Consequently, an irreducible Jordan-form realization of the system $S$ can be found as
\[
\begin{bmatrix}
\dot{x}_1 \\
\dot{x}_2 \\
\dot{x}_3
\end{bmatrix} =
\begin{bmatrix}
-1 & 0 & 0 \\
0 & -1-2i & 0 \\
0 & 0 & -1+2i
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix} +
\begin{bmatrix}
1 \\
1 \\
1
\end{bmatrix} u
\]
\[
y = \begin{bmatrix} 2 & -1 & -1 \end{bmatrix} x
\]

By using (D-5) through (D-8), a discrete-time state equation can be computed as
\[
x(k+1) =
\begin{bmatrix}
e^{-T} & 0 & 0 \\
e^{-T-2iT} & 0 & 0 \\
e^{-T+2iT} & 0 & 0
\end{bmatrix}
x(k) +
\begin{bmatrix}
1 - e^{-T} \\
1 + 2i (1 - e^{-T - 2iT}) \\
1 - 2i (1 - e^{-T + 2iT})
\end{bmatrix} u(k)
\]

We conclude from Theorem D-1 that the discrete-time state equation (D-15) is controllable if and only if
\[
T \neq \frac{2\pi \alpha}{2} = \pi \alpha
\]
and
\[
T \neq \frac{2\pi \alpha}{4} = \frac{\pi \alpha}{2}, \quad \alpha = \pm 1, \pm 2, \ldots
\]

This fact can also be verified directly from (D-15) by using either the criterion $\rho(\tilde{B} A \tilde{B}^T - \tilde{A}^T \tilde{B}^T) = 3$ or Corollary 5-21.

### Problems

**D-1** Consider the continuous-time state equation
\[
\dot{x}(t) =
\begin{bmatrix}
-1 & 0 & 0 \\
0 & -1-2i & 0 \\
0 & 0 & -1+2i
\end{bmatrix}
\begin{bmatrix}
x_1(t) \\
x_2(t) \\
x_3(t)
\end{bmatrix} +
\begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix} u(t)
\]

Show that its discretized state equation is always controllable for any $T$ including $T = \frac{2\pi \alpha}{4}$ for $\alpha = \pm 1, \pm 2, \ldots$. This shows that the conditions on the eigenvalues in Theorem D-1 are not necessary for a multiple-input discretized equation to be controllable.

**D-2** Show that sufficient conditions for the discrete-time dynamical equation in (D-5) to be observable are that the dynamical equation in (D-1) is observable and $\text{Im} \left[ \lambda_j(A) - \lambda_j(\tilde{A}) \right] \neq 2\pi n / T$ for $\alpha = \pm 1, \pm 2, \ldots$. Whenever $\text{Re} \left[ \lambda_j(A) - \lambda_j(\tilde{A}) \right] = 0$. For the single-output case ($\eta = 1$), the conditions are necessary as well.
where \( x \) and \( y \) are any vectors in \((\mathbb{C}^n, \cdot, \cdot)^*\), then the hermitian form can be written as
\[
x^* M x = \langle x, M x \rangle = \langle M^* x, x \rangle = \langle M x, x \rangle
\] (E-4)
where, in the last step, we have used the fact that \( M^* = M \).

**Theorem E-1**

All the eigenvalues of a hermitian matrix \( M \) are real.

**Proof**

Let \( \lambda \) be any eigenvalue of \( M \) and let \( e \) be an eigenvector of \( M \) associated with \( \lambda \); that is, \( Me = \lambda e \). Consider
\[
\langle e, Me \rangle = \langle e, \lambda e \rangle = \lambda \langle e, e \rangle
\] (E-5)
Since \( \langle e, Me \rangle \) is a real number and \( \langle e, e \rangle \) is a positive real number, from (E-5) we conclude that \( \lambda \) is a real number.

Q.E.D.

**Theorem E-2**

The Jordan-form representation of a hermitian matrix \( M \) is a diagonal matrix.

**Proof**

Recall from Section 2.6 that every square matrix which maps \((\mathbb{C}^n, \cdot, \cdot)^*\) into itself has a Jordan-form representation. The basis vectors that give a Jordan-form representation consist of eigenvectors and generalized eigenvectors of the matrix. We show that if a matrix is hermitian, then there is no generalized eigenvector of grade \( k \geq 2 \); we show this by contradiction. Suppose there exists a vector \( e \) such that \((M - \lambda_1 I)^{k-1} e = 0 \) and \((M - \lambda_2 I)^{k-1} e \neq 0 \) for some eigenvalue \( \lambda_i \) of \( M \). Consider now, for \( k \geq 2 \),
\[
0 = \langle (M - \lambda_1 I)^{k-1} e, (M - \lambda_2 I)^{k-1} e \rangle = \langle (M - \lambda_1 I)^{k-1} e \rangle^2
\]
which implies \((M - \lambda_2 I)^{k-1} e = 0 \). This is a contradiction. Hence there is no generalized eigenvector of grade \( k \geq 2 \). Consequently, there is no Jordan block whose order is greater than one. Hence the Jordan-form representation of a hermitian matrix is a diagonal matrix. In other words, there exists a nonsingular matrix \( P \) such that \( M^* = M \) and \( \tilde{M} \) is a diagonal matrix with eigenvalues on the diagonal.

Q.E.D.

Two vectors \( x, y \) are said to be orthogonal if and only if \( \langle x, y \rangle = 0 \). A vector \( x \) is said to be normalized if and only if \( \langle x, x \rangle = \frac{1}{\|x\|^2} = 1 \). It is clear that every vector \( x \) can be normalized by choosing \( \hat{x} = (1/\|x\|) x \). A set of basis vectors \( \{q_1, q_2, \ldots, q_n\} \) is said to be an orthonormal basis if and only if
\[
\langle q_i, q_j \rangle = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}
\] (E-6)

Now we show that the basis of a Jordan-form representation of a hermitian matrix can be chosen as an orthonormal basis. This is derived in part from the following theorem.

**Theorem E-3**

The eigenvectors of a hermitian matrix \( M \) corresponding to different eigenvalues are orthogonal.

**Proof**

Let \( e_i \) and \( e_j \) be the eigenvectors of \( M \) corresponding to the distinct eigenvalues \( \lambda_i \) and \( \lambda_j \), respectively; that is, \( Me_i = \lambda_i e_i \) and \( Me_j = \lambda_j e_j \). Consider
\[
\langle e_i, Me_j \rangle = \langle e_i, \lambda_j e_j \rangle = \lambda_j \langle e_i, e_j \rangle
\] (E-7)
and
\[
\langle e_j, Me_i \rangle = \langle e_j, \lambda_i e_i \rangle = \lambda_i \langle e_j, e_i \rangle
\] (E-8)
where we have used the fact that the eigenvalues are real. Subtracting (E-8) from (E-7), we obtain \((\lambda_i - \lambda_j) \langle e_i, e_j \rangle = 0 \). Since \( \lambda_i \neq \lambda_j \), we conclude that \( \langle e_i, e_j \rangle = 0 \).

Q.E.D.

Since every eigenvector can be normalized and since eigenvectors of a hermitian matrix \( M \) associated with distinct eigenvalues are orthogonal, then the eigenvectors associated with different eigenvalues can be made to be orthonormal. We consider now the linearly independent eigenvectors associated with the same eigenvalue. Let \( \{e_1, e_2, \ldots, e_n\} \) be a set of linearly independent eigenvectors associated with the same eigenvalue. Now we shall obtain a set of orthonormal vectors from the set \( \{e_1, e_2, \ldots, e_n\} \). Let
\[
u_1 = e_1,
q_1 = \frac{\nu_1}{\|\nu_1\|}
\]
\[
u_2 = e_2 - \langle e_2, e_1 \rangle q_1,
q_2 = \frac{\nu_2}{\|\nu_2\|}
\]
\[
u_3 = e_3 - \langle e_3, e_2 \rangle q_1 - \langle e_3, e_1 \rangle q_2,
q_3 = \frac{\nu_3}{\|\nu_3\|}
\]
\[
u_m = e_m - \sum_{k=1}^{m-1} \langle e_m, e_k \rangle q_k,
q_m = \frac{\nu_m}{\|\nu_m\|}
\]
The procedure for defining \( q_i \) is illustrated in Figure E-1. It is called the Schmidt orthonormalization procedure. By direct verification, it can be shown that \( \langle q_i, q_j \rangle = 0 \) for \( i \neq j \).

From Theorem E-3 and the Schmidt orthonormalization procedure we conclude that for any hermitian matrix there exists a set of orthonormal vectors with respect to which the hermitian matrix has a diagonal-form representation; or equivalently, for any hermitian matrix \( M \), there exists a nonsingular matrix \( Q \)
If the rank of $H$ is $r$, so is the rank of $H^*H$ (Problem E-3). Hence we have $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_r^2 > 0$ and $\lambda_{r+1}^2 = \lambda_{r+2}^2 = \cdots = \lambda_n^2 = 0$. Let $q_i, i = 1, 2, \ldots, n$, be the orthonormal eigenvectors of $H^*H$ associated with $\lambda_i^2$. Define

$$Q = [q_1, \ q_2, \ \cdots, \ q_r, \ q_{r+1}, \ \cdots, \ q_n] = [Q_1 : Q_2]$$

Then Theorem E-4 implies

$$Q^*H^*HQ = \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$

(11)

where $\Sigma^2 = \text{diag}(\lambda_1^2, \lambda_2^2, \ldots, \lambda_r^2)$. Using $Q = [Q_1 : Q_2]$, (11) can be written as

$$Q^*H^*HQ_2 = 0$$

(12)

and

$$Q^*H^*HQ_1 = \Sigma^2$$

which implies

$$\Sigma^{-1}Q^*H^*HQ_1 \Sigma^{-1} = I$$

(13)

where $\Sigma = \text{diag} \{\lambda_1, \lambda_2, \ldots, \lambda_r\}$. Define the $m \times r$ matrix $R_1$ by

$$R_1 = HQ_1$$

(14)

Then (13) becomes $R_1^*R_1 = I$ which implies that the columns of $R_1$ are orthonormal. Let $R_2$ be chosen so that $R = [R_1 : R_2]$ is unitary. Consider

$$R^*HQ = \begin{bmatrix} R_1^* \\ R_2^* \end{bmatrix} H[Q_1 : Q_2] = \begin{bmatrix} R_1^*HQ_1 \\ R_2^*HQ_2 \end{bmatrix}$$

(15)

Clearly, (14) implies $HQ_1 = R_1 \Sigma$ and (12) implies $HQ_2 = 0$. Because $R$ is orthonormal, we have $R_1^*R_1 = I$ and $R_2^*R_2 = 0$. Consequently, (15) becomes

$$R^*HQ = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$$

This is stated as a theorem.

Theorem E-5 (Singular value decomposition)

Every $m \times n$ matrix $H$ of rank $r$ can be transformed into the form

$$R^*HQ = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$$

or

$$H = R \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} Q^*$$

where $R^*R = RR^* = I_m$, $Q^*Q = QQ^* = I_n$, and $\Sigma = \text{diag} \{\lambda_1, \lambda_2, \ldots, \lambda_r\}$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$.

Although $\Sigma$ is uniquely determined by $H$, the unitary matrices $R$ and $Q$ are not necessarily unique. Indeed let $\lambda_i^2$ be a multiple eigenvalue of $H^*H$; then the corresponding columns of $Q$ may be chosen as any orthonormal basis for the space spanned by the eigenvectors of $H^*H$ corresponding to $\lambda_i^2$. Hence $Q$
is not unique. Once Q is chosen, \( R_1 \) can be computed from (E-14). The choice of \( R_2 \) again may not be unique so long as \([ R_1 \quad R_2 ]\) is unitary.

The singular value decomposition has found many applications in linear systems. In Section 6-5, we use it to find an irreducible realization from a Hankel matrix. It can also be used to find simplified or approximated models of systems. See References S141, S161, and S171. The singular value decomposition is also essential in the study of sensitivity and stability margin of multivariable systems. See References S34 and S194. For computer programs, see Reference S82.

The elements of matrices in this appendix are permitted to assume real or complex numbers. Certainly all results still apply if they are limited to real numbers. For a real matrix \( M \) we have \( M^* = M' \), where the prime denotes the transpose. A real matrix with \( M = M' \) is called a symmetric matrix; a real matrix with \( M^{-1} = M' \) is called an orthogonal matrix. With these modifications in nomenclature, all theorems in this appendix apply directly to real matrices.

Problems

E-9 Show that all the singular values of a unitary matrix (including orthogonal matrix) are equal to 1.

E-10 What are the eigenvalues and singular values of the elementary matrices

\[
E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & 0 \\ 0.5 & 1 \\ 0.8 & 0 \end{bmatrix}
\]

E_1 may arise in gaussian elimination without any pivoting; \( E_2 \) may arise in gaussian elimination with partial pivoting. If the condition number of a matrix is defined as the ratio of the largest and smallest singular values, which matrix has a larger condition number? Roughly speaking, a condition number gives the amplification factor of the relative errors in the computation.

E-11 Show that the controllability gramian

\[
W_c = \int_0^T e^{tB}BB^*e^{t^*} \, dt
\]

is positive definite if and only if \([A, B] \) is controllable.

E-12 Show that Theorem E-5 reduces to Theorem E-4 if \( H \) is square, hermitian, and positive semidefinite. If \( H \) is square and hermitian (without being positive semidefinite), what are the differences between Theorem E-5 and Theorem E-4?

E-13 What are the singular values of the following matrices?

\[
A_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}
\]
clearly $x$ is a right eigenvector of $A$ and $y$ is a left eigenvector of $B$. Applying the operator $\mathcal{A}$ on the $n \times n$ matrix $xy$ that is clearly a nonzero matrix, we obtain

$$\mathcal{A}(xy) = Axy + xyB = \lambda_i xy + \mu_j xy = (\lambda_i + \mu_j)xy$$

Hence $\lambda_i + \mu_j$ is an eigenvalue of $\mathcal{A}$ for $i = 1, 2, \ldots, l$, and $j = 1, 2, \ldots, m$. See Definition 2-12.

Next we prove that all eigenvalues of $\mathcal{A}$ are of the form $\lambda_i + \mu_j$. Let us assume that $\eta_k$ is an eigenvalue of $\mathcal{A}$. Then, by definition, there exists a $M \neq 0$ such that

$$\mathcal{A}(M) = AM + MB = \eta_k M$$

or

$$(\eta_k I - A)M = MB \quad (F-1)$$

We now show that the matrices $\eta_k I - A$ and $B$ have at least one eigenvalue in common. We prove this by contradiction. Let

$$\Delta(s) = s^p + a_{p-1}s^{p-1} + \cdots + a_1s + a_0$$

be the characteristic polynomial of $\eta_k I - A$, that is $\Delta(s) = \det (sI - \eta_k I + A)$. Then we have

$$\Delta(\eta_k I - A) = 0 \quad (F-2)$$

If $\eta_k I - A$ and $B$ have no common eigenvalue, then the matrix

$$\Delta(B)$$

is nonsingular. Indeed, if $\mu_i, i = 1, 2, \ldots, n$, are eigenvalues of $B$, then $\Delta(\mu_i)$ are eigenvalues of $\Delta(B)$ (Problem 2-32) and $\det \Delta(B) = \prod \Delta(\mu_i)$ (Problem 2-22).

If $\eta_k I - A$ and $B$ have no common eigenvalue, then $\Delta(\mu_i) \neq 0$ for all $i$, and

$$\Delta(B) \neq 0.$$ Hence $\Delta(B)$ is nonsingular.

From (F-1), we can develop the following equalities

$$(\eta_k I - A)^2M = (\eta_k I - A)MB = MB^2$$

$$(\eta_k I - A)^pM = MB^p$$

The summation of the products of $(\eta_k I - A)M = MB$ and $a_i$, for $i = 0, 1, 2, \ldots, n$, with $a_0 = 1$, yields

$$\Delta(\eta_k I - A)M = M\Delta(B)$$

which, together with (F-2), implies

$$0 = M\Delta(B) \quad (F-3)$$

Since $\Delta(B)$ is nonsingular, (F-3) implies $M = 0$. This contradicts the assumption that $M \neq 0$. Hence the matrices $\eta_k I - A$ and $B$ have at least one common eigenvalue. Now the eigenvalue of $\eta_k I - A$ is of the form $\eta_k - \lambda_i$. Consequently, for some $i$ and for some $j$,

$$\eta_k = \lambda_i + \mu_j$$

Q.E.D.
transition matrix of the equation is

$$
\Phi(t, 0) = \begin{bmatrix}
  e^{-t} & (e^{-t} - e^{-t})/2 \\
  0 & e^{-t}
\end{bmatrix}
$$

(as in Problem 4-1 or by direct verification), whose norm tends to infinity as $t \to \infty$.

8-5 Lyapunov Theorem

The asymptotic stability of $\dot{x} = Ax$ can be determined by first computing the characteristic polynomial of $A$ and then applying the Routh-Hurwitz criterion. If all the roots of the characteristic polynomial have negative real parts, then the zero state of $\dot{x} = Ax$ is asymptotically stable. There is one more method of checking the asymptotic stability of $\dot{x} = Ax$ without computing explicitly the eigenvalues of $A$. We shall discuss such a method in this section and then apply it to establish the Routh-Hurwitz criterion.

Before proceeding, we need the concept of positive definite and positive semidefinite matrices. An $n \times n$ matrix $M$ with elements in the field of complex numbers is said to be a hermitian matrix if $M^* = M$, where $M^*$ is the complex conjugate transpose of $M$. If $M$ is a real matrix, $M$ is said to be symmetric. The matrix $M$ can be considered as an operator that maps $(\mathbb{C}^n, \mathbb{C})$ into itself.

It is shown in Theorem E-1 that all the eigenvalues of a hermitian matrix are real, and that there exists a nonsingular matrix $P$, called a unitary matrix, such that $P^{-1} = P^*$ and $M = PMP^*$, where $M$ is a diagonal matrix with eigenvalues on the diagonal (Theorem E-4). We shall use this fact to establish the following theorem.

Theorem 8-18

Let $M$ be a hermitian matrix and let $\lambda_{min}$ and $\lambda_{max}$ be the smallest and largest eigenvalues of $M$, respectively. Then

$$
\lambda_{min} ||x||^2 \leq x^*Mx \leq \lambda_{max} ||x||^2
$$

(8-35)

for any $x$ in the $n$-dimensional complex vector space $\mathbb{C}^n$, where

$$
||x||^2 = (x, x) \triangleq x^*x = \sum_{i=1}^{n} |x_i|^2
$$

and $x_i$ is the $i$th component of $x$.

Proof

Note that $x^*Mx$ is a real number for any $x$ in $\mathbb{C}^n$. Let $P$ be the nonsingular matrix such that $P^{-1} = P^*$ and $M = PMP^*$, where $M$ is a diagonal matrix with eigenvalues of $M$ on the diagonal. Let $\check{x} = Px$ or $x = P^{-1} \check{x} = P^* \check{x}$, then

$$
x^*Mx = \check{x}^*PMP^*\check{x} = \check{x}^*M\check{x} = \sum_{i=1}^{n} \lambda_i |\check{x}_i|^2
$$

where the $\lambda_i$'s are the eigenvalues of $M$. It follows that

$$
\lambda_{min} \sum_{i=1}^{n} |\check{x}_i|^2 \leq x^*Mx = \check{x}^*M\check{x} = \sum_{i=1}^{n} \lambda_i |\check{x}_i|^2 \leq \lambda_{max} \sum_{i=1}^{n} |\check{x}_i|^2
$$

(8-36)

The fact that $P^{-1} = P^*$ implies that

$$
||x||^2 = x^*x = \check{x}^*x = \sum_{i=1}^{n} |x_i|^2
$$

Hence, the inequality (8-36) implies (8-35).

Q.E.D.

Definition 8-6

A hermitian matrix $M$ is said to be positive definite if and only if $x^*Mx > 0$ for all nonzero $x$ in $\mathbb{C}^n$. A hermitian matrix $M$ is said to be positive semidefinite or nonnegative definite if and only if $x^*Mx \geq 0$ for all $x$ in $\mathbb{C}^n$, and the equality holds for some nonzero $x$ in $\mathbb{C}^n$.

Theorem 8-19

A hermitian matrix $M$ is positive definite (positive semidefinite) if and only if any one of the following conditions holds:

1. All the eigenvalues of $M$ are positive (nonnegative).
2. All the leading principal minors of $M$ are positive (all the principal minors of $M$ are nonnegative).
3. There exists a nonsingular matrix $N$ (a singular matrix $N$) such that $M = N^*N$.

* The principal minors of the matrix

$$
M = \begin{bmatrix}
  m_{11} & m_{12} & m_{13} \\
  m_{21} & m_{22} & m_{23} \\
  m_{31} & m_{32} & m_{33}
\end{bmatrix}
$$

are $m_{11}$, $m_{22}$, $m_{33}$, $\det \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$, $\det \begin{bmatrix} m_{11} & m_{13} \\ m_{31} & m_{33} \end{bmatrix}$, $\det \begin{bmatrix} m_{22} & m_{23} \\ m_{32} & m_{33} \end{bmatrix}$, and $\det M$, that is, the minors whose diagonal elements are also diagonal elements of the matrix. The leading principal minors of $M$ are $m_{11}$, $\det \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$, and $\det M$, that is, the minors obtained by deleting the last $k$ columns and the last $k$ rows, for $k = 2, 1$, and 0.

9 It is shown in Reference 39 that if all the leading principal minors of a matrix are positive, then all the principal minors are positive. However, it is not true that if all the leading principal minors are nonnegative, then all the principal minors are nonnegative. For a counterexample, try Problem 8-32b.

* If $M$ is a lower triangular matrix, it is called the Cholesky decomposition. Subroutines are available in LINPACK and IBM Scientific Subroutine Package to carry out this decomposition.

Suppose $P$ is a matrix with $\text{Chol}(P) = 1$.
Proof

Condition 1 follows directly from Theorem 8-18. A proof of condition 2 can be found, for example, in References 5 and 39. For a proof of condition 3, see Problem 8-24. Q.E.D.

With these preliminaries, we are ready to introduce the Lyapunov theorem and its extension. They will be used to prove the Routh-Hurwitz criterion.

Theorem 8-20 (Lyapunov theorem)

All the eigenvalues of A have negative real parts or, equivalently, the zero state of $\dot{x} = Ax$ is asymptotically stable if and only if for any given positive definite hermitian matrix N, the matrix equation

$$A^*M + MA = -N$$

has a unique hermitian solution $M$ and $M$ is positive definite.

Corollary 8-20

All the eigenvalues of A have negative real parts, or equivalently, the zero state of $\dot{x} = Ax$ is asymptotically stable, if and only if for any given positive semi-definite hermitian matrix $N$ with the property $\{A, N\}$ observable, the matrix equation

$$A^*M + MA = -N$$

has a unique hermitian solution $M$ and $M$ is positive definite.

The implication of Theorem 8-20 and Corollary 8-20 is that if $A$ is asymptotically stable and if $N$ is positive definite or positive semidefinite, then the solution M of (8-37) must be positive definite. However, it does not say that if $A$ is asymptotically stable and if $M$ is positive definite, then the matrix $N$ computed from (8-37) is positive definite or positive semidefinite.

Before proving the Lyapunov theorem, we make a few comments. Since Theorem 8-20 holds for any positive definite hermitian matrix $N$, the matrix $N$ in (8-37) is often chosen to be a unit matrix. Since $M$ is a hermitian matrix, there are $n^2$ unknown numbers in $M$ to be solved. If $M$ is a real symmetric matrix, there are $(n^2 + 1)/2$ unknown numbers in $M$ to be solved. Hence the matrix equation (8-37) actually consists of $n^2$ linear algebraic equations. To apply Theorem 8-20, we first solve these $n^2$ equations for $M$, and then check whether or not $M$ is positive definite. This is not an easy task. Hence Theorem 8-20 and its corollary are generally not used in determining the stability of $\dot{x} = Ax$. However, they are very important in the stability study of nonlinear time-varying systems by using the so-called second method of Lyapunov. Furthermore, we shall use it to prove the Routh-Hurwitz criterion.

We give now a physical interpretation of the Lyapunov theorem. If the hermitian matrix $M$ is positive definite, the plot of $V(x)$

$$V(x) \triangleq x^*Mx$$

will be bowl shaped, as shown in Figure 8-6. Consider now the successive values taken by $V$ along a trajectory of $\dot{x} = Ax$. We like to know whether the value of $V$ will increase or decrease with time as the state moving along the trajectory. Taking the derivative of $V$ with respect to $t$ along any trajectory of $\dot{x} = Ax$, we obtain

$$\frac{d}{dt} V(x(t)) = \frac{d}{dt} (x^*(t)Mx(t)) = \left(\frac{d}{dt} x^*(t)\right) Mx(t) + x^*(t)M \left(\frac{d}{dt} x(t)\right)$$

$$= x^*(t)A^*Mx(t) + x^*(t)Mx(t) = x^*(t)(A^*M + MA)x(t)$$

$$= -x^*(t)Nx(t)$$

(8-39)

where $N \triangleq -(A^*M + MA)$. This equation gives the rate of change of $V(x)$ along any trajectory of $\dot{x} = Ax$. Now if $N$ is positive definite, the function $-x^*(t)Nx(t)$ is always negative. This implies that $V(x(t))$ decreases monotonically with time along any trajectory of $\dot{x} = Ax$; hence $V(x(t))$ will eventually approach zero as $t \to \infty$. Now since $V(x)$ is positive definite, we have $V(x) = 0$ only at $x = 0$; hence we conclude that if we can find positive definite matrices $M$ and $N$ that are related by (8-37), then every trajectory of $\dot{x} = Ax$ will approach the zero state as $t \to \infty$. The function $V(x)$ is called a Lyapunov function of $\dot{x} = Ax$. A Lyapunov function can be considered as a generalization of the concept of distance or energy. If the "distance" of the state along any trajectory of $\dot{x} = Ax$ decreases with time, then $x(t)$ must tend to $0$ as $t \to \infty$.

Proof of Theorem 8-20

Sufficiency: Consider $V(x) = x^*Mx$. Then we have

$$\dot{V}(x) \triangleq \frac{d}{dt} V(x) = -x^*Nx$$

along any trajectory of $\dot{x} = Ax$. From Theorem 8-18, we have

$$\frac{\dot{V}}{V} = \frac{x^*Nx}{x^*Mx} \leq \frac{\|x\|^2}{\lambda_{max}(M)}$$

(8-40)

Figure 8-6  A Lyapunov function $V(x)$. 

0
where \( \lambda_{\text{min}} \) is the smallest eigenvalue of \( N \) and \( \lambda_{\text{max}} \) is the largest eigenvalue of \( M \). From Theorem 8-19 and from the assumption that the matrices \( M \) and \( N \) are positive definite, we have \( \lambda_{\text{min}} > 0 \) and \( \lambda_{\text{max}} > 0 \). If we define

\[
\alpha \triangleq \frac{\lambda_{\text{min}}}{\lambda_{\text{max}}}
\]

then inequality (8-40) becomes \( \dot{V} \leq -\alpha V \), which implies that \( V(t) \leq e^{-\alpha t} V(0) \).

It is clear that \( \alpha > 0 \); hence \( V \) decreases exponentially to zero on every trajectory of \( \dot{x} = Ax \). Now \( V(x) = 0 \) only at \( x = 0 \); hence we conclude that the response of \( \dot{x} = Ax \) due to any initial state \( x_0 \) tends to zero as \( t \to \infty \). This proves that the zero state of \( \dot{x} = Ax \) is asymptotically stable. \textit{Necessity:} If the zero state of \( \dot{x} = Ax \) is asymptotically stable, then all the eigenvalues of \( \lambda \) have negative real parts. Consequently, for any \( N \), there exists a unique matrix \( M \) satisfying

\[
AM + MA = -N
\]

and \( M \) can be expressed as

\[
M = \int_0^\infty e^{At}Ne^{At}dt \quad (8-41)
\]

(see Appendix F). Now we show that if \( N \) is positive definite, so is \( M \). Let \( H \) be a nonsingular matrix such that \( N = H^*NH \) (Theorem 8-19). Consider

\[
x(t)Mx(t) = \int_0^\infty x(t)e^{At}H^*Ne^{At}x(t)dt = \int_0^\infty ||He^{At}x_0||^2 dt \quad (8-42)
\]

Since \( H \) is nonsingular and \( e^{At} \) is nonsingular for all \( t \), we have \( He^{At}x_0 \neq 0 \) for all \( t \) unless \( x_0 = 0 \). Hence we conclude that \( x(t)Mx(t) > 0 \) for all \( x_0 \neq 0 \), and \( M \) is positive definite. This completes the proof of this theorem. Q.E.D.

In order to establish Corollary 8-20, we show that if \( N \) is positive semidefinite and if \( \{A, N\} \) is observable, then \( x(t)N(t)x(t) \) cannot be identically zero along any nontrivial trajectory of \( \dot{x} = Ax \) (any solution due to any nonzero initial state \( x_0 \)). First we use Theorem 8-19 to write \( N = H^*NH \). Then the observability of \( \{A, N\} \) implies the observability of \( \{A, H\} \) (Problem 5-56).

Consider

\[
x^*(t)N(t)x(t) = x^*(t)e^{At}H^*He^{At}x(t)dt
\]

Since \( \{A, H\} \) is observable, all rows of \( He^{At} \) are linearly independent on \([0, \infty)\). Hence we have that \( He^{At}x_0 = 0 \) for all \( t \) if and only if \( x_0 = 0 \). Because \( e^{At} \) is analytic over \([0, \infty)\), we conclude that for any \( x_0 \neq 0 \), \( He^{At}x_0 \) can never be identically zero over any finite interval, no matter how small; otherwise it would be identically zero over \([0, \infty)\). See Theorem B-1. Note that \( He^{At}x_0 = 0 \), at some discrete instants of time, is permitted.

With the preceding discussion, we are ready to establish Corollary 8-20. Consider the Lyapunov function \( V(x) \) defined in (8-38) and \( dV(x)/dt = -x^*(t)N(t)x(t) \) in (8-39). If \( x_0 \neq 0 \) and \( dV(x)/dt \leq 0 \) and the equality holds only at some discrete instants of time; hence \( V(x(t)) \) will decrease with time, not necessarily monotonically at every instant of time, and will eventually approach zero as \( t \to \infty \). This shows the sufficiency of the corollary. The necessary part can be similarly proved as in Theorem 8-20 by using (8-42).

Theorem 8-20 and its corollary are generally not used in checking the asymptotic stability of \( \dot{x} = Ax \). However, they are important by their own right and are basic in the stability study of nonlinear systems. They also provide a simple proof of the Routh-Hurwitz criterion.

**A proof of the Routh-Hurwitz criterion.** Consider the polynomial

\[
D(s) = a_0s^n + a_1s^{n-1} + a_2s^{n-2} + \cdots + a_n \quad a_0 > 0
\]

with real coefficients \( a_i \), \( i = 0, 1, 2, \ldots, n \). We form the polynomials

\[
D_1(s) = a_0s^p + a_1s^{p-1} + \cdots + a_p \quad a_p > 0
\]

and compute

\[
\frac{D(s)}{D_1(s)} = \frac{a_{n-p} + \frac{a_{n-p+1}}{a_{n-p}} + \frac{a_{n-p+2}}{a_{n-p}} + \cdots + 1}{a_{n-p} + \frac{a_{n-p+1}}{a_{n-p}} + \frac{a_{n-p+2}}{a_{n-p}} + \cdots + 1 + \frac{a_{n-p+1}}{a_{n-p}} + \frac{a_{n-p+2}}{a_{n-p}} + \cdots + 1 + \frac{a_{n-p+2}}{a_{n-p}} + \cdots + \frac{a_n}{a_{n-p}}}
\]

For convenience, we shall restate the theorem here.

**Theorem 8-6**

The polynomial \( D(s) \) is a Hurwitz polynomial if and only if all the \( n \) numbers \( a_1, a_2, \ldots, a_n \) are positive.

**Proof**

First we assume that all the \( a_i \)'s are different from zero. Consider the rational function

\[
g(s) \triangleq \frac{D(s)}{D_1(s)} = \frac{D_2(s)}{D_1(s)} + \frac{1}{D_1(s)}
\]

The assumption \( a_{i,0} \neq 0 \), for \( i = 1, 2, \ldots, n \), implies that there is no common factor between \( D_2(s) \) and \( D_1(s) \). Consequently, there is no common factor between \( D_1(s) \) and \( D_2(s) \); in other words, \( g(s) \) is irreducible.

Consider the block diagram shown in Figure 8-7. We show that the transfer function from \( u \) to \( y \) is \( g(s) \). Let \( \hat{H}_1(s) \) be the transfer function from \( x_0 \) to \( x_0 \), or equivalently, from the terminal \( E \) to the terminal \( F \), as shown in Figure 8-7.

\[1\text{This follows Reference 90.}\]
SOME FACTS FROM MATRIX THEORY

Appendix

We have collected together, in the form of groups of exercises, all the matrix facts that will be needed for the first part of this book. It is not necessary to actually work out any of the exercises, at least on a first reading. But you should try to do so after a result in the text provides some motivation for its significance and value. If the proof seems difficult, do not hesitate to look up the relevant topic in a book or a cited paper. Use your good sense on how far to pursue something—the goal is not the mathematics per se but to learn how to find and use mathematics for your needs. By using the Appendix this way, by the end of this course you will have learned a surprisingly large amount of matrix theory, and you will be better prepared and motivated to profit from regular courses in these subjects.

There are of course several good books on matrix theory and linear algebra, and we give an annotated list of some of our favorites (11-19) in the References. Note, however, that none of the books has everything we shall need, though [1] and [2] come very close.

1 BASIC OPERATIONS

A.1 Transformations and the Usual Rule for Matrix Multiplication

We have the relations

\[ y_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \]
\[ y_2 = a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \]
\[ y_3 = a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \]

or

\[ y_i = \sum_{j=1}^{3} a_{ij}x_j, \quad i = 1, 2, 3 \]
Appendix  Some Facts from Matrix Theory

and

\[
\begin{align*}
    z_1 &= b_{11}y_1 + b_{12}y_2 + b_{13}y_3, \quad i = 1, 2, \\
    z_2 &= b_{21}y_1 + b_{22}y_2 + b_{23}y_3.
\end{align*}
\]

or \( z_i = \sum_{j=1}^{3} b_{ij} y_j \), \( i = 1, 2 \).

Let \( x, y, z \) be column matrices with components \( \{x_i\}, \{y_i\}, \{z_i\} \), respectively. Represent in matrix notation the relations between (1) \( y \) and \( x \), (2) \( z \) and \( y \), and (3) \( z \) and \( x \). This example explains the origin of the usual rule for matrix multiplication. We say usual or special applications other rules can be given, e.g., the so-called Schur product, defined by

\[
[A \circ B]_{ij} = a_{ij}b_{ij}
\]

or the Lie product, defined by

\[
[A, B]_{ij} = \sum_{k=1}^{n} [a_{ki}b_{kj} - b_{ki}a_{kj}]
\]

or the Kronecker product, defined by

\[
A \otimes B = \begin{bmatrix}
    a_{11}B & a_{12}B & \cdots & a_{1n}B \\
    \vdots & \vdots & & \vdots \\
    a_{m1}B & a_{m2}B & & a_{mn}B
\end{bmatrix}
\]

A.2 Vectors and Column Matrices

Let \( \bar{x} \) be a vector in 2-space, with \( Ox \) and \( Oy \) the (standard) coordinate axes for the space. Let the components of \( \bar{x} \) along these axes be \( x_1 \) and \( x_2 \). The column matrix \( x \) with elements \( x_1 \) and \( x_2 \) will be a representation of \( \bar{x} \). Similarly, let \( y \) be a representation of \( \bar{y} \).

1. Show that

\[
\langle \bar{x}, \bar{y} \rangle \triangleq \text{scalar product of } \bar{x} \text{ and } \bar{y}
\]

\[
= x'y' = y'x' \triangleq \langle x, y \rangle
\]

\[
||\bar{x}|| \triangleq \text{length of } \bar{x} = \sqrt{\bar{x}'\bar{x}} = ||x||
\]

\[
\cos \theta \triangleq \text{cosine of angle between } \bar{x} \text{ and } \bar{y} = x'y' ||x|| ||y||
\]

If \( x'y' = 0 \), we say the vectors \( \bar{x} \) and \( \bar{y} \), and the (column) matrices \( x \) and \( y \), are orthogonal. (Note well: If \( \{x, y\} \) have complex entries, we must use the Hermitian transpose above; i.e., we must also take complex conjugates when we transpose; otherwise we shall not have \( ||x|| \) nonnegative. For notational convenience, we shall assume throughout

this book that we are dealing with real-valued vectors, leaving the reader to make the appropriate adjustments as above when complex vectors are encountered, e.g., as eigenvectors.)

2. A vector \( \bar{y} \) is obtained from \( \bar{x} \) by rotating \( \bar{x} \) through an angle \( \alpha \). Find the relation between \( \bar{x} \), \( \bar{y} \), and \( \bar{x} \).

A.3 The Cauchy-Schwarz Inequality

Let \( x \) and \( y \) be two \( N \)-component column matrices. Show that

\[
|\langle x, y \rangle|^2 = \langle x, x \rangle \langle y, y \rangle \leq ||x||^2 \cdot ||y||^2
\]

Under what conditions will equality hold?

A.4 Gram-Schmidt Orthogonalization and Triangularization

1. Let \( \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n\} \) be a set of vectors. Show how to find a set of orthonormal vectors \( \{\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_m, M \leq N\} \), spanning the same space as the \( \{\bar{x}_i\} \). (The space spanned by the \( \{\bar{x}_i\} \) is the space of all linear combinations of the \( \{\bar{x}_i\} \).

   Hint: Let \( \bar{y}_1 = \bar{x}_1/||\bar{x}_1||, \bar{y}_2 = \bar{x}_2 - a_2 \bar{y}_1 / ||\bar{x}_2 - a_2 \bar{y}_1||, \) where \( a \) is chosen so that \( \langle \bar{y}_2, \bar{y}_2 \rangle = 0 \). Consider the geometric interpretation of this procedure, which is known as the Gram-Schmidt technique.

2. Using the above, show that if \( A = m \times n \), there exists an \( n \times n \) upper triangular matrix \( R \) (i.e., one with zero entries below the main diagonal) and an \( m \times n \) matrix \( Q \) with orthonormal columns (i.e., columns that are orthogonal and have unit length) such that \( A = QR \). Hint: Apply the Gram-Schmidt technique to the columns of \( A \) and write the result in matrix notation. This is an important exercise. Work out a numerical example for yourself, taking, say, four different \( 3 \times 1 \) vectors.

Alternative procedures for orthogonalization and triangularization—modified Gram-Schmidt, Householder, and Givens—are described, for example, in [12]-[19] and, more briefly, in [2, pp. 276–282]; they generally have better numerical properties than the Gram-Schmidt method (however also see the remarks in [17, p. 9.26]).

A.5 Complex Numbers and 3-Vectors as Matrices

1. Verify that complex numbers can be represented by matrices via the correspondence

\[
a + ib \sim \begin{bmatrix} a & -b^* \\ b & a \end{bmatrix}
\]
with the usual rules for matrix manipulation. This representation is 
useful in doing complex arithmetic on a digital computer.

2. Consider the association of 3-vectors with matrices, say,

\[
\mathbf{x} = x_1 \mathbf{i} + x_2 \mathbf{j} + x_3 \mathbf{k} \sim \begin{bmatrix} 0 & x_3 & x_2 \\ -x_3 & 0 & x_1 \\ -x_2 & -x_1 & 0 \end{bmatrix} = \mathbf{X}
\]

where \((\mathbf{i}, \mathbf{j}, \mathbf{k})\) are the orthogonal unit vectors in 3-space. Show that the Lie product rule (cf. Exercise A.1), \([X, Y] = XY - YX\), gives us the matrix associated with the “vector product” of \(\mathbf{x}\) and \(\mathbf{y}\). This representation is useful for calculations in electromagnetic theory, inertial navigation, etc.

### A.6 Toeplitz Matrices

A matrix is Toeplitz if its \((i, j)\)th entry depends only on the value of \(i - j\). Thus such a matrix is “constant along the diagonals.”

Note that a lower (upper) triangular Toeplitz matrix is completely specified by the elements of the first column (row). Among many other things, Toeplitz matrices provide a convenient representation for polynomial multiplication.

1. Thus, show that if \(a(y)b(y) = c(y)\), with \(a(y)\) of degree \(n\) and \(b(y)\) of degree \(m\), then \(T(a)b = c\), where (in an obvious notation) \(b' = \begin{bmatrix} b_0 & \cdots & b_m \end{bmatrix}\), \(c' = \begin{bmatrix} c_0 & \cdots & c_{m+n} \end{bmatrix}\) and \(T(a)\) is an \((m+n+1) \times (m+1)\) lower triangular Toeplitz matrix with first column \([a_0 \cdots a_n 0 \cdots 0]\).

2. Show that we can also write \(c = T(b)a\), where \(T(b)\) is an \((m+n+1) \times (n+1)\) matrix with first column \([b_0 \cdots b_n 0 \cdots 0]\).

3. What can you say about the commutativity of Toeplitz matrices from the fact that polynomials commute?


### 2 SOME DETERMINANT FORMULAS

We assume that the reader is familiar with the elementary properties of determinants. We recall only that the determinant can be evaluated by using Laplace’s expansion,

\[
det A = \sum_i a_{ii} \gamma_i
\]

for any \(i = 1, 2, \ldots, n\)

where \(\gamma_i\) denotes the cofactor corresponding to \(a_{ii}\) and is equal to \((-1)^{i+1}\)

\([\mathbf{A} \setminus \mathbf{a}_i]_{ij} = \mathbf{A}_{ij} \text{ is the } (n-1) \times (n-1) \text{ matrix obtained by deleting the } i\text{th row and } j\text{th column of } \mathbf{A}. \det \mathbf{A}_{ij} \text{ is called the } ij\text{th minor of the matrix; the leading or principal minors are obtained when } i = j.\]

An important consequence of Laplace’s expansion is that the determinant of a triangular matrix is equal to the product of the diagonal elements.

To assist in the computation of \(\det A\), we often use the above fact and some combination of the following results. The operations described in 1–3 below are called elementary (row or column) operations.

1. If any column (or row) of \(A\) is multiplied by a scalar \(c\) and the resulting matrix is denoted by \(\mathbf{A}\), then \(\det \mathbf{A} = c \det A\). Hence it is easy to deduce that \(\det [cA] = c^n \det A\).

2. If \(\mathbf{A}\) is the matrix obtained from \(A\) by interchanging any two rows (or columns) of \(A\), then \(\det \mathbf{A} = -\det A\).

3. If \(\mathbf{A}\) is obtained from \(A\) by adding a multiple of any one row (or column) to another, then \(\det \mathbf{A} = \det A\).

4. \(\det A' = \det A\) (transposition does not affect the determinant).

5. \(\det (A + B) = \det A \det B\) for any two square matrices, then \(\det AB = \det A \det B\).

### A.7 Vandermonde Determinants

Given \([\lambda_1, \ldots, \lambda_n]\), the Vandermonde matrix is one that has \(i\)th row \([1 \ \lambda_i \ \lambda_i^2 \ \cdots \ \lambda_i^{n-1}]\). By direct evaluation (or by row operations), show that the \(3 \times 3\) Vandermonde determinant is given by \((\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)(\lambda_3 - \lambda_1)\).

Then show that the \(n \times n\) Vandermonde determinant is equal to \(\prod_{i=1}^{n} (\lambda_i - \lambda_j)\), where the product is taken over all \((i, j)\) such that \(1 \leq i < j \leq n\).

### A.8 Cramer's Rule

There is a famous formula, useful chiefly for theoretical analyses, that gives the inverse of a matrix as \(A^{-1} = \text{Adj } A/\det A\), where \(\text{Adj } A\) is the adjugate matrix of \(A\), defined by (note the transpose)

\[
\text{Adj } A = [a_{ji}']
\]

and \(a_{ji}'\) is, as noted earlier, the cofactor of \(a_{ij}\).

Now suppose we have to solve \(Ax = y\). Show that \(x = [x_1 \ \cdots \ x_n]'\) can be computed as \(x = \det A/\det A\) with the \(i\)th column replaced by \(y\).

### A.9 The Binet-Cauchy Formula [5]

If \(A\) is an \(m \times n\) matrix, let \([A]_{rk} = \mathbf{r} \times \mathbf{r}\) minor formed by selecting \(r\) rows, say \([i_1, \ldots, i_r]\), and \(r\) columns, say \([k_1, \ldots, k_r]\), of \(A\), where \(r\) is any integer such that \(r \leq \min(m, n)\). Then if \(B\) is any \(n \times p\) matrix, the Binet-
Cauchy formula says that

$$|AB|_k = \sum |A|_k |B|_k$$

where $l$ ranges over all sets of $r$ integers, and $1 \leq r \leq \min(m, n, p)$.

1. Show that if $A$ and $B$ are both square, then $\det AB = \det A \det B = \det BA$.
2. If $A$ is $m \times n$ and $B$ is $n \times m$, $m < n$, find an expression for $\det AB$.

3 BLOCK MATRICES AND THEIR DETERMINANTS

We shall often encounter matrices whose elements are themselves matrices. No additional special rules are necessary for such matrices; we just have to be sure that we are not violating any rules for ordinary matrices.

A.10 Multiplication of Block Matrices

Specify the dimensions that will make the following formula valid when $A$ is $n \times m$ and $H$ is $p \times q$:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} E & F \\ G & H \end{bmatrix} = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix}$$

A.11 Determinants of Block Matrices

1. Show that when $A$ and $B$ are square

$$\det \begin{bmatrix} A & 0 \\ C & B \end{bmatrix} = \det A \det B$$

*Hint:* If $A$ or $B$ is singular, the identity is trivial. Otherwise, observe that

$$\begin{bmatrix} A \quad 0 \\ C \quad B \end{bmatrix} = \begin{bmatrix} A \quad 0 \\ 0 \quad I \end{bmatrix} \begin{bmatrix} I \\ 0 \quad B \end{bmatrix} B^{-1} C I$$

2. Hence, if $A$ is nonsingular, show that

$$\det \begin{bmatrix} A & D \\ C & B \end{bmatrix} = \det A \det [B - CA^{-1} D]$$

What if we only know that $B$ is nonsingular?

A.12 A Useful Identity for Determinants of Products

We have stated that $\det AB = \det BA$ if $A$ and $B$ are square but that this does not hold for nonsquare $A$ and $B$. However, show that if $A$ is $n \times m$ and $B$ is $m \times n$, then $\det [I_m - AB] = \det [I_n - BA]$, where $I_p$ is the $p \times p$ identity matrix. The special case when $m = 1$, so that $A$ is a column and $B$ is a row, is especially useful. *Hint:* Apply elementary row and column operations and Exercise A.11 to the block matrix with first row $[I_n, A]$ and second row $[B, I_m]$.

A.13 Useful Representations for Transfer Functions

1. Show that if $A$ is $n \times n$, $b$ is $n \times 1$, $c$ is $1 \times n$, and $d$ is a scalar, then

$$\det (sI - A)[c(sI - A)^{-1} b + d] = \det \begin{bmatrix} sI - A & b \\ -c & d \end{bmatrix}$$

The two sides are sometimes said to be reciprocal forms (see [10, p. 162]).

2. If $G(s) = c(sI - A)^{-1} b$, show that we can write

$$G(s) = \frac{\det (sI - A + bc)}{\det (sI - A)}$$

A.14 Some Remarks on Linear Equations

Many questions in system theory reduce ultimately to the study of the linear equation

$$A \cdot x = y$$

$m \times n \quad n \times 1 \quad m \times 1$
for an unknown vector \( x \). Depending on the values of the matrices \( A \) and \( y \), this equation may have a unique solution, or many solutions, or no solution.

If we recall that \( Ax \) can be regarded as a linear combination of the columns of \( A \), with weights equal to the corresponding components of \( x \), then it is clear that the equations will have a solution if and only if the right-hand side \( y \) is some linear combination of the columns of \( A \). In this case, the equations are said to be consistent, and \( y \) is said to lie in the range space of \( A \), denoted \( \mathcal{R}(A) \). The word space is used because of the following closure property: if \( y_1 \) and \( y_2 \) are in the range of \( A \) (i.e., if they are linear combinations of the columns of \( A \)), then any linear combination of them, \( c_1 y_1 + c_2 y_2 \), where \( c_1 \) and \( c_2 \) are scalars, will also be in the range of \( A \).

If the equations are consistent, when can we have more than one solution? Note that if \( x_1 \) and \( x_2 \) are any two solutions (i.e., \( Ax_1 = y \) and \( Ax_2 = y \)), then

\[
A(x_1 - x_2) = 0
\]

It is then clear that \( x_1 \) will be different from \( x_2 \), and we shall have more than one solution if and only if \( A(x_1 - x_2) = 0 \) has a nontrivial solution, i.e., a solution not all of whose components are zero. The null-space of \( A \), denoted \( \mathcal{N}(A) \), is the space of all solutions of the equation

\[
Ax = 0
\]

and this space will be empty if and only if no nontrivial combination of the columns of \( A \) can be zero.† Any set of columns with this property is said to be a linearly independent set, and any matrix whose columns are linearly independent will be said to have full column rank. The null-space of \( A \) will be nonempty (i.e., will contain something besides the zero column) if and only if \( A \) does not have full column rank. It should now be clear that the equations \( Ax = y \) will have a unique solution if and only if \( A \) has full column rank.

An \( m \times n \) matrix will be said to have column rank \( r \) if at most \( r \) columns of \( A \) are linearly independent.

So far we have talked only about the columns of \( A \) and the equation \( Ax = y \). Clearly, similar statements can be made about the rows of \( A \) and the equation

\[
xA = y
\]

(where again the context is used to set the dimensions). Thus the row rank of \( A \) is the maximum number of linearly independent rows of \( A \).

The special case of square matrices (\( m = n \)) is particularly important. The properties of determinants show that a square, full- (column or row)

†The null-space of \( A \) is often also called the kernel of \( A \) and written \( \ker(A) \).

rank matrix can never have a zero determinant and therefore will be nonsingular. A nonsingular matrix is always invertible, so that for nonsingular \( A \) the equation \( Ax = y \) will always be consistent for any \( y \); a solution is \( A^{-1}y \). Moreover, the solution will be unique because \( Ax = 0 \) will imply \( x = A^{-1}0 = 0 \).

On the other hand, if \( A \) is square and not of full rank, its determinant will be zero and \( A \) will be singular. Now the null-space will not be empty, but whether or not we have a solution will depend on whether or not \( y \) lies in \( \mathcal{R}(A) \).

A.14 A Simple Linear Equation

Let

\[
A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 1 \\ 4 & 1 & 1 \end{bmatrix}
\]

1. Does the equation \( Ax = y \) have a solution for \( y = [2 \ 3 \ 5]' \)?
2. If there is more than one solution, give a simple way of describing the family of all possible solutions.
3. Repeat the above questions for \( y = [2 \ 3 \ 6]' \).
4. How many linearly independent columns does \( A \) have? How many linearly independent rows?
5. Repeat all the above questions for the matrix obtained from \( A \) by changing the (2, 2) entry from 1 to 2.

A.15 Range Spaces and Null Spaces

1. Suppose \( x \) is a vector in \( \mathcal{R}(A) \), the null-space of \( A \). Prove that \( x \) is orthogonal to every vector in \( \mathcal{R}(A') \), the (column) range space of \( A' \).
2. Prove the converse: if \( x \) is orthogonal to every vector in \( \mathcal{R}(A') \), then \( x \) is in \( \mathcal{R}(A) \).
3. Make and explore similar statements about \( \mathcal{R}(A) \) and \( \mathcal{R}(A') \).

There are many other results in the theory of linear equations that we shall not enter into here, though some more geometric aspects are also explored in Sec. 6.2.

Elementary Operations and Matrix Factorizations. The question of ways to solve the equation \( Ax = y \) is a vast one (see, e.g., [1]-[3], [12]-[19]), but in all methods it is useful to be able to "simplify" the given matrix \( A \) by performing some elementary, reversible operations on it. As might already be familiar (e.g., from knowledge of determinants), there are three basic elementary row operations: (1) multiplying a row by a constant, (2) interchanging two rows, and (3) adding a multiple of one row to another.
We can also analogously define elementary column operations. These elementary operations are clearly reversible, and combinations of them can be used to induce various desirable properties for a matrix.

A.16 Elementary Matrices

We know that premultiplying a matrix $A$ on the left describes certain operations on the rows of $A$.

1. Find the matrices that describe the above elementary row operations on $A$. These will be called elementary matrices.
2. Show that multiplication on the right by these elementary matrices will correspond to the elementary column operations on $A$.
3. Suppose $A$ is $m \times n$ and has row rank $r < m$. Show that by elementary row operations we can transform $A$ to a matrix with its first $r$ rows linearly independent and the remaining $m - r$ rows identically zero. Why is it that we cannot make more than $m - r$ rows identically zero?

A.17 LDU Decompositions

1. If the first $r$ rows of an $m \times n$ matrix $A$ are linearly independent, where $r$ is the row rank of $A$, show that we can always find a sequence of elementary operations that will transform $A$ to an upper triangular matrix.

   Hence, show that such a matrix $A$ can always be written in the form $A = LDU$, where $L$ is an $m \times r$ lower triangular matrix with diagonal elements equal to unity, $U$ is an $r \times n$ upper triangular matrix with unity diagonal entries, and $D$ is an $r \times r$ diagonal matrix. (Note, of course, that the $D$ can be absorbed into $L$ or $U$ to obtain an $LU$ factorization.)

2. Assume that $m = n = r$. Show that if there are two factorizations $A = L_1 D_1 U_1 = L_2 D_2 U_2$, then we must have $L_1 = L_2$, $D_1 = D_2$, $U_1 = U_2$. That is, the $LDU$ factorization (or decomposition) is unique.

   Hint: $L_1^{-1} L_2 = D_2^{-1} D_1^{-1} U_1^{-1} U_2$.

3. If $A$ is $n \times n$, of full rank, and if $A = LDU$, what can you say about the $LDU$-factorization of $A_n$, the matrix formed by taking the first $m$ rows and first $m$ columns of $A$, $1 \leq m \leq n$.

5 SOME RESULTS ON RANK

A perhaps surprising, but fundamental, result of the theory of linear equations is that for any matrix $A$,

\[
\text{the column rank = the row rank}
\]

Proofs can be found in [1, Chap. 5] and [2, Chap. 2]; a short elegant proof is given by H. Liebeck (Am. Math. Mon., Vol. 73, p. 1114, 1966). It is also true that these ranks are equal to the so-called determinantal rank of $A$, which is $\alpha_1$ if all submatrices formed from a $1 \times 1$ submatrix are singular and at least one $2 \times 2$ submatrix is nonsingular.

In future, therefore, we can just talk of the rank of a matrix $A$, which we shall write as $r(A)$.

A.18 Rank of Products

1. Prove that $r(AB) \leq \min \{r(A), r(B)\}$ unless $B$ is nonsingular, in which case $r(AB) = r(A)$.
2. If $A$ is $m \times n$, $m > n$ and of full rank, prove that $A^T A$ is nonsingular.

   Hint: Consider $x^T A^T A x$.

3. Sylvester’s inequality [1, Sec. 5.5]. If $A$ is $m \times n$ and $B$ is $n \times p$, show that $r(A) + r(B) - n \leq r(AB) \leq \min \{r(A), r(B)\}$.

A.19 Matrices of Rank $\alpha$

1. If $u$ is $n \times 1$ and $v$ is $m \times 1$, show that the matrix $uv^T$ has rank 1.

   Show, conversely, that any matrix of rank 1 can be expressed as the product of a column matrix and a row matrix.

2. Show that in general $r(A + B) \leq r(A) + r(B)$.

3. Show that a matrix $A$ has rank $\alpha$ if and only if it can be written in the form

\[
A = \sum_i \alpha_i u_i v_i^T
\]

where $[u_i]$ and $[v_i]$ are sets of independent column vectors.

It is important to note that the actual determination of the rank of a matrix is a notoriously dangerous numerical problem; the rank is very sensitive to small changes in the element values, a very simple example being the matrix diag $[1, e]$. The $QR$ factorization of Exercise A.4 is often used for rank determination, though singular value criteria (cf. Sec. 15, p. 667) are being increasingly advocated. For some perspective on these choices, see the remarks in [17, p. 11.23].

6 SOME FORMULAS ON INVERSES

A.20 A Matrix Inverse

Let $A$ be a nonsingular matrix. Let $u$ and $v$ be column matrices, and assume that $A + uv^T$ is nonsingular. Verify by using the definition of the inverse of a matrix that

\[
(A + uv^T)^{-1} = A^{-1} - \frac{(A^{-1} u v^T A^{-1})}{1 + v^T A^{-1} u}
\]
The point of this formula is that if $A^{-1}$ is known, the inverse of $A$ augmented by a rank-1 matrix can be obtained by a simple modification of the known $A^{-1}$.

A.21 A Generalization—The Modified Matrices Formula

$A$ and $C$ are nonsingular $m \times m$ and $n \times n$ matrices, respectively. Verify, by using the definition of the matrix inverse, that

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(\lambda A^{-1}B + C^{-1})^{-1}D A^{-1}.$$

This general formula has many applications in system theory, especially in the form

$$[I + C(sI - A)^{-1}B]^{-1} = I - C(sI - A + BC)^{-1}B.$$

Its exact origin is hard to trace; it is usually attributed to Woodbury (1950), but according to Householder [15], it may be in the work of Schur. In any case, it is a very useful result. See also Gill et al., Math. Comp., 29, pp. 1051-1077, 1975, for methods of modifying LDU factorizations.

A.22 Inverse of Block Matrices

1. Show, when $A^{-1}$ and $B^{-1}$ exist, that

$$
\begin{bmatrix}
A & 0 \\
C & B
\end{bmatrix}^{-1} =
\begin{bmatrix}
A^{-1} & 0 \\
-B^{-1}CA^{-1} & B^{-1}
\end{bmatrix}
$$

and

$$
\begin{bmatrix}
A & D \\
0 & B
\end{bmatrix}^{-1} =
\begin{bmatrix}
A^{-1} - A^{-1}DB^{-1} & B^{-1} \\
0 & B^{-1}
\end{bmatrix}.
$$

2. If $A^{-1}$ exists, show that

$$
\begin{bmatrix}
A & D \\
C & B
\end{bmatrix}^{-1} =
\begin{bmatrix}
A^{-1} + EA^{-1}F & -EA^{-1} \\
-\Delta^{-1}F & \Delta^{-1}
\end{bmatrix},
$$

where $\Delta = B - CA^{-1}D$, $E = A^{-1}D$, and $F = CA^{-1}$. Show that if $B^{-1}$ exists, the $(1, 1)$ block element of the inverse can also be written as

$$[A - DB^{-1}C]^{-1}$. $\Delta$ is known as the Schur complement of $A$.

7. CHARACTERISTIC POLYNOMIALS AND RESOLVENTS

For any square $n \times n$ matrix $A$, the polynomial

$$
\det(A - sI) = \frac{1}{n} \det(sI - A)
$$

is called the characteristic polynomial of $A$. The $n$ roots of $\det(A - sI)$ [i.e., the solutions of the characteristic equation $\det(A - sI) = 0$] are called the eigenvalues of $A$. Thus we can write

$$
\det(A - sI) = (s - \lambda_1)(s - \lambda_2) \cdots (s - \lambda_n)
$$

where the $\lambda_k$ are the eigenvalues of $A$. The existence of $n$ roots is, of course, due to the fact that we allow the coefficients and the roots of $\det(A)$ to be complex numbers.

The matrix $(sI - A)^{-1}$ is known as the resolvent of $A$.

A.23 Some Resolvent Identities

1. Verify (by multiplying both sides by $sI - A$) that

$$
\text{Adj}(sI - A) = [s^{n-1} + (A + a_1i) s^{n-2} + \cdots + (A^{n-1} + a_{n-1}i)]
$$

$$
= [s^n + (s + a_n)s^{n-1} + \cdots + (s^{n-1} + a_2 s^{n-2} + \cdots + a_1) s + a_0]
$$

(Note that these expressions differ only by the substitution $sI \rightarrow A$.

The real reason for this can be seen from an alternative, deeper approach to this whole topic, based on some results from polynomial matrix theory—see Exercise 6.3.7.)

2. Show that the coefficients, say $S_k$, of the powers $s^{-1}$ in the first expression can be recursively computed as

$$
S_1 = I, \quad S_2 = S_1 A + a_1 I, \quad S_3 = S_2 A + a_1 S_1, \quad \cdots,
$$

$$
S_{n+1} = S_n A + a_{n-1} I, \quad 0 = S_n A + a_n I.
$$

These formulas show that the adjugate matrix can be recursively determined from knowledge of $A$ and the coefficients of $\det(A)$.

3. Show that the above formulas can be used to obtain a recursive method of simultaneously calculating the coefficients $[a_i]$ of the characteristic polynomial and the matrices $[S_i]$ that enter into the numerator of $(sI - A)^{-1}$ via the identities

$$
\frac{1}{n} \frac{1}{(n-1)!} \text{tr}(S_{n-1} A) = \frac{1}{n} \frac{1}{(n-2)!} \text{tr}(S_{n-2} A) = \cdots = S_1 = I.
$$

$$
\frac{1}{n} \frac{1}{(n-1)!} \text{tr}(S_{n-1}) = \frac{1}{n} \frac{1}{(n-2)!} \text{tr}(S_{n-2}) = \cdots = S_1 = I.
$$