{x}} = \underline{\mathbf{A}}(\lambda)\underline{\boldsymbol{x}} = \underline{\mathbf{0}} \tag{6}$$

implying

$$\sum_{j=1}^{n} (\lambda\delta_{ij} - a_{ij})\underline{\boldsymbol{x}}_j = \underline{\mathbf{0}} \qquad (i = 1, \ldots, n)$$

If the rank of $\underline{\mathbf{A}}(\lambda)$ is $r \leq n$, then it has r linearly independant nontrivial eigenvectors and a maximum of r distinct eigenvalues. For at least one nontrivial solution of (6) we must have the scalar equation

$$\det \underline{\mathbf{A}}(\lambda) = \Delta_{\underline{\mathbf{A}}}(\lambda) = |(\lambda \underline{\mathbf{I}} - \underline{\mathbf{A}})| = 0 \tag{7}$$

$\Delta_{\underline{\mathbf{A}}}(\lambda)$ is called the *characteristic polynomial* in λ of degree n, and equation (7) is called the *characteristic equation.* We have

$$\begin{aligned} \Delta_{\underline{\mathbf{A}}}(\lambda) = P(\lambda) &= \begin{bmatrix} \lambda - a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ -a_{n1} & \cdots & \lambda - a_{nn} \end{bmatrix} \\ &= \lambda^n + a_1 \lambda^{n-1} + \cdots + a_n \end{aligned} \tag{8}$$

According to the fundamental theorem of algebra (8) has n roots $\lambda_1, \ldots, \lambda_n$, not necessarily all distinct. These roots λ_i are the eigenvalues of $\underline{\mathbf{A}}$ belonging to the corresponding eigenvector $\underline{\boldsymbol{x}}_i$.

Elementary Symmetric Functions

Let

$$P(\lambda) = a_0 \lambda^n + a_1 \lambda^{n-1} + a_2 \lambda^{n-2} + \cdots + a_n = \prod_{i=1}^{n} (\lambda - \lambda_i) \tag{9}$$

Then

$$\begin{aligned} a_0 &= 1 \\ (-1) a_1 &= \sum_{i=1}^{n} \lambda_i \\ (-1)^2 a_2 &= \frac{1}{2!} \sum_{i,j=1}^{n}{}' \lambda_i \lambda_j \\ &\vdots \\ (-1)^m a_m &= \frac{1}{m!} \sum_{i_1, i_2, \ldots, i_m = 1}^{n}{}' \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_m} \\ &. \\ (-1)^n a_n &= \prod_{i=1}^{n} \lambda_i \end{aligned}$$

where a prime on the summation implies a sum only over distinct subscripts.

Two important quantities associated with $\underline{\mathbf{A}}$, its trace (also called *Spur*) and determinant, can then be expressed as

$$\operatorname{Tr} \underline{\mathbf{A}} = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i$$

$$\det \underline{\mathbf{A}} = \Delta_{\underline{\mathbf{A}}} = \Delta_{\underline{\mathbf{A}}}(0) = \prod_{i=1}^{n} \lambda_i$$

Generalized Eigenvectors of Multiplicity *k*

Let

$$(\lambda \underline{\mathbf{I}} - \underline{\mathbf{A}})^k \underline{\boldsymbol{x}} = \underline{\mathbf{0}}, \quad (\lambda \underline{\mathbf{I}} - \underline{\mathbf{A}})^{k-1} \neq \underline{\mathbf{0}} \qquad (k \leq n) \tag{10}$$

The vector $\underline{\boldsymbol{x}}$ is called a *generalized eigenvector* of $\underline{\mathbf{A}}$ of multiplicity k.

If the generalized eigenvector $\underline{\boldsymbol{x}}_i$ of order k belongs to the eigenvalue λ_i, then the chain of k generalized eigenvector $\{\underline{\boldsymbol{x}}_i, (\lambda \underline{\mathbf{I}} - \underline{\mathbf{A}})\underline{\boldsymbol{x}}_i, \ldots, (\lambda \underline{\mathbf{I}} - \underline{\mathbf{A}})^{k-1} \underline{\boldsymbol{x}}_i\}$ are linearly independent and can be utilized as k linearly independent eigenvectors of $\underline{\mathbf{A}}$.

Observe that:

1. Generalized eigenvectors of a matrix corresponding to different eigenvalues are linearly independent.
2. Eigenvalues of a Hermitian matrix are real, and eigenvectors corresponding to different eigenvalues are orthogonal. This result plays a important role in physics, particularly in quantum mechanics.

Norm of a Matrix

A *norm* of a matrix $\underline{\mathbf{A}}$, denoted by $\|\underline{\mathbf{A}}\|$, corresponding to the "greatest stretching" of vectors under its mapping. Three main useful norms are

$$\|\underline{\mathbf{A}}\|_m = \max_i \sum_j |a_{ij}|, \qquad \|\boldsymbol{x}\|_m = \max_i |\mathbf{x}_i| \qquad (m - \text{norm}) \tag{11a}$$

$$\|\underline{\mathbf{A}}\|_l = \max_j \sum_i |a_{ij}|, \qquad \|\boldsymbol{x}\|_l = \sum_j |\mathbf{x}_j| \qquad (l - \text{norm}) \tag{11b}$$

$$\|\underline{\mathbf{A}}\|_k = \left(\sum_{i,j} |a_{ij}|^k \right)^{1/k}, \qquad \|\boldsymbol{x}\|_k = \left[\sum_j |\mathbf{x}_j|^k \right]^{1/k} \qquad (k - \text{norm}) \tag{11c}$$

Geometric Series

For any matrix $\underline{\mathbf{A}}$,

$$\underline{\mathbf{I}} + \underline{\mathbf{A}} + \underline{\mathbf{A}}^2 + \cdots = \sum_{k=1}^{\infty} \underline{\mathbf{A}}^k = (\underline{\mathbf{I}} - \underline{\mathbf{A}})^{-1}, \qquad \|\underline{\mathbf{A}}\| < 1$$

$$\|(\underline{\mathbf{I}} - \underline{\mathbf{A}})^{-1}\| \leq (1 - \|\underline{\mathbf{A}}\|)^{-1}$$

If

$$f(\lambda) = \lambda^n + a_1 \lambda^{n-1} + \cdots + a_n$$

then

$$f(\mathbf{A}) = \underline{\mathbf{A}}^n + a_1 \underline{\mathbf{A}}^{n-1} + \cdots + a_n \underline{\mathbf{I}}$$

Eigenvalues of a Function of $\underline{\mathrm{A}}$

If λ_i is an eigenvalue of $\underline{\mathbf{A}}$ (denoted as $\underline{\mathbf{A}} \to \lambda_i$), then

$$\begin{aligned} \underline{\mathbf{A}}^{-1} &\longrightarrow \lambda_i^{-1}, & \underline{\mathbf{A}}^k &\longrightarrow \lambda_i^k \\ \mathbf{A}^{\mathrm{T}} &\longrightarrow \lambda_i, & f(\underline{\mathbf{A}}) &\longrightarrow f(\lambda_i) \\ & & \overline{\underline{\mathbf{A}}} &\longrightarrow \overline{\lambda}_i \end{aligned}$$

Sylvesters Theorem

For a quadratic form $[\underline{\boldsymbol{x}}^{\mathrm{T}} \underline{\mathbf{A}} x]$ to be positive definite it is necessary and sufficient that all the principal minors (along the main diagonal) of $\underline{\mathbf{A}}$ be positive.

DIAGONALIZATION OF MATRICES

An $n \times n$ matrix $\underline{\mathbf{A}}$ can be diagonalized if and only if it has n linearly independent eigenvectors. This is always possible if all of its n eigenvalues are all distinct. Then

$$\underline{\mathbf{A}}\underline{\boldsymbol{x}}_i = \lambda_i \underline{\boldsymbol{x}}_i \qquad (i = 1, \ldots, n)$$

where $\underline{\boldsymbol{x}}_i$ is the eigenvector belonging to λ_i; thus

$$\underline{\mathbf{A}}[\underline{\boldsymbol{x}}_1 \quad \cdots \quad \underline{\boldsymbol{x}}_n] = [\underline{\boldsymbol{x}}_1 \quad \cdots \quad \underline{\boldsymbol{x}}_n]\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

Let

$$\underline{\mathbf{P}} = [\underline{\boldsymbol{x}}_1 \quad \cdots \quad \underline{\boldsymbol{x}}_n] \qquad \text{(modal matrix)}$$

$$\underline{\boldsymbol{\Lambda}} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \qquad \text{(a diagonal matrix)}$$

Then

$$\underline{\mathbf{AP}} = \underline{\mathbf{P\Lambda}}, \qquad \underline{\mathbf{A}} = \underline{\mathbf{P}}^{-1}\underline{\mathbf{\Lambda P}}$$

In general two matrices $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$ are similar if one can find a nonsingular matrix $\underline{\mathbf{P}}$ such that

$$\underline{\mathbf{A}} = \underline{\mathbf{P}}^{-1}\underline{\mathbf{BP}}$$

Observe that:

1. Similar matrices $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$ have the same eigenvalues, equal determinants, and the same characteristic polynomials:

$$\underline{\mathbf{A}} \longrightarrow \lambda_i \quad \text{also means} \quad \underline{\mathbf{B}} \longrightarrow \lambda_i$$

$$\Delta_{\underline{\mathbf{A}}} = \Delta_{\underline{\mathbf{B}}}$$

$$\Delta_{\underline{\mathbf{A}}}(\lambda) = \Delta_{\underline{\mathbf{B}}}(\lambda)$$

2. Every Hermitian matrix is diagonalizable, and its modal matrix $\underline{\mathbf{P}}$ is unitary:

$$\underline{\mathbf{P}}^*\underline{\mathbf{P}} = \underline{\mathbf{I}}$$

3. If $\underline{\mathbf{A}}_c$ is a *companion matrix*

$$\underline{\mathbf{A}}_c = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}$$

then

$$\Delta_{\underline{\mathbf{A}}_c}(\lambda) = a_0\lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + a_n = \prod_{i=1}^{n}(\lambda - \lambda_i)$$

If the eigenvalues of $\underline{\mathbf{A}}_c$ are distinct, then

$$\underline{\mathbf{P}} = \underline{\mathbf{V}}_n = \begin{bmatrix} 1 & \cdots & 1 \\ \lambda_1 & \cdots & \lambda_n \\ \vdots & & \vdots \\ \lambda_1^{n-1} & \cdots & \lambda_n^{n-1} \end{bmatrix}$$

the *Vandermonde matrix,* is nonsingular, and

$$\det \underline{\mathbf{V}}_n = \prod_{i=2}^{n}\left(\prod_{j=1}^{i-1}(\lambda_i - \lambda_j)\right)$$

4. If a matrix has eigenvalues of multiplicity greater than one, then for diagonalization these eigenvalues should induce the same number of linearly independent eigenvectors as the multiplicity; otherwise the similarity transformation produces not the diagonal but the *Jordan* form. As discussed earlier, we produce a set of generalized eigenvectors for the same eigenvalues, which are linearly independent and transform a matrix into *Jordan canonical form.*

THE JORDAN CANONICAL FORM

When the characteristic polynomial of a matrix has multiple roots, it may not be possible to diagonalize the matrix. Nevertheless, it is possible to transform the matrix into a canonical form called *Jordan canonical form,* via similarity transformations. We shall limit ourself to the procedure for arriving at this canonical form.

Let

$$\Delta_{\underline{\mathbf{A}}}(\lambda) = |(\lambda\underline{\mathbf{I}} - \underline{\mathbf{A}})| = \prod_{i=1}^{r}(\lambda - \lambda_i)^{k_i}, \qquad \sum_{i=1}^{r} k_i = n$$

The matrix $\underline{\mathbf{A}}$ can be transformed to a matrix $\underline{\mathbf{J}}$ with canonical superboxes $\underline{\mathbf{J}}_i$ $(i = 1, \ldots, r)$: $\underline{\mathbf{J}} = \dot{\sum}_{i=1}^{r} \underline{\mathbf{J}}_i$. These superboxes $\underline{\mathbf{J}}_i$ are further divided into boxes $\underline{\mathbf{J}}_{ij}$ $(j = 1, \ldots, r_i;\ r_i \le k_i)$: $\underline{\mathbf{J}}_i = \dot{\sum}_{j=1}^{r_i} \underline{\mathbf{J}}_{ij}$. Namely,

$$\underline{\mathbf{J}} = \begin{bmatrix} \underline{\mathbf{J}}_1 & & & & \\ & \ddots & & \underline{\mathbf{0}} & \\ & & \underline{\mathbf{J}}_i & & \\ & \underline{\mathbf{0}} & & \ddots & \\ & & & & \underline{\mathbf{J}}_r \end{bmatrix}$$

$$\underline{\mathbf{J}}_i = \begin{bmatrix} \underline{\mathbf{J}}_{i1} & & & & \\ & \ddots & & & \\ & & \underline{\mathbf{J}}_{ij} & & \\ & & & \ddots & \\ & & & & \underline{\mathbf{J}}_{ir_i} \end{bmatrix}$$

The dimension of the box $\underline{\mathbf{J}}_i$ is $k_i \times k_i$ ($\sum_{i=1}^{r} k_i = n$). The dimension of the ijth box within $\underline{\mathbf{J}}_i$ is $l_{ij} \times l_{ij}$ ($\sum_{j=1}^{r_i} l_{ij} = k_i$).

The following procedure is used to determine

$$\begin{array}{ll} \lambda_i & (i = 1, \ldots, r) \\ k_i & (i = 1, \ldots, r) \\ l_{ij} & (i = 1, \ldots, r; \quad j = 1, \ldots, r_i) \end{array}$$

Step 1. Determine the characteristic polynomial $\Delta_{\underline{\mathbf{A}}}(\lambda)$ of $\underline{\mathbf{A}}$ [with order $\det(\lambda\underline{\mathbf{I}} - \underline{\mathbf{A}})$] and its roots λ_i ($i = 1, \ldots, r$).

Step 2. Determine the multiplicity indices k_i ($i = 1, \ldots, r$) such that $(\lambda - \lambda_i)^{k_i}$ is a factor of $\Delta_{\underline{\mathbf{A}}}(\lambda)$ but $(\lambda - \lambda_i)^{k_i+1}$ is not.

Step 3. Consider all the minors of order $n - j$ ($i = 1, \ldots, r; j = 1, \ldots, r_i$). If the greatest common divisor (gcd) of any one of these minors contains a factor $(\lambda - \lambda_i)^{k_{i,j}}$ but not $(\lambda - \lambda_i)^{k_{i,j+1}}$, then $l_{ij} = k_{i,j-1} - k_{i,j}$ ($i = 1, \ldots, r; j = 1, \ldots, r_i; k_{i,0} = k_i$)

The minors of order $n - k_i - 1$ contain no factor $\lambda - \lambda_i$. each Jordan subbox $\underline{\mathbf{J}}_{ij}$ appears as

$$\underline{\mathbf{J}}_{ij} = \begin{bmatrix} \lambda_i & 1 & 0 & \ldots & 0 \\ 0 & \lambda_i & 1 & \ldots & 0 \\ 0 & 0 & \lambda_i & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix}, \qquad l_{ij} \times l_{ij}$$

In practice we use the method of *elementary divisors* to arrive at the structure of Jordan canonical form. We transform $\underline{\mathbf{A}}$ into Jordan form $\underline{\mathbf{J}}$ via a similarity transformation $\underline{\mathbf{P}}$:

$$\underline{\mathbf{A}} = \underline{\mathbf{P}}\,\underline{\mathbf{J}}\,\underline{\mathbf{P}}^{-1}, \qquad \underline{\mathbf{A}}^k = \underline{\mathbf{P}}\,\underline{\mathbf{J}}^k\,\underline{\mathbf{P}}^{-1}$$

The modal matrix $\underline{\mathbf{P}}$ is made up of the chain of generalized eigenvectors

$$\underline{\boldsymbol{x}}_{ij},\ (\lambda_i\underline{\mathbf{I}} - \underline{\mathbf{A}}\,\underline{\boldsymbol{x}}_{ij}),\ \ldots,\ (\lambda_i\underline{\mathbf{I}} - \underline{\mathbf{A}}^{r_i-1})\underline{\boldsymbol{x}}_{ij} \qquad (i = 1, \ldots, r, \quad j = 1, \ldots, r_i)$$

Every square matrix $\underline{\mathbf{A}}$ can be transformed into Jordan form.

The *minimal polynomial* of $\underline{\mathbf{J}}$ (or $\underline{\mathbf{A}}$) is $\underline{\mathbf{P}}_m(\lambda) = \prod_{i=1}^{r^{(i)}}(\lambda - \lambda_i)^{l_{i1}}$, where l_{i1} is the size of the largest Jordan subbox associated with λ_i

Using Dg for block-diagonal matrices, we have

$$\begin{aligned} \underline{\mathbf{J}} &= \mathrm{Dg}[\underline{\mathbf{J}}_1, \underline{\mathbf{J}}_2, \ldots, \underline{\mathbf{J}}_i, \ldots, \underline{\mathbf{J}}_n] \\ \underline{\mathbf{J}}_i &= \mathrm{Dg}[\underline{\mathbf{J}}_{i1}, \underline{\mathbf{J}}_{ij}, \ldots, \underline{\mathbf{J}}_{ir_i}] \qquad (i = 1, \ldots, r) \end{aligned}$$

$(\lambda - \lambda_i)^{l_{ij}}$ are known as *elementary divisors* of $\underline{\mathbf{A}}$ ($i = 1, \ldots, r, j = 1, \ldots, r_i$)

$(\lambda_i\underline{\mathbf{I}} - \underline{\mathbf{A}})$ acts as an *elevator matrix.* It raises an eigenvector to the next higher eigenvector until the last vector in the chain is reached, and then annihilates it.

CAYLEY–HAMILTON THEOREM

This remarkable theorem states: "A matrix satisfies its own characteristic equation". Specifically, if

$$\Delta_{\underline{\mathbf{A}}}(\lambda) = p(\lambda) = |(\lambda I - \underline{\mathbf{A}})| = \lambda^n + a_1\lambda^{n-1} + \cdots$$

where $\underline{\mathbf{A}}$ is an $n \times n$ matrix, then $\underline{\mathbf{A}x} \in \mathbf{E}_n$ when $\underline{\boldsymbol{x}} \in \mathbf{E}_n$. Then $p(\underline{\mathbf{A}})\underline{\boldsymbol{x}} = \underline{\mathbf{0}}$, implying

$$p(\underline{\mathbf{A}}) \equiv \underline{\mathbf{A}}^n + a_1\underline{\mathbf{A}}^{n-1} + \cdots + a_n\underline{\mathbf{I}} = \underline{\mathbf{0}}$$

Proof:

$$[\underline{\mathbf{A}}(\lambda)]^{-1} = (\lambda\underline{\mathbf{I}} - \underline{\mathbf{A}})^{-1} = \frac{1}{\Delta_{\underline{\mathbf{A}}}(\lambda)}\underline{\mathbf{A}}^*(\lambda) = \frac{1}{p(\lambda)}\underline{\mathbf{B}}(\lambda) \tag{12}$$

where $\underline{\mathbf{B}}(\lambda) = \underline{\mathbf{A}}^*(\lambda)$ is a polynomial matrix in λ of degree $(n - 1)$

$$\underline{\mathbf{B}}(\lambda) \equiv \underline{\mathbf{B}}_1\lambda^{n-1} + \underline{\mathbf{B}}_2\lambda^{n-2} + \cdots + \underline{\mathbf{B}}_n = \sum_{i=0}^{n-1}\underline{\mathbf{B}}_{n-i}\lambda^i \tag{13}$$

From (12) and (13),

$$(\lambda\underline{\mathbf{I}} - \underline{\mathbf{A}})\underline{\mathbf{B}}(\lambda) = p(\lambda)\underline{\mathbf{I}}$$

Equating powers of λ on both sides,

$$\begin{aligned} \underline{\mathbf{0}} - \underline{\mathbf{AB}}_{\mathbf{n}} &= a_n\underline{\mathbf{I}} \\ \underline{\mathbf{B}}_{\mathbf{n}} - \underline{\mathbf{AB}}_{\mathbf{n-1}} &= a_{n-1}\underline{\mathbf{I}} \\ &\vdots \\ \underline{\mathbf{B}}_2 - \underline{\mathbf{AB}}_1 &= a_1\underline{\mathbf{I}} \\ \underline{\mathbf{B}}_1 - \underline{\mathbf{0}} &= \underline{\mathbf{I}} \end{aligned} \tag{14}$$

Multiplying these equations by $\underline{\mathbf{I}}, \underline{\mathbf{A}}, \ldots, \underline{\mathbf{A}}^n$ respectively and adding,

$$\underline{\mathbf{0}} \equiv \underline{\mathbf{A}}^n + a_1\underline{\mathbf{A}}^{n-1} + \cdots + a_n\underline{\mathbf{I}} \equiv p(\underline{\mathbf{A}}) \tag{15}$$

This theorem is very significant in system theory, for it implies that all matrices $\underline{\mathbf{A}}^k$ ($k \geq n$) can be expressed as a linear combination of matrices $\underline{\mathbf{A}}^j$ ($j < n$).

COMPUTATION OF A POLYNOMIAL FUNCTION OF THE MATRIX $\underline{\mathbf{A}}$

Let

$$F(\underline{\mathbf{A}}) = \sum_{k=1}^{m} c_k\underline{\mathbf{A}}^k, \qquad m \geq n \tag{16}$$

where λ_i ($i = 1, \ldots, n$) are eigenvalues of $\underline{\mathbf{A}}$. Then

$$\frac{F(\lambda)}{\Delta_{\underline{\mathbf{A}}}(\lambda)} = Q(\lambda) + \frac{R(\lambda)}{\Delta_{\underline{\mathbf{A}}}(\lambda)} \tag{17}$$

by long division, where $R(\lambda)$ is polynomial of degree less than n. Then

$$F(\lambda) = Q(\lambda)\Delta_{\underline{\mathbf{A}}}(\lambda) + R(\lambda)$$
$$F(\lambda_i) = R(\lambda_i), \quad \Delta_{\underline{\mathbf{A}}}(\lambda_i) = 0, \qquad i = 1, \ldots, n$$

Compute the coefficients of $R(\lambda_i)$ from $F(\lambda_i)$. If λ_i is an eigenvalue of multiplicity m_i, then not only does $\Delta_{\underline{\mathbf{A}}}(\lambda_i) = 0$, but the first $m_i - 1$ derivatives of $\Delta_{\underline{\mathbf{A}}}(\lambda_i)$ with respect to λ computed at $\lambda = \lambda_i$ also vanish, resulting in

$$\frac{d^k}{d\lambda^k}F(\lambda)\bigg|_{\lambda=\lambda_i} = \frac{d^k}{d\lambda^k}R(\lambda)\bigg|_{\lambda=\lambda_i} \qquad (k = 0, 1, \ldots, m_{i-1})$$

For the matrix exponential we have

$$e^{\underline{\mathbf{A}}t} = \sum_{k=0}^{\infty} \frac{\underline{\mathbf{A}}^k t^k}{k!}$$

(not generally recommended for computing),

$$e^{\underline{\mathbf{A}}t} = \sum_{i=0}^{n-1} \alpha_i(t)\underline{\mathbf{A}}^i, \qquad \alpha_0(0) = 1, \quad \alpha_i(0) = 0\ (i = 2, \ldots)$$

and

$$e^{(\underline{\mathbf{A}}+\underline{\mathbf{B}})t} = e^{\underline{\mathbf{A}}t}e^{\underline{\mathbf{B}}t} \qquad \text{if} \quad \underline{\mathbf{AB}} = \underline{\mathbf{BA}}$$

For a series

$$g(\lambda) = \sum_{k=0}^{\infty} g_k\lambda^k$$

For a series $|\lambda| \le r \le 1$ implies convergence.

$$g(\underline{\mathbf{A}}) = \sum_{k=0}^{\infty} g_k\underline{\mathbf{A}}^k, \qquad \underline{\mathbf{A}} \text{ with eigenvalues } \lambda_i$$

$|\lambda_i| \le r \le 1$ ($i = 1, 2 \ldots, n$) implies convergence.

From complex integration

$$f(\boldsymbol{A}) = \frac{1}{2\pi j}\oint_c (\lambda\underline{\mathbf{I}} - \underline{\mathbf{A}})^{-1} f(\lambda)\,d\lambda, \qquad |\lambda_{\mathbf{i}}| \le c$$

For

$$\underline{\mathbf{J}} = \begin{bmatrix} \lambda_1 & 1 & & \\ & \lambda_1 & 1 & \\ & & \lambda_1 & \ddots \\ & & & \ddots \end{bmatrix}$$

We obtain

$$f(\underline{\mathbf{J}}) = \begin{bmatrix} f(\lambda_1) & \dfrac{f'(\lambda_1)}{1!} & \dfrac{f''(\lambda_1)}{2!} & \cdots \\ & f(\lambda_1) & \dfrac{f'(\lambda_1)}{1!} & \ddots \\ & & f(\lambda_1) & \\ & & & \ddots \end{bmatrix}$$

$$e^{\underline{\mathbf{J}}t} = \begin{bmatrix} e^{\lambda_1 t} & te^{\lambda_1 t} & t^2e^{\lambda_1 t} & \cdots \\ & e^{\lambda_1 t} & te^{\lambda_1 t} & \\ & & e^{\lambda_1 t} & \\ & & & \ddots \end{bmatrix}$$

$$\underline{\mathbf{\Lambda}} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

$$f(\underline{\mathbf{\Lambda}}) = \begin{bmatrix} f(\lambda_1) & & & \\ & f(\lambda_2) & & \\ & & \ddots & \\ & & & f(\lambda_n) \end{bmatrix}$$

$$e^{\underline{\mathbf{\Lambda}}t} = \begin{bmatrix} e^{\lambda_1 t} & & & \\ & e^{\lambda_2 t} & & \\ & & \ddots & \\ & & & e^{\lambda_n t} \end{bmatrix}$$

$$\underline{\mathbf{A}} = \underline{\mathbf{P}}^{-1}\underline{\mathbf{\Lambda}}\underline{\mathbf{P}}, \qquad f(\underline{\mathbf{A}}) = \underline{\mathbf{P}}^{-1}\mathbf{f}(\underline{\mathbf{\Lambda}})\underline{\mathbf{P}}$$
$$\underline{\mathbf{A}} = \underline{\mathbf{S}}^{-1}\underline{\mathbf{JS}}, \qquad f(\underline{\mathbf{J}}) = \underline{\mathbf{S}}^{-1}\mathbf{f}(\underline{\mathbf{J}})\underline{\mathbf{S}}$$

If an $n \times n$ matrix $\underline{\mathbf{A}}$ has minimal polynomial of degree $m < n$, then

$$e^{\underline{\mathbf{A}}t} = \alpha_0(t)\underline{\mathbf{I}} + \alpha_1(t)\underline{\mathbf{A}} \cdots \alpha_{m-1}(t)\underline{\mathbf{A}}^{m-1}$$

where $\alpha_j(t)$ ($j = 0, \ldots m - 1$) can be computed from the eigenvalues, distinct or multiple.

A matrix $\underline{\mathbf{A}}$ is called *stable* if the real parts of its eigenvalues λ_i ($i = 1, \ldots, n$) are negative.

For the *Riccati equation*

$$\underline{\mathbf{AS}} + \underline{\mathbf{SA}}^{\mathrm{T}} = -\underline{\mathbf{Q}}, \qquad \underline{\mathbf{S}}, \underline{\mathbf{Q}} \text{ symmetric}$$

we have the solution

$$\underline{\mathbf{S}} = \int_0^{\infty} \underline{\mathbf{e}}^{\underline{\mathbf{A}}t}\underline{\mathbf{Q}}\,\underline{\mathbf{e}}^{\underline{\mathbf{A}}^{\mathrm{T}}t}\,dt$$

COMPANION MATRICES

A companion matrix has the form

$$\underline{\mathbf{A}}_c = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \dots & -a_1 \end{bmatrix}$$

$$\triangle_{\underline{\mathbf{A}}_c}(\lambda) = |\lambda \underline{\mathbf{I}} - \underline{\mathbf{A}}_c| = \lambda^n + a_1\lambda^{n-1} + \dots + a_n = p(\lambda)$$

The polynomial $\triangle_{\underline{\mathbf{A}}_c}(\lambda)$ can be associated with the companion matrix $\underline{\mathbf{A}}_c$.

The following special properties are associated with companion matrices:

1. If λ_i is an eigenvalue of multiplicity one (distinct), the associated eigenvector is

$$\underline{\boldsymbol{p}}_i^{\mathrm{T}} = [1 \quad \lambda_i \quad \lambda_i^2 \quad \cdots \quad \lambda_i^{n-1}]$$

2. If λ_i is an eigenvalue of multiplicity $k \le n$ [$(\lambda - \lambda_i)^k$ is, factor of $\triangle_{\underline{\mathbf{A}}_c}(\lambda)$ but $(\lambda - \lambda_i)^{k+1}$ is not], then this eigenvalue has k *generalized* eigenvectors and one and only one Jordan block of size $k \times k$ belonging to the eigenvalue λ_i. This implies that companion matrix is nonderogatory, and we have

$$\begin{aligned} \underline{\boldsymbol{p}}_{i1}^{\mathrm{T}} &= [1 \quad \lambda_1 \quad \lambda_1^2 \quad \dots \quad \lambda_1^{n-1}] \\ \underline{\boldsymbol{p}}_{i1}^{\mathrm{T}} &= [0 \quad 1 \quad 2\lambda_1 \quad \dots \quad (n-1)\lambda_1^{n-2}] \\ &\vdots \\ \underline{\boldsymbol{p}}_{in}^{\mathrm{T}} &= \left[0 \quad 0 \quad 0 \quad \dots \quad \prod_{j=1}^{k-1}(n-j)\lambda_i^{n-k}\right] \end{aligned}$$

3. An nth-order Linear differential equation.

$$x^{(n)} + a_1 x^{(n-1)} \dots + a_{n-1}\dot{x} + a_n x = 0$$

can be written as

$$\underline{\dot{\boldsymbol{x}}} = \underline{\mathbf{A}}_c \underline{\boldsymbol{x}}$$

where $\underline{\mathbf{A}}_c$ is a companion matrix.

4. A matrix $\underline{\mathbf{A}}$ is similar to the companion matrix $\underline{\mathbf{A}}_c$ [of its characteristic polynomial $\triangle_{\underline{\mathbf{A}}}(\lambda)$] if and only if the minimal and the characteristic polynomial of $\underline{\mathbf{A}}$ are the same. This implies $\underline{\mathbf{A}}$ is nonderogatory.

CHOLESKY DECOMPOSITION (ALSO KNOWN AS LU DECOMPOSITION)

This is a convenient scheme for machine computation of $\underline{\mathbf{A}}\underline{x} = \underline{\boldsymbol{b}}$, where $\underline{\mathbf{A}}$ is $n \times n$ of rank n, and $\underline{\boldsymbol{b}}$ is $n \times 1$.

Write $\underline{\mathbf{A}} = (a_{ij})$ as $\underline{\mathbf{A}} = \underline{\mathbf{LU}}$, where $\underline{\mathbf{L}}$ is lower triangular and $\underline{\mathbf{U}}$ is upper triangular:

$$\underline{\mathbf{L}} = \begin{bmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \dots & l_{nn} \end{bmatrix}, \qquad \underline{\mathbf{U}} = \begin{bmatrix} 1 & c_{12} & \dots & c_{1n} \\ 0 & 1 & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

l_{ij} and u_{ij} are computed as

$$\begin{aligned} l_{i1} &= a_{i1},\ u_{1j} = \frac{a_{1j}}{l_{11}} && (i = 1, \dots, n;\ j = 1, \dots, n) \\ l_{ij} &= a_{ij} - \sum_{k=1}^{j-1}(l_{ik}u_{kj}) && (i \ge j < 1) \\ u_{ij} &= \frac{1}{l_{ii}}\left(a_{ij} - \sum_{k=1}^{i-1} l_{ik}u_{kj}\right) && (j > i > 1), \qquad u_{ii} = 1 \end{aligned}$$

Knowing $\underline{\mathbf{L}}$ and $\underline{\mathbf{U}}$, solve the two sets of equations

$$\underline{\mathbf{Ux}} = \underline{\boldsymbol{y}}, \qquad \underline{\mathbf{Ly}} = \underline{\mathbf{b}}$$

where $\underline{\mathbf{A}}$ is symmetric. The computation of $\underline{\mathbf{U}}$ is simplified as

$$u_{ij} = \frac{1}{l_{ii}} l_{ji} \qquad (i \le j)$$

JACOBI AND GAUSS–SEIDEL METHODS

When all the diagonal elements of $\underline{\mathbf{A}}$ are nonzero, we can decompose $\underline{\mathbf{A}}$ as

$$\underline{\mathbf{A}} = \underline{\mathbf{L}} + \underline{\mathbf{D}} + \underline{\mathbf{U}}$$

with

$\underline{\mathbf{U}}$ upper triangular with zero on the diagonal
$\underline{\mathbf{L}}$ lower triangular with zero on the diagonal
$\underline{\mathbf{D}}$ diagonal

The iterative schemes for solving $\underline{\mathbf{A}}\underline{x} = \underline{\boldsymbol{b}}$, with initial guess $\underline{\boldsymbol{x}}^{(0)}$, are

$$\underline{\boldsymbol{x}}^{(i+1)} = \underline{\mathbf{D}}^{-1}\underline{\boldsymbol{b}} - \underline{\mathbf{D}}^{-1}(\underline{\mathbf{L}} + \underline{\mathbf{U}})\underline{\boldsymbol{x}}^{(i)} \quad (i = 0, 1, 2, 3, \dots) \quad \text{(Jacobi)}$$

$$\underline{\boldsymbol{x}}^{(i+1)} = (\underline{\mathbf{L}} + \underline{\mathbf{D}})^{-1}\underline{\boldsymbol{b}} - (\underline{\mathbf{L}} + \underline{\mathbf{D}})^{-1}\underline{\mathbf{U}\boldsymbol{x}}^{(i)} \quad \text{(Gauss–Seidel)}$$

LEAST-SQUARES BEST-FIT PROBLEM (ALSO KNOWN AS THE PSEUDOINVERSE PROBLEM)

Given:

$\underline{\mathbf{A}}\underline{x} = \underline{\boldsymbol{b}}$, subject to the condition $\underline{\mathbf{B}}\underline{x} = 0$
$\underline{\mathbf{A}}$ $n \times p$, rank $\underline{\mathbf{A}} = p$;
$\underline{\mathbf{B}}$ $r \times p$, rank $\underline{\mathbf{B}} = r$ $(r \le p \le n)$

we are to solve for $\underline{\boldsymbol{x}}$.

Define $(\underline{\mathbf{A}}^{\mathrm{T}}\underline{\mathbf{A}})^{-1}\underline{\mathbf{A}}^{\mathrm{T}} = \underline{\mathbf{A}}^{+}$ (*pseudoinverse*), $(\underline{\mathbf{A}}^{\mathrm{T}}\underline{\mathbf{A}})^{-1}\underline{\mathbf{B}}^{\mathrm{T}} = \underline{\mathbf{B}}_1$. The least-squares solution is

$$\hat{\underline{\boldsymbol{x}}} = [\underline{\mathbf{A}}^{+} - \underline{\mathbf{B}}_1(\underline{\mathbf{B}}\underline{\mathbf{B}}_1)^{-1}\underline{\mathbf{B}}\underline{\mathbf{A}}^{+}]\underline{\boldsymbol{b}}$$

HERMITIAN (OR SYMMETRIC REAL) MATRICES AND DEFINITE FUNCTIONS

1. Let $\underline{\mathbf{A}}$ be Hermitian. If for all $\underline{\boldsymbol{x}}$ we have

$$\overline{\underline{\mathbf{x}}}^{\mathrm{T}}\underline{\mathbf{A}\boldsymbol{x}} = \underline{\boldsymbol{x}}^{*}\underline{\mathbf{A}\boldsymbol{x}} > 0$$

then $\underline{\mathbf{A}}$ is *positive definite.* If for all $\underline{\boldsymbol{x}}$ we have

$$\underline{\boldsymbol{x}}^{*}\underline{\mathbf{A}\boldsymbol{x}} \geq 0$$

then $\underline{\mathbf{A}}$ is *positive semidefinite.* If for some $\underline{\boldsymbol{x}}$ we have $\underline{\boldsymbol{x}^{*}\mathrm{A}x} > 0$ and for other $\underline{\boldsymbol{x}}$ we have $\underline{\boldsymbol{x}^{*}\mathrm{A}x} < 0$, then $\underline{\mathbf{A}}$ is *indefinite.*

2. Hermitian (or symmetric real) matrices have distinct eigenvalues, and their eigenvectors are mutually orthogonal. If in addition the matrix is positive definite, then all its eigenvalues are necessarily positive. If λ_1 is the largest and λ_n is the smallest eigenvalue of $\underline{\mathbf{A}}$, then

$$\lambda_n(\underline{\boldsymbol{x}}^{*}\underline{\boldsymbol{x}}) \leq (\underline{\boldsymbol{x}}^{*}\underline{\mathbf{A}\boldsymbol{x}}) \leq \lambda_1(\underline{\boldsymbol{x}}^{*}\underline{\boldsymbol{x}})$$

In fact, any Hermitian (or real symmetric) matrix can be diagonalized by a similarity transformation $\underline{\mathbf{P}}$ in which all the columns are mutually orthonormal (called a *unitary* matrix). All the eigenvalues of a Hermitian (or symmetric real) positive definite matrix are strictly positive. The coefficients of the characteristic polynomial $|(\lambda\underline{\mathbf{I}} - \underline{\mathbf{A}})|$ of a positive definite matrix alternate in sign, yielding a necessary and sufficient condition for positive definiteness. The principal diagonal minors of the determinant of a positive definite Hermitian matrix must be strictly positive. If two Hermitian matrices commute, then they can be simultaneously diagonalized.

3. For the *simultaneous diagonalization* of two real matrices $\underline{\mathbf{R}} > 0$ and $\underline{\mathbf{Q}} \leq 0$, choose a nonsingular $\underline{\mathbf{W}}$, the *square-root matrix of* $\underline{\mathbf{R}}$, such that $\underline{\mathbf{R}} = \underline{\mathbf{W}}^{\mathrm{T}}\underline{\mathbf{W}}$. Choose an orthogonal matrix $\underline{\mathbf{O}}$ such that $\underline{\mathbf{O}}^{\mathrm{T}}\underline{\mathbf{W}}^{\mathrm{T}}\underline{\mathbf{Q}\mathbf{W}\mathbf{O}} = \underline{\mathbf{D}}$ ($\underline{\mathbf{D}} \leq 0$ is a diagonal matrix).

4. *Liapunov Stability Theorem.* Given an $n \times n$ real matrix $\underline{\mathbf{A}}$ with eigenvalues λ_i, if there exists a matrix $\underline{\mathbf{S}} \geq 0$ such that $\underline{\mathbf{A}}^{\mathrm{T}}\underline{\mathbf{S}} + \underline{\mathbf{S}\mathbf{A}} \leq 0$, then Re $\lambda_i < 0$ ($i = 1, \ldots, n$).

SOME USEFUL FACTS AND IDENTITIES

1. $(\underline{\mathbf{A}}^{-1} - \underline{\mathbf{B}}^{-1})^{-1} = \underline{\mathbf{A}} - \underline{\mathbf{C}\mathbf{A}}$, where $\underline{\mathbf{C}} = (\underline{\mathbf{A}} - \underline{\mathbf{B}})^{-1}$.
2. $(\underline{\mathbf{I}} - \underline{\mathbf{A}\mathbf{B}})^{-1} = \underline{\mathbf{I}} - \underline{\mathbf{A}}(\underline{\mathbf{I}} + \underline{\mathbf{B}\mathbf{A}})^{-1}\underline{\mathbf{B}}$, $\underline{\mathbf{B}\mathbf{A}}$ nonsingular (*Woodbury's form*).

If $\underline{\mathbf{B}} = \underline{\boldsymbol{x}}$, an $n \times 1$ vector, and $\underline{\mathbf{A}} = \underline{\boldsymbol{y}}^{\mathrm{T}}$, a $1 \times n$ column vector, then

$$(\underline{\mathbf{I}} + \underline{\boldsymbol{x}\boldsymbol{y}}^{\mathrm{T}})^{-1} = \underline{\mathbf{I}} - \frac{1}{\beta}\underline{\boldsymbol{x}\boldsymbol{y}}^{\mathrm{T}}, \qquad \beta = (1 + \underline{\mathbf{x}}^{\mathrm{T}}\underline{\mathbf{y}})$$

(the *Sherman–Morrison formula*),

$$(\underline{\mathbf{A}} + \underline{\boldsymbol{x}\boldsymbol{y}}^{\mathrm{T}})^{-1} = \underline{\mathbf{A}}^{-1} - \frac{1}{\alpha}\underline{\mathbf{A}}^{-1}\underline{\boldsymbol{x}\boldsymbol{y}}^{\mathrm{T}}\underline{\mathbf{A}}^{-1}, \qquad \alpha = 1 + T_r(\underline{\boldsymbol{x}\boldsymbol{y}}^{\mathrm{T}}\underline{\mathbf{A}}^{-1})$$

3. Suppose $\underline{\mathbf{A}} = \sum_{i=1}^{n} \lambda_i \underline{\mathbf{x}}_i\underline{\mathbf{y}}_i^{\mathrm{T}}$ is an $n \times n$ real matrix with distinct eigenvalues λ_i ($i = 1, \ldots, n$), $\underline{\mathbf{x}}_i$ the corresponding eigenvectors of $\underline{\mathbf{A}}$, and $\underline{\mathbf{y}}_i$ the corresponding eigenvectors of $\underline{\mathbf{A}}^{\mathrm{T}}$. If $\underline{\mathbf{A}}$ is Hermitian, then we have $\underline{\boldsymbol{y}}_i = \underline{\boldsymbol{x}}_i^{\mathrm{T}} = \underline{\boldsymbol{x}}_i^{*}$.
4. $\underline{\mathbf{A}} = \underline{\boldsymbol{x}\boldsymbol{y}}^{\mathrm{T}}$ implies $\underline{\mathbf{A}}$ is of rank one.
5. *Gerschgorin Circles.* Given an $n \times n$ nonsingular matrix $\underline{\mathbf{A}} = (a_{ij})$ with eigenvalues λ_k ($k = 1, \ldots, n$), then

$$|a_{ij}| > \sum_{i \neq j} |a_{ij}| \qquad (i = 1, \ldots, n)$$

$$|\lambda_k - a_{ii}| \leq \sum_{i \neq j} (a_{ij}) \qquad \text{for at least one } k, \quad i = 1, \ldots, n$$

6. *Bordering Matrices.* These matrices are useful in sequential filtering problems. Let

$$\underline{\tilde{\mathbf{A}}} = \begin{bmatrix} \underline{\mathbf{A}} & \underline{\boldsymbol{x}} \\ \underline{\boldsymbol{y}}^{\mathrm{T}} & \alpha \end{bmatrix}$$

Then

$$\underline{\tilde{\mathbf{A}}}^{-1} = \begin{bmatrix} \left(\underline{\mathbf{A}} - \frac{1}{\alpha}\underline{\boldsymbol{x}\boldsymbol{y}}^{\mathrm{T}}\right)^{-1} & -\frac{1}{\beta}\underline{\mathbf{A}}^{-1}\underline{\boldsymbol{x}} \\ -\frac{1}{\beta}\underline{\boldsymbol{y}}^{\mathrm{T}}\underline{\mathbf{A}}^{-1} & -\frac{1}{\beta} \end{bmatrix}, \qquad \beta = \alpha - \underline{\boldsymbol{y}}^{\mathrm{T}}\underline{\mathbf{A}}^{-1}\underline{\boldsymbol{x}}$$

If $\underline{\mathbf{A}}$ is Hermitian (diagonalizable, $\underline{\mathbf{A}} = \underline{\mathbf{P}\Lambda\mathbf{P}^{*}}$, $\underline{\mathbf{P}}$ unitary) and $\underline{\boldsymbol{y}} = \underline{\boldsymbol{x}}$, then the eigenvalues $\tilde{\lambda}$ of $\underline{\tilde{\mathbf{A}}}$ are computed from

$$\underline{\boldsymbol{x}}^{*}\underline{\mathbf{P}}(\tilde{\lambda}\underline{\mathbf{I}} - \underline{\Lambda})^{-1}\underline{\mathbf{P}}^{*}\underline{\boldsymbol{x}} = \tilde{\lambda} - \alpha \qquad (\underline{\tilde{\mathbf{A}}} \text{ is also Hermitian})$$

If $\underline{\mathbf{A}} > 0$ and $\alpha > \underline{\boldsymbol{y}}^{\mathrm{T}}\underline{\mathbf{A}}^{-1}\underline{\boldsymbol{x}}$, $\underline{\boldsymbol{y}} = \underline{\boldsymbol{x}}$, then $\underline{\tilde{\mathbf{A}}} > 0$.

7. *Kronecker Product.* For $m \times n$ $\underline{\mathbf{A}}$ and $p \times q$ $\underline{\mathbf{B}}$,

$$\underline{\mathbf{A}} \otimes \underline{\mathbf{B}} = \begin{bmatrix} a_{11}\underline{\mathbf{B}} & \ldots & a_{1n}\underline{\mathbf{B}} \\ \vdots & & \vdots \\ a_{m1}\underline{\mathbf{B}} & \ldots & a_{mn}\underline{\mathbf{B}} \end{bmatrix}$$

is an $mp \times nq$ matrix called the *Kronecker product.* There are mn blocks of this matrix, and the ijth block is $a_{ij}\underline{\mathbf{B}}$. We have

$$(\underline{\mathbf{A}} \otimes \underline{\mathbf{B}})(\underline{\mathbf{C}} \otimes \underline{\mathbf{D}}) = (\underline{\mathbf{A}\mathbf{C}} \otimes \underline{\mathbf{B}\mathbf{C}})$$

provided $\underline{\mathbf{A}\mathbf{C}}$ and $\underline{\mathbf{B}\mathbf{D}}$ exist

Furthermore,

$$(\underline{\mathbf{A}} \otimes \underline{\mathbf{B}})^{\mathrm{T}} = (\underline{\mathbf{A}}^{\mathrm{T}} \otimes \underline{\mathbf{B}}^{\mathrm{T}}), \qquad (\underline{\mathbf{A}} \otimes \underline{\mathbf{B}})^{-1} = (\underline{\mathbf{A}}^{-1} \otimes \underline{\mathbf{B}}^{-1})$$

Finally, we can express the Liapunov matrix equation

$$\underline{\mathbf{AS}} + \underline{\mathbf{A}}^{\mathrm{T}} = \underline{\mathbf{Q}} \tag{18}$$

(all matrices are $n \times n$) in Kronecker product form. $\underline{\mathbf{S}}$ and $\underline{\mathbf{Q}}$ are symmetric. We leave $\underline{\mathbf{A}}$ alone and express $\underline{\mathbf{Q}}$ and $\underline{\mathbf{S}}$ as direct sums of n vectors each, yielding

$$\underline{\mathbf{Q}} = [\underline{\boldsymbol{q}}_1 | \cdots | \underline{\boldsymbol{q}}_n], \quad \underline{\mathbf{S}} = [\underline{\boldsymbol{s}}_1 | \ldots | \underline{\boldsymbol{s}}_n]$$

$$\underline{\mathbf{q}} = \begin{bmatrix} \underline{\boldsymbol{q}}_1 \\ \vdots \\ \underline{\boldsymbol{q}}_n \end{bmatrix}, \qquad \underline{\mathbf{s}} = \begin{bmatrix} \underline{\boldsymbol{s}}_1 \\ \vdots \\ \underline{\boldsymbol{s}}_n \end{bmatrix}$$

The matrix equation (18) takes the form

$$(\underline{\mathbf{I}} \otimes \underline{\mathbf{A}} + \underline{\mathbf{A}} \otimes \underline{\mathbf{I}})\underline{\mathbf{s}} = \underline{\mathbf{q}}$$

8. *Hadamard Product.* If $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$ are $n \times n$, their *Hadamard product* is defined as

$$\underline{\mathbf{H}} = \underline{\mathbf{A}} * \underline{\mathbf{B}}, \qquad \underline{\mathbf{H}} = (h_{ij}) = (a_{ij} b_{ij}) \quad (i, j = 1, \ldots, n)$$

9. *Tridiagonal Form.* If an $n \times n$ matrix $\underline{\mathbf{A}}$ is symmetric, it can be transformed via similarity transformation into a *tridiagonal* form having nonzero entries only on, directly below, or directly above the main diagonal.

10. *Binet–Cauchy Theorem.* A very useful theorem in electrical network theory states the algorithm for computing the determinant of the product $\underline{\mathbf{AB}}$ where $\underline{\mathbf{A}}$ is $m \times n$ and $\underline{\mathbf{B}}$ is $n \times m$, $m < n$. Define the *major* of $\underline{\mathbf{A}}$ (or of $\underline{\mathbf{B}}$) as the determinant of the submatrix of maximum order (in this case m). Then according to *Binet–Cauchy theorem,*

$$\det(\underline{\mathbf{AB}}) = \sum_{\text{all majors}} (\text{products of corresponding majors of } \underline{\mathbf{A}} \text{ and } \underline{\mathbf{B}})$$

11. *Lancaster's Formula.* One has

$$p(\underline{\boldsymbol{x}}) = e^{-f(\underline{\boldsymbol{x}})}, \qquad f(\underline{\boldsymbol{x}}) = \frac{1}{2}\underline{\boldsymbol{x}}^{\mathrm{T}}\underline{\mathbf{R}}^{-1}\underline{\boldsymbol{x}} > 0$$

$$\int_{-\infty}^{\infty} p(\underline{\boldsymbol{x}})\, d\underline{\boldsymbol{x}} = (2\pi)^{-n/2}\Delta_{\underline{\mathbf{R}}}$$

12. *Singular-Value Decomposition.* Suppose $\underline{\mathbf{A}}$ is an $n \times m$ real matrix with $n > m$, with rank $r \le m$. Form

    $\underline{\mathbf{S}} = \underline{\mathbf{AA}}^{\mathrm{T}}$, an $n \times n$ matrix with orthogonal eigenvector $\underline{\boldsymbol{e}}_1, \ldots, \underline{\boldsymbol{e}}_n$

    $\underline{\mathbf{R}} = \underline{\mathbf{A}}^{\mathrm{T}}\underline{\mathbf{A}}$, an $m \times m$ matrix with orthogonal eigenvector $\underline{\boldsymbol{f}}_1, \ldots, \underline{\boldsymbol{f}}_m$

    $\underline{\mathbf{U}} = [\underline{\boldsymbol{e}}_1 | \cdots | \underline{\boldsymbol{e}}_n]$, $\underline{\mathbf{V}} = [\underline{\boldsymbol{f}}_1 | \cdots | \underline{\boldsymbol{f}}_m]$

    $\underline{\boldsymbol{\Sigma}} = Dg[\sigma_1, \sigma_2, \ldots, \sigma_r, \mathbf{0}, \mathbf{0}]$, $\sigma_1 > \sigma_2 > \ldots > \sigma_r$ nonnegative

    $= Dg$[square roots of eigenvalues of $\underline{\mathbf{A}}^{\mathrm{T}}\underline{\mathbf{A}}$]

    Then the *singular–value decomposition* of $\underline{\mathbf{A}}$ is written as $\underline{\mathbf{A}} = \underline{\mathbf{U\Sigma V}}$. The solution to the linear equation $\underline{\mathbf{A}}\underline{x} = \underline{\boldsymbol{b}}$ is

$$\underline{\boldsymbol{x}} = \underline{\boldsymbol{x}}_a + \underline{\boldsymbol{x}}_b, \qquad \underline{\mathbf{x}}_a = \sum_{i=1}^{r} (\underline{\boldsymbol{e}}_i^{\mathrm{T}}\underline{\boldsymbol{b}})\sigma_i^{-1}\underline{\boldsymbol{f}}_i,$$

$$\underline{\boldsymbol{x}}_b = \sum_{i=r+1}^{m} c_i \underline{\boldsymbol{f}}_i \ (c_i \text{ arbitrary})$$

$\underline{\boldsymbol{x}}_b$ represents the auxiliary part of $\underline{\boldsymbol{x}}$, which can be taken as zero.

13. *Schur–Cohen Criteria.* In order that the roots of a polynomial

$$p(\lambda) = a_0\lambda^n + a_1\lambda^{n-1} + \cdots + a_n$$

lie within the unit circle, it is necessary and sufficient that the following conditions be satisfied:

$$(-1)^n p(-1) > 0$$
$$p(1) > 0$$
$$\det(\underline{\mathbf{X}}_i + \underline{\mathbf{Y}}_i) > 0$$
$$\det(\underline{\mathbf{X}}_i - \underline{\mathbf{Y}}_i) > 0$$

$$\underline{\mathbf{X}}_i = \begin{bmatrix} a_0 & a_1 & \ldots & a_{i-1} \\ & a_0 & \ldots & a_{i-2} \\ 0 & & \ddots & \vdots \\ & & & a_0 \end{bmatrix}$$

$$\underline{\mathbf{Y}}_i = \begin{bmatrix} & & & a_n \\ 0 & & a_n & a_{n-1} \\ & \iddots & & \vdots \\ a_n & & \ldots & a_{n-i+1} \end{bmatrix}$$

14. *Hankel, Toeplitz, and Circulant Matrices.* A matrix $\underline{\mathbf{H}}$ is *Hankel* if its (i, j)th entry depends only on the value of $i + j$. Similiarly, a matrix $\underline{\mathbf{T}}$ is *Toeplitz* if its (i, j)th entry depends only on the value of $i - j$. A *circulant matrix* $\underline{\mathbf{C}}$ is defined by $(\underline{\mathbf{C}})_{ij} = c_{j+1-i}$, where subscripts are mod n. Thus 4×4 matrices of those kinds are of the form

$$\underline{\mathbf{H}} = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 \\ h_2 & h_3 & h_4 & h_5 \\ h_3 & h_4 & h_5 & h_6 \\ h_4 & h_5 & h_6 & h_7 \end{bmatrix}$$

$$\underline{\mathbf{T}} = \begin{bmatrix} t_0 & t_{-1} & t_{-2} & t_{-3} \\ t_1 & t_0 & t_{-1} & t_{-2} \\ t_2 & t_1 & t_0 & t_{-1} \\ t_3 & t_2 & t_1 & t_0 \end{bmatrix}$$

$$\underline{\mathbf{C}} = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 \\ c_4 & c_1 & c_2 & c_3 \\ c_3 & c_4 & c_1 & c_2 \\ c_2 & c_3 & c_4 & c_1 \end{bmatrix}$$

Such matrices play an important role in system theory involving state-space realizations.

Nehari's Theorem. Hankel matrices can be used to compute an important bound on a function $f(t)$. Given $f(t)$, compute its

Fourier coefficients c_i ($i = -n, \ldots, -1, 0, 1, \ldots n$) and the associated complex symmetric $(n + 1) \times (n + 1)$ Hankel matrix $\underline{\mathbf{H}}$ such that $(\underline{\mathbf{H}})_{ij} = (c_{i+j})$ ($i = -n, -n + 1, \ldots, 0; j = 0, 1, \ldots, n$). A very useful result due to Nehari states that

$$|\underline{\boldsymbol{x}}^* \underline{\mathbf{H}} \underline{\boldsymbol{x}}| \leq k \underline{\boldsymbol{x}}^* \underline{\boldsymbol{x}}$$

where k is the bound of the function $f(t)$.

BIBLIOGRAPHY

A. C. Aitken, *Determinants and Matrices,* New York: Interscience, 1942. This is excellent reading.

R. Bellman, *Introduction to Matrix Algebra,* New York: McGraw-Hill, 1960. Very comprehensive coverage of current applications, very readable, lots of references.

G. D. Dudley, *Mathematical Foundations for Electromagnetic Theory,* Piscataway, NJ: IEEE Press, 1994, has a very good chapter on linear vector spaces.

F. R. Gantmacher, *The Theory of Matrices,* Vols. 1, 2, translated from Russian by K. A. Hirsch, New York: Chelsea, 1959. This is a very complete work.

R. A. Horn and C. R. Johnson, *Matrix Analysis,* Cambridge, UK: Cambridge Univ. Press, 1993. Must reading for Hermitian and symmetric matries.

T. Kailath, *Linear Systems,* Englewood Cliffs, NJ: Prentice-Hall, 1980. This is a very complete collection of matrix theory for the study of linear dynamical systems.

M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities,* New York: Dover, 1964. A very good survey, very precise and concise.

K. S. Miller, *Some Eclectic Matrix Theory,* Krieger Publishing Co., has a very good treatment of matrix inverses.

A. D. Myskis, *Advanced Mathematics for Engineers: Special Course,* translated from Russian by V. M. Volosov and I. G. Volosova, Moscow: Mir, 1975.

B. Noble, *Applied Linear Algebra,* Englewood Cliffs, NJ: Prentice-Hall, 1969.

L. S. Pontryagin, *Ordinary Differential Equations,* translated from Russian by L. Kacinskas and W. B. Counts, Reading, MA: Addison-Wesley, 1962. Excellent chapter on linear algebra related to matrix differential equations.

V. I. Smirnov, *Linear Algebra and Group Theory,* translated from Russian and revised by R. A. Silverman, New York: McGraw-Hill, 1961. Excellent reading, particularly for the Jordan canonical form.

H. S. Wilf, *Mathematics for the Physical Sciences,* New York: Wiley, 1962.

J. H. Wilkinson, *The Algebraic Eigenvalue Problem,* London: Oxford Univ. Press, 1965. A very valuable book for computation methods.

N. PURI
Rutgers University

LINEAR ALGEBRAIC SOLVERS. See PARALLEL NUMERICAL ALGORITHMS AND SOFTWARE.