PURE AND APPLIED MATHEMATICS

A Wiley-Interscience Series of Texts, Monographs, and Tracts

Founded by RICHARD COURANT
Editors Emeriti: PETER HILTON and HARRY HOCHSTADT
Editors: MYRON B. ALLEN III, DAVID A. COX, PETER LAX,
JOHN TOLAND

A complete list of the titles in this series appears at the end of this volume.

Positive Linear Systems

Theory and Applications

LORENZO FARINA
SERGIO RINALDI

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York / Chichester / Weinheim / Brisbane / Singapore / Toronto

Appendix B: Elements of Linear Systems Theory

B.1 DEFINITION OF LINEAR SYSTEMS

Linear systems are a particular, but very important, class of dynamic systems. As such, they are characterized by *input*, *state*, and *output* variables, denoted by u, x, and y, respectively. The symbol t denotes time, and can be either an integer (*discrete-time system*) or a real number (*continuous-time system*). We will consider *finite dimensional* systems with a single *input* and a single *output*, that is,

$$u(t) \in \mathbb{R} \quad x(t) \in \mathbb{R}^n \quad y(t) \in \mathbb{R}$$

where the dimension n of the state vector is called *order* of the system.

In discrete-time linear systems, the state vector is updated through a linear equation, called a *state equation*,

$$x(t+1) = Ax(t) + bu(t) \tag{B.1}$$

where A is an $n \times n$ matrix and b is an $n \times 1$ vector, while the output depends on the state and input through a linear equation, called an *output transformation*,

$$y(t) = c^T x(t) + du(t) \tag{B.2}$$

where c^T is a $1 \times n$ row vector and d is a real. Written for each component of the state vector, (B.1) corresponds to

$$\begin{aligned} x_1(t+1) &= a_{11}x_1(t) + \ldots + a_{1n}x_n(t) + b_1u(t) \\ x_2(t+1) &= a_{21}x_1(t) + \ldots + a_{2n}x_n(t) + b_2u(t) \\ &\vdots \\ x_n(t+1) &= a_{n1}x_1(t) + \ldots + a_{nn}x_n(t) + b_nu(t) \end{aligned}$$

while (B.2) becomes

$$y(t) = c_1x_1(t) + \ldots + c_nx_n(t) + du(t)$$

Next, we will consider only time-invariant systems, that is, systems with A, b, c^T, and d constant over time.

Analogously, we can define continuous-time linear systems as systems with the following state equation:

$$\dot{x}(t) = Ax(t) + bu(t) \tag{B.3}$$

where $\dot{x}(t)$ is the derivative of $x(t)$ with respect to time, and

$$y(t) = c^Tx(t) + du(t) \tag{B.4}$$

Thus, continuous-time and discrete-time linear systems are identified by the quadruple (A, b, c^T, d), which can be conveniently ordered as follows:

$$A = \boxed{} \quad b = \boxed{}$$

$$c^T = \boxed{} \quad d = \boxed{}$$

They are often graphically represented in one of the forms shown in *Fig. B.1*. The first form shows only the input and output variables, called *external* variables, since they are those through which the system interacts with the rest of the world. The second form also shows the state variables, called *internal*.

In many systems, the input does not directly influence the output, that is, $d = 0$. Such systems, called *proper*, are identified by the triple (A, b, c^T), while those with $d \neq 0$, called *improper*, are identified by the quadruple $(A, b, c^T d)$. Systems without input $(b = 0, d = 0)$, are called *autonomous* and described, by the pair (A, c^T).

Next, we will discuss the main features of linear systems, starting from those depending only on the matrix A (reversibility and internal stability), and continuing with those characterized by the pair (A, b) (reachability), or by the pair (A, c^T) (observability), and ending with those depending on the triple (A, b, c^T) or on the quadruple (A, b, c^T, d) (external stability, minimum phase, minimality, etc.).

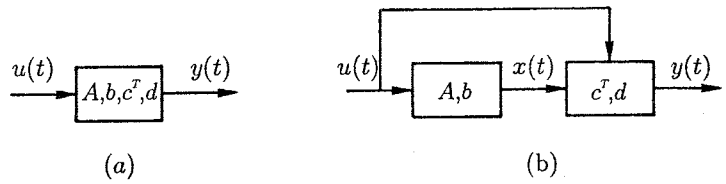

Figure B.1 Representations of a linear system: (a) compact form; (b) disaggregated form in which the first block represents the state equation and the second the output transformation.

EXAMPLE 1 (Newton's law)
Suppose that a point mass m moves without friction along a straight line and that a force $u(t)$ is applied to it along the same direction. If $y(t)$ is the position of the point mass, measured with respect to a fixed point, Newton's law states that

$$\dot{x}_1(t) = x_2(t)$$

$$\dot{x}_2(t) = \frac{1}{m}u(t)$$

while the output transformation is

$$y(t) = x_1(t)$$

In conclusion, Newton's law is described by a proper linear system identified by the triple

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad b = \begin{pmatrix} 0 \\ 1/m \end{pmatrix}$$

$$c^T = \begin{pmatrix} 1 & 0 \end{pmatrix}$$

♣

EXAMPLE 2 (Fibonacci's rabbits)
Maybe the oldest example of a discrete-time linear system is that concerning a rabbit population described by *Leonardo Fibonacci* from Pisa (1180–1250) in the book *Liber Abaci*. Let us denote the year by t, the number of pairs of young and adult rabbits at the beginning of year t by $x_1(t)$ and $x_2(t)$, the number of pairs of adult rabbits killed by hunters during year t by $u(t)$, and the total number of pairs of rabbits with $y(t)$. The assumptions made by Fibonacci (some of them are a bit extreme) are the following:

- Young rabbits do not reproduce.
- Young rabbits become adult after 1 year.

- Adult rabbits reproduce once a year.
- Each pair of adult rabbits generate a pair of young rabbits.
- Rabbits do not die.

Under these assumptions, a simple balance of young and adult rabbits leads to the following state equation:

$$\begin{aligned} x_1(t+1) &= x_2(t) \\ x_2(t+1) &= x_1(t) + x_2(t) - u(t) \end{aligned}$$

while the output transformation is

$$y(t) = x_1(t) + x_2(t)$$

Thus, the system is proper and identified by the triple

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \quad b = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$
$$c^T = \begin{pmatrix} 1 & 1 \end{pmatrix}$$

If we assume that at time $t = 0$ there is only one pair of young rabbits, that is,

$$x(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and that there are no hunters [$u(t) = 0$ for all t], the state equation and the output transformation can be used recursively to determine the growth of the population, namely, the output sequence $y(0), y(1), y(2), \ldots$. The reader can easily verify that each element of the sequence is equal to the sum of the two previous elements (Fibonacci's series). ♣

B.2 ARMA MODEL AND TRANSFER FUNCTION

The definition of linear systems given in the previous paragraph is often termed *internal* since it explicitly refers to the state of the system. For the same reason, the alternative definition involving only input and output variables is called *external*. The definition is as follows: In a discrete-time system of order n, a weighted sum of $(n+1)$ subsequent input values equals, at any time t, a weighted sum of the corresponding output values, namely,

$$y(t)+\alpha_1 y(t-1)+\cdots+\alpha_n y(t-n) = \beta_0 u(t)+\beta_1 u(t-1)+\cdots+\beta_n u(t-n) \quad \text{(B.5)}$$

If $\beta_0 \neq 0$, the input $u(t)$ directly influences the output $y(t)$ and, therefore, the system is improper. If, on the contrary, $\beta_0 = 0$ the system is proper. Equation (B.5) is often used in the form

$$y(t) = \sum_{i=1}^{n}(-\alpha_i)y(t-i) + \sum_{i=0}^{n}\beta_i u(t-i) \quad \text{(B.6)}$$

in which the first term of the right-hand side is called *autoregression* and the second *moving average*. For this reason, Eq. (B.5) is known as the autoregressive moving average model often abbreviated as the *ARMA model*. The continuous-time analog of Eq. (B.5) is the differential equation of order n

$$y^{(n)}(t)+\alpha_1 y^{(n-1)}(t)+\cdots+\alpha_n y^{(0)}(t)=\beta_0 u^{(n)}(t)+\beta_1 u^{(n-1)}(t)+\cdots+\beta_n u^{(0)}(t) \quad \text{(B.7)}$$

where $u^{(i)}(t)$ and $y^{(i)}(t)$ are the ith derivatives of input and output. Also, this model will be called (even if unproperly) the ARMA model.

The interpretation of Newton's law (see *Example* 1) can be completed by noting that the relationship

$$\ddot{y}(t) = \frac{1}{m}u(t)$$

is a particular case of Eq. (B.7) (MA model, *i.e.* ARMA model without the autoregressive term). As for the Fibonacci's rabbits (*Example* 2) the ARMA model is (easy to check)

$$y(t) - y(t-1) - y(t-2) = -u(t-1) - u(t-2)$$

This ARMA model can be used to recursively generate the Fibonacci's series [by annihilating the input and setting $y(0) = y(1) = 1$].

Equations (B.5) and (B.7) can be written in the general form

$$D(p)y(t) = N(p)u(t) \quad \text{(B.8)}$$

where $D(\cdot)$ and $N(\cdot)$ are two polynomials of degree n

$$D(p) = p^n + \alpha_1 p^{n-1} + \cdots + \alpha_n$$
$$N(p) = \beta_0 p^n + \beta_1 p^{n-1} + \cdots + \beta_n$$

and p is a "shift" operator for the discrete-time case (*i.e.*, $py(t) = y(t+1)$, $p^2y(t) = y(t+2), \ldots$,) and a "derivative" operator for the continuous-time case [*i.e.*, $py(t) = \dot{y}, p^2y(t) = \ddot{y}(t), \ldots$]. An ARMA model is therefore equivalent to two polynomials $D(\cdot)$ and $N(\cdot)$ or, alternatively, to $2n+1$ parameters β_0 and $\{\alpha_i, \beta_i\}, i = 1, \ldots, n$. Usually, the symbol p in (B.8) is replaced by $z[s]$ when dealing with discrete [continuous] -time systems. Thus, for example, Newton's law (*Example* 1) is described by

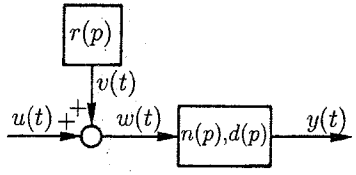

Figure B.2 Decomposition of an unreduced ARMA model $[D(p), N(p)]$ in a reduced ARMA model $[n(p), d(p)]$, and in an AR model $[r(p)]$.

$$D(s) = s^2 \quad N(s) = \frac{1}{m}$$

while Fibonacci's rabbits (*Example* 2) are described by

$$D(z) = z^2 - z - 1 \quad N(z) = -z - 1$$

If the two polynomials $D(\cdot)$ and $N(\cdot)$ are coprime (*i.e.*, they do not have common zeros), the ARMA model is said to be in *reduced form* or, simply, *reduced*. In such a case, the polynomial $D(\cdot)$ being monic, the knowledge of the pair $[D(\cdot), N(\cdot)]$ is equivalent to the knowledge of the ratio $D(\cdot)/N(\cdot)$, called the *transfer function* and denoted by $G(\cdot)$, that is,

$$G(p) = \frac{N(p)}{D(p)} \tag{B.9}$$

If the ARMA model is not in reduced form, namely, if

$$\begin{aligned} D(p) &= r(p)d(p) \\ N(p) &= r(p)n(p) \end{aligned} \tag{B.10}$$

with $n(\cdot)$ and $d(\cdot)$ coprime, the transfer function (B.9) is equal to $n(p)/d(p)$. The roots of $n(\cdot)$ and $d(\cdot)$ are called, respectively, *zeros* and *poles* of the transfer function. If we take into account Eq. (B.8) and (B.10), one can check that an unreduced ARMA model can be decomposed, as shown in *Fig. B.2*, in a reduced ARMA model identified by the pair of coprime polynomials $[d(\cdot), n(\cdot)]$.

$$d(p)y(t) = n(p)w(t) \tag{B.11}$$

and in an AR model determined by the polynomial $r(\cdot)$,

$$r(p)v(t) = 0 \tag{B.12}$$

In fact, if (B.11) is multiplied by $r(p)$ and (B.12) is taken into account together with the fact that

$$w(t) = v(t) + u(t)$$

Eq. (B.8) with $D(p)$ and $N(p)$ given by (B.10) is obtained. *Figure B.2* clearly shows that the transfer function $G(p) = n(p)/d(p)$ exclusively describes the reduced part of the ARMA model. In other words, if only the transfer function is known, it is not possible to compute the output of the system from its input, unless the signal $v(\cdot)$ is identically zero, which occurs when the initial condition of the AR model (B.12) is zero.

B.3 COMPUTATION OF TRANSFER FUNCTIONS AND REALIZATION

Since we have given two different definitions of a dynamical system (one internal and one external) it is important to show how it is possible to move from one description to the other.

The problem of the computation of the ARMA model and of the transfer function of a system, given the quadruple (A, b, c^T, d), can be fully understood only after introducing the notions of reachability and observability. For the moment, notice that Eqs. (B.1) and (B.3), recalling the meaning of the operator p, can be written in the form

$$px(t) = Ax(t) + bu(t)$$

so that

$$x(t) = (pI - A)^{-1}bu(t)$$

From (B.2) and (B.4), it follows that

$$y(t) = [c^T(pI - A)^{-1}b + d]u(t)$$

which, compared with (B.8) and (B.9) yields

$$G(p) = c^T(pI - A)^{-1}b + d \tag{B.13}$$

The inverse of the $n \times n$ matrix $(pI - A)$ can be written in the form

$$(pI - A)^{-1} = \frac{1}{\Delta_A}P(p)$$

where $P(p)$ is an $n \times n$ matrix of polynomials with a degree $< n$ and $\Delta_A(p)$ is the characteristic polynomial of A. Then, $\Delta_A(p)$ and $P(p)$ can be computed using the following formulas (due to Souriau):

$$\Delta_A(p) = p^n + \alpha_1 p^{n-1} + \cdots + \alpha_n$$

$$P(p) = P_0 p^{n-1} + P_1 p^{n-2} + \cdots + P_{n-1}$$

where

$$
\begin{aligned}
P_0 &= I & \alpha_1 &= -\mathrm{tr}(P_0 A) \\
P_1 &= P_0 A + \alpha_1 I & \alpha_2 &= -\frac{1}{2}\mathrm{tr}(P_1 A) \\
P_2 &= P_1 A + \alpha_2 I & \alpha_3 &= -\frac{1}{3}\mathrm{tr}(P_2 A) \\
&\vdots & &\vdots \\
P_{n-1} &= P_{n-2} A + \alpha_{n-1} I & \alpha_n &= -\frac{1}{n}\mathrm{tr}(P_{n-1} A)
\end{aligned}
$$

If the transfer function $G(p) = n(p)/d(p)$ computed by means of (B.13) has the polynomial $d(p)$ with degree n, then, from Souriau's formulas it follows that:

$$d(p) = D(p) = \Delta_A(p)$$

that is the ARMA model $[D(p), N(p)]$ of the system is in reduced form and the poles of the transfer function are n and coincide with the eigenvalues of matrix A. In contrast, if the degree of $d(\cdot)$ is $< n$, the poles of the transfer function are $< n$ but still coincide with some of the eigenvalues of matrix A.

The problem of the computation of the quadruple (A, b, c^T, d) from an ARMA model $[D(p), N(p)]$ is known as the *realization* problem [the quadruple (A, b, c^T, d), which solves the problem, is also called realization]. The solution of such a problem is not unique, so that it is particularly interesting to determine the realization with minimal dimension. In order to deal with this problem, it is necessary, however, to be aware of the notions of reachability and observability. For the moment, let us state that a particular realization, called *control canonical form*, of a reduced ARMA model

$$
\begin{aligned}
D(p) &= p^n + \alpha_1 p^{n-1} + \cdots + \alpha_n \\
N(p) &= \beta_0 p^n + \beta_1 p^{n-1} + \cdots + \beta_n
\end{aligned}
\tag{B.14}
$$

is the quadruple

$$
A_c = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & & \dots \\ 0 & 0 & 0 & \dots & 1 \\ -\alpha_n & -\alpha_{n-1} & -\alpha_{n-2} & \dots & -\alpha_1 \end{pmatrix} \quad b_c = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}
$$

$$
c_c^T = \begin{pmatrix} \gamma_n & \gamma_{n-1} & \gamma_{n-2} & \dots & \gamma_1 \end{pmatrix} \quad d_c = \beta_0
$$

with

$$\gamma_i = \beta_i - \beta_0 \alpha_i \quad i = 1, \dots, n$$

Notice that the order n of the control canonical form is equal to the "memory" of the autoregressive component of the ARMA model.

A second realization, called *reconstruction canonical form*, is the following

$$
A_r = \begin{pmatrix} 0 & 0 & \dots & 0 & -\alpha_n \\ 1 & 0 & \dots & 0 & -\alpha_{n-1} \\ 0 & 1 & \dots & 0 & -\alpha_{n-2} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -\alpha_1 \end{pmatrix} \quad b_r = \begin{pmatrix} \gamma_n \\ \gamma_{n-1} \\ \gamma_{n-2} \\ \vdots \\ \gamma_1 \end{pmatrix}
$$

$$
c_r^T = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \end{pmatrix} \quad d_r = \beta_0
$$

with

$$\gamma_i = \beta_i - \beta_0 \alpha_i \quad i = 1, \dots, n$$

It is worth noting that

$$(A_r, b_r, c_r^T, d_r) = (A_c^T, c_c, b_c^T, d_c)$$

which is a formula that we will recall when discussing the duality principle.

EXAMPLE 3 (Fibonacci's rabbits)
Consider the ARMA model

$$D(z) = z^2 - z - 1 \quad N(z) = -z - 1$$

which, as previously shown, is the ARMA model describing the growth of the Fibonacci's rabbits (see *Example* 2). The control and reconstruction canonical forms of this ARMA model are

$$
\begin{aligned}
A_c &= \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \qquad b_c = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
c_c^T &= \begin{pmatrix} -1 & -1 \end{pmatrix}
\end{aligned}
$$

and

$$
\begin{aligned}
A_r &= \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \quad b_r = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \\
c_r^T &= \begin{pmatrix} 0 & 1 \end{pmatrix}
\end{aligned}
$$

and are, therefore, different from the triple (A, b, c^T) used in *Example* 2.

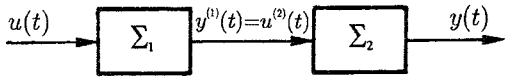

Figure B.3 Two systems connected in series.

B.4 INTERCONNECTED SUBSYSTEMS AND MASON'S FORMULA

Very often a dynamic system Σ is composed of interconnected subsystems Σ_i. Two dynamic systems, Σ_1 and Σ_2, can be interconnected in three ways: in *series*, in *parallel*, and in *feedback*. If $x^{(1)}$ and $x^{(2)}$ are the state vectors of Σ_1 and Σ_2, the state vector x of Σ is $x = [x^{(1)^T}\ x^{(2)^T}]^T$. Thus, if $\Sigma_i = (A_i, b_i, c_i^T, d_i)$, $i = 1, 2$, are the two subsystems, we are interested in the determination of the interconnected system $\Sigma = (A, b, c^T, d)$.

Series

Two systems are connected in series (*Fig. B.3*) when the output of the first system is the input of the second system.

The state equations of Σ are therefore

$$\dot{x}^{(1)}(t) = A_1 x^{(1)}(t) + b_1 u(t)$$

$$\dot{x}^{(2)}(t) = A_2 x^{(2)}(t) + b_2(c_1^T x^{(1)}(t) + d_1 u(t))$$

while the output transformation is

$$y(t) = c_2^T x^{(2)}(t) + d_2(c_1^T x^{(1)}(t) + d_1 u(t))$$

In conclusion, Σ is identified by the following quadruple:

$$A = \begin{pmatrix} A_1 & \mathbf{0} \\ b_2 c_1^T & A_2 \end{pmatrix} \quad b = \begin{pmatrix} b_1 \\ b_2 d_1 \end{pmatrix}$$
$$c^T = \begin{pmatrix} d_2 c_1^T & c_2^T \end{pmatrix} \quad d = (d_1 d_2)$$

Observe that matrix A is in block triangular form, so that its eigenvalues are those of matrices A_1 and A_2.

Parallel

Two systems are connected in parallel (*Fig. B.4*) when they have the same input and the output of the overall system is the sum of their outputs.

It is straightforward to check that Σ is identified by the following four matrices

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$
$$c^T = \begin{pmatrix} c_1^T & c_2^T \end{pmatrix} \quad d = (d_1 + d_2)$$

Also, in this case matrix A is block triangular (actually, diagonal) so that its eigenvalues are those of matrices A_1 and A_2.

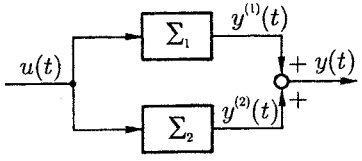

Figure B.4 Two systems connected in parallel.

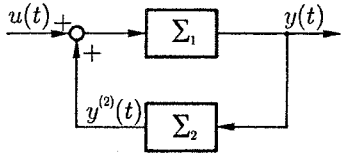

Figure B.5 Two systems connected in feedback (Σ_1 is the forward path and Σ_2 the feedback path).

Feedback

Two systems are connected in feedback (*Fig. B.5*) when the input of the first system is the sum of an external input u and of the output of the second system and the input of the second system is the output of the first one.

Obviously, interconnected subsystems can also be studied from the point of view of their external behavior. Actually, the ARMA model and the transfer function of a system Σ can be easily determinated from the ARMA models and the transfer functions of all its subsystems Σ_i. To verify this statement, we first analyze the cases of series, parallel, and feedback connections of two subsystems.

Series

With reference to *Fig. B.3*, let $\Sigma_1 = [D_1(p), N_1(p)]$ and $\Sigma_2 = [D_2(p), N_2(p)]$.

This means that the ARMA model of the first subsystem is

$$D_1(p)y^{(1)}(t) = N_1(p)u(t)$$

By applying to both sides of this equation the operator $N_2(p)$ and by noting that $y^{(1)} = u^{(2)}$, we obtain

$$N_2(p)D_1(p)u^{(2)}(t) = N_2(p)N_1(p)u(t)$$

But $N_2 D_1 = D_1 N_2$ and $N_2 N_1 = N_1 N_2$ since deriving (or shifting) a function first r times and then s times is equivalent to deriving (or shifting) it first s times and then r times. Thus, we can write

$$D_1(p)N_2(p)u^{(2)}(t) = N_1(p)N_2(p)u(t)$$

On the other hand, the ARMA equation of the second subsystem is

$$D_2(p)y(t) = N_2(p)u^{(2)}(t)$$

so that, finally, we obtain

$$D_1(p)D_2(p)y(t) = N_1(p)N_2(p)u(t)$$

In other words, if two systems Σ_1 and Σ_2 are connected in series, the resulting system Σ is characterized by an ARMA model identified by the following two polynomials:

$$D(p) = D_1(p)D_2(p) \quad N(p) = N_1(p)N_2(p)$$

This means that the transfer function $G(p) = N(p)/D(p)$ of Σ can be obtained by multiplying the two transfer functions $G_1(p)$ and $G_2(p)$ of the two subsystems, that is,

$$G(p) = G_1(p)G_2(p)$$

This result allows one to conclude that the order in which the two systems are connected is not relevant when computing the transfer function of the resulting system.

Parallel

With reference to *Fig. B.4*, proceeding as in the case of the series connection, it is easy to show that the transfer function of Σ is

$$G(p) = G_1(p) + G_2(p)$$

In other words, the transfer function of a system composed of two systems connected in parallel is the sum of their transfer functions.

Feedback

In the case of two systems Σ_1 and Σ_2 connected in feedback, as shown in *Fig. B.5*, one obtains

$$G(p) = \frac{G_1(p)}{1 - G_1(p)G_2(p)}$$

This formula is very useful in the analysis of feedback systems. It holds for the connection shown in *Fig. B.5* where the *feedback* is called *positive* since the signal $y^{(2)}$ coming from the feedback path is summed to the external signal u. In contrast, if one considers the *negative feedback*

$$u^{(1)} = u - y^{(2)}$$

the formula to be used is the following:

$$G(p) = \frac{G_1(p)}{1 + G_1(p)G_2(p)}$$

This formula is often described by saying that the transfer function of a system with a negative feedback is the ratio between the transfer function of the direct path (G_1) and the loop transfer function (G_1G_2) plus one (G_1G_2 is the loop transfer function because, in the loop, the two systems Σ_1 and Σ_2 are connected in series).

Mason's formula

Mason's formula allows one to compute the transfer function $G(p)$ of any system composed of interconnected subsystems. Under the nonlimiting assumption that signals are only summed (and not subtracted) the formula is

$$G(p) = \frac{\sum_k C_k(p)\Delta_k(p)}{\Delta(p)}$$

where $C_k(p)$, $\Delta(p)$, and $\Delta_k(p)$ are called, respectively, *transfer function of the kth direct input–output path, determinant* of the system, and *reduced determinant* with respect to the kth direct path. The transfer function $C_k(p)$ is simply the product of the transfer functions of all the systems composing the kth direct (*i.e.*, not containing cycles) path from input to output. The determinant $\Delta(p)$ is given by

$$\Delta(p) = 1 - \sum_i L_i(p) + \sum_i \sum_j L_i(p)L_j(p) - \sum_i \sum_j \sum_k L_i(p)L_j(p)L_k(p) + \cdots$$

where $L_i(p)$ is the transfer function of the ith closed path (loop), that is, the product of the transfer functions of all the subsystems composing the ith closed path exiting in the system. The first sum in the formula concerns all the loops, the second all the disjoint pairs of loops (*i.e.*, loops that do not touch each other) and so on. Finally, the reduced determinant Δ_k is the determinant Δ without all the terms corresponding to loops that are touched by the kth direct path. On occasions, it may not be easy to find all the direct paths and all the loops by inspection of the graph representing the interconnected system. However, in many cases of practical interest, Mason's formula is straightforward to apply, particularly when there are no disjoint loops.

B.5 CHANGE OF COORDINATES AND EQUIVALENT SYSTEMS

The quadruple (A, b, c^T, d) describing a linear system depends on the units chosen for time, input, state and output variables and on the order in which the state variables are listed. But, the choice of the variables to be considered as state variables is also not unique and has an impact on the quadruple (A, b, c^T, d) that identifies the system. For example, in a chemical reactor characterized by two species, one can consider as state variables the concentrations x_1 and x_2 of such species or, alternatively, their sum z_1 and their difference z_2. Obviously, the quadruple (A, b, c^T, d) corresponding to the state variables (x_1, x_2) is different from that corresponding to

the state variables (z_1, z_2) while the system, from a physical point of view, is the same. For this reason, the two quadruples are called *equivalent*. In order to find the relationship among equivalent quadruples, it is necessary to determine the effect of a change of coordinates

$$z = Tx$$

In the case of the chemical reactor, for example, the change of coordinates $z = Tx$ is given by

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

It is straightforward to check that a change of coordinates $z = Tx$ transforms the discrete-time system (B.1), (B.2) into the equivalent system

$$z(t+1) = TAT^{-1}z(t) + Tbu(t)$$
$$y(t) = c^T T^{-1} x(t) + du(t)$$

Analogously, the continuous-time system (B.3), (B.4) is transformed into the system

$$\dot{z}(t) = TAT^{-1}z(t) + Tbu(t)$$
$$y(t) = c^T T^{-1} x(t) + du(t)$$

In conclusion, a change of coordinates $z = Tx$ transforms the quadruple (A, b, c^T, d) into the quadruple $(TAT^{-1}, Tb, c^T T^{-1}, d)$.

B.6 MOTION, TRAJECTORY, AND EQUILIBRIUM

Once the initial state $x(0)$ and the input $u(t)$ for $t \geq 0$ are fixed, the state equations (B.1) and (B.3) admit a unique solution $x(t)$ for $t \geq 0$ (this is pretty obvious for discrete-time systems while, for continuous-time systems, it follows from results of existence and uniqueness of ordinary differential equations). The function $x(\cdot)$ thus obtained, is called *motion*, while the set $\{x(t), t \geq 0\}$ in the space $\mathbb{R}^n$ is called *trajectory*. The trajectory of a continuous-time system is a line originating from point $x(0)$ and with a specified direction [see *Fig. B.6*(a)]. In the case of discrete-time systems, the trajectory is a sequence of points $\{x(0), x(1), \ldots\}$ that, for the sake of clarity, are often linked one to the next with a segmented straight line, as shown in *Fig. B.6*(b).

As shown in *Fig. B.6*(a), it may happen that the trajectory passes through the same point x at different instants of time t_1, t_2, and so on, and that the vectors tangent to the trajectory at that point are different. In fact, the tangent vector is $\dot{x}$ and therefore, the case shown in *Fig. B.6*(a) can occurs if

$$Ax + bu(t_1) \neq Ax + bu(t_2)$$

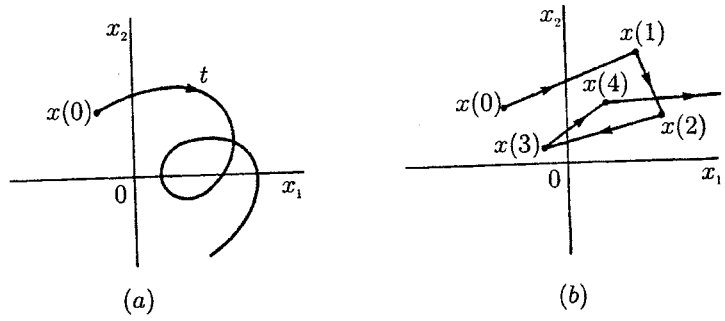

Figure B.6 Trajectories of second-order systems: (a) continuous-time systems; (b) discrete-time systems.

that is,

$$bu(t_1) \neq bu(t_2)$$

Obviously, such conditions cannot hold if the input u is constant over time.

It may occur that the motion $x(\cdot)$ corresponding to a particular initial state $x(0)$ and to a particular input function is periodic with period T, that is,

$$x(t) = x(t+T) \quad \forall t$$

In this case, the trajectory is a closed line (cycle) repeatedly visited every T time units. If x is periodic, then so is $\dot{x}$, so that

$$Ax + bu(t) = Ax(t+T) + bu(t+T) \quad \forall t$$

and therefore

$$bu(t) = bu(t+T) \quad \forall t \tag{B.15}$$

This means that a cycle can possibly be obtained only if the input function satisfies condition (B.15). It is important to note that condition (B.15) holds for every periodic input function with period T and, consequently, for any constant input.

A degenerate case occurs when the state of the system does not change over time, so that the cycle is represented by a point $\bar{x}$ called an *equilibrium state*. To this purpose, we give *Definition 1*.

DEFINITION 1 *(equilibrium)*

A system is said to be at *equilibrium* if input and state (and, therefore, also output) are constant, that is, if

$$u(t) = \bar{u} \quad x(t) = \bar{x} \quad y(t) = \bar{y} \quad \forall t$$

Vector $\bar{x}$ is called an equilibrium state.

Since for continuous-time systems, $x(t) = \bar{x}\ \forall t$ implies $\dot{x}(t) = 0$, it follows that in such systems

$$A\bar{x} + b\bar{u} = 0 \tag{B.16}$$

$$\bar{y} = c^T\bar{x} + d\bar{u} \tag{B.17}$$

If A is nonsingular (*i.e.*, if $\det A \neq 0$ or, equivalently, if A has no zero eigenvalues), then there exists a unique solution $\bar{x}$ to Eq. (B.16) for each $\bar{u}$, and, therefore, there is also only one solution $\bar{y}$ of (B.17), which are formally given by

$$\bar{x} = -A^{-1}b\bar{u} \qquad \bar{y} = (d - c^TA^{-1}b)\bar{u} \tag{B.18}$$

In the case A is singular [$\det A = 0$] and $\bar{u}$ is fixed, either no solutions $\bar{x}, \bar{y}$ of (B.16), (B.17) exist or they are infinite.

For discrete-time systems, Eq. (B.16) and (B.17) must be replaced by the relations

$$(I - A)\bar{x} = b\bar{u}$$

$$\bar{y} = c^T\bar{x} + d\bar{u}$$

so that uniqueness of the equilibrium state (and output) for any fixed $\bar{u}$ is guaranteed by nonsingularity of the matrix $(I - A)$, that is,

$$\det(I - A) \neq 0$$

or, equivalently, from the fact that A has no unitary eigenvalues. In such cases, one has

$$\bar{x} = (I - A)^{-1}b\bar{u} \qquad \bar{y} = (d + c^T(I - A)^{-1}b)\bar{u} \tag{B.19}$$

Equations (B.18) and (B.19) show that in nonsingular cases the relationship between input and output at the equilibrium is linear. Since for single input and single output systems it is usual to define the *gain* of the system as the ratio μ between output and input at the equilibrium

$$\mu = \frac{\bar{y}}{\bar{u}}$$

then, for continuous-time systems the following formula holds:

$$\mu = d - c^TA^{-1}b$$

while for discrete-time systems one has

$$\mu = d + c^T(I - A)^{-1}b$$

Obviously, the same formulas show that it is meaningless to define a gain in singular cases.

It is important to note that the calculation of the gain is straightforward whenever the input–output relation (B.7) of a continuous-time system is known, since the equilibrium condition implies $y^{(i)} = u^{(i)} = 0, i = 1, \ldots, n, y^{(0)} = \bar{y}$ and $u^{(0)} = \bar{u}$, so that

$$\mu = \frac{\beta_n}{\alpha_n} \tag{B.20}$$

For discrete-time systems, one has [see (B.6)]

$$\mu = \frac{\sum_{i=0}^{n}\beta_i}{1 + \sum_{i=1}^{n}\alpha_i} \tag{B.21}$$

Note that, denoting with $G(s)$ the transfer function of a continuous-time system, Eq. (B.20) is equivalent to $\mu = G(0)$, so that the gain μ is equal to the value of the transfer function for $s = 0$. For discrete-time systems with transfer function $G(z)$, from (B.21) it follows that $\mu = G(1)$, that is, the gain is equal to the value of the transfer function for $z = 1$.

The gain is also easy to compute in the case of systems composed of interconnected subsystems. It is in fact immediate to verify that the gain μ of a system composed of two subsystems connected in series is the product of the gains of the two subsystems, that is,

$$\mu = \mu_1\mu_2$$

while for parallel connections, the following formula holds

$$\mu = \mu_1 + \mu_2$$

and for feedback connections we have

$$\mu = \frac{\mu_1}{1 + \mu_1\mu_2}$$

B.7 LAGRANGE'S FORMULA AND TRANSITION MATRIX

From state equations of a linear system, it follows that the state at time t is a function of the initial state at time $t = 0$, of the input during the interval of time $[0, t)$ and, obviously, of the considered interval of time t. Obtaining an explicit solution, in the usual sense, of the state equations is possible only in particularly simple cases (typically, for first- and second-order systems). The solution can be, however, specified and written in a particularly useful form for the understanding of many problems and for the proof of several properties. In the case of continuous-time systems, the formula is credited to Lagrange; for the sake of simplicity, we will give the same name to the corresponding formula holding for discrete-time systems.

THEOREM 1 *(Lagrange formula)*

In a continuous-time linear system

$$\dot{x}(t) = Ax(t) + bu(t)$$

the state $x(t)$ for $t \geq 0$ is given by (*Lagrange formula*)

$$x(t) = e^{At}x(0) + \int_0^t e^{A(t-\xi)}bu(\xi)d\xi \tag{B.22}$$

where

$$e^{At} = I + At + A^2\frac{t^2}{2!} + A^3\frac{t^3}{3!} + \cdots$$

Analogously, in a discrete-time linear system

$$x(t+1) = Ax(t) + bu(t)$$

for $t > 0$, the following holds:

$$x(t) = A^t x(0) + \sum_{i=0}^{t-1} A^{t-i-1}bu(i) \tag{B.23}$$

Equation (B.23) is also called the Lagrange formula.

The Lagrange formulas (B.22) and (B.23) can be rewritten in a more compact form as follows:

$$x(t) = \Phi(t)x(0) + \Psi(t)u_{[0,t)}(\cdot) \tag{B.24}$$

where $\Phi(t)$ and $\Psi(t)$ are linear transformations acting, respectively, on the initial state $x(0)$ and on the segment $u_{[0,t)}(\cdot)$ of the input function $u(\cdot)$. By comparing Eq. (B.24) with (B.22) and (B.23), it follows that the matrix $\Phi(t)$, called the *transition matrix*, is given by

$$\Phi(t) = \begin{cases} e^{At} & \text{for continuous-time systems} \\ A^t & \text{for discrete-time systems} \end{cases}$$

Equation (B.24) states that the state of the system is at any time given by the sum of two terms, the first linearly depending on the initial state and the second linearly depending on the input. These two contributions to the motion of a dynamic system are called, respectively, *free motion* and *forced motion*. The reason for such names is obvious: $\Phi(t)x(0)$ represents the evolution of the "free" system, that is, of the system with a null input (or without input, as usually said), while $\Psi(t)u_{[0,t)}(\cdot)$ represents the evolution of the system initially at rest $[x(0) = 0]$ but forced by the input $u(\cdot)$. By applying the output transformation given by (B.24), one obtains

$$y(t) = c^T\Phi(t)x(0) + c^T\Psi(t)u_{[0,t)}(\cdot) + du(t) \tag{B.25}$$

which clearly shows that the output is the sum of free and forced evolutions.

Equation (B.25), if appropriately interpreted, allows us to formulate the so-called *superposition principle*, often recalled when talking about dynamic linear systems.

THEOREM 2 *(superposition principle)*

If the pair $(x'(0), u'(\cdot))$ gives rise to the output $y'(\cdot)$ and the pair $(x''(0), u''(\cdot))$ to the output $y''(\cdot)$, then the pair $[\alpha x'(0) + \beta x''(0), \alpha u'(\cdot) + \beta u''(\cdot)]$ gives rise to the same linear combination $\alpha y'(\cdot) + \beta y''(\cdot)$ of the outputs.

Among the most frequently used formulas of any discipline, it is almost invariably possible to find some that are nothing but the Lagrange formula applied to simple first- or second-order systems. The law of the falling of bodies, the law of charge and discharge of a capacitor, the law governing the rise of temperature in a thermometer, and the one describing the release from a reservoir, are just examples of the application of the Lagrange formula to continuous-time systems. But the same holds for laws regarding discrete-time systems, as shown in *Example 4*.

EXAMPLE 4 (amortization)

If a debt D is amortized by returning for N consequent years an amount A, the debt x varies during the years according to the equation

$$x(t+1) = (1+\rho)x(t) - A$$

where ρ is the annual interest rate. One can then apply the Lagrange formula (B.23) with $t = N$ and $u(i) = A$ to such a system thus obtaining

$$x(N) = (1+\rho)^N D - A\sum_{i=0}^{N-1}(1+\rho)^{N-i-1}$$

By imposing the final condition $x(N) = 0$ and by solving with respect to A, one obtains the famous amortization formula

$$A = \frac{\rho}{1-(1+\rho)^{-N}}D$$

♣

The Lagrange formula should *not* be regarded as a formula useful for the calculation of the state evolution of a linear system. Such a statement is particularly simple to illustrate in the case of discrete-time systems. In fact, in such systems, the state evolution can be computed by a repeated use for $t = 0, 1, 2$, and so on, of the equation

$$x(t+1) = Ax(t) + bu(t)$$

By doing so, at each step n^2+n multiplications are needed and about the same number of sums, so that the computation of $x(1), x(2), \ldots, x(N)$ requires $Nn(n+1)$

elementary operations. The computation of the same vectors by means of (B.23) is indeed much more onerous since the evaluation the matrix powers $A^2, A^3, \ldots, A^N$ is an operation requiring Nn^3 elementary operations. The importance of the Lagrange formula is then mostly related to conceptual and formal aspects of the theory of linear systems.

B.8 REVERSIBILITY

In a dynamic system, the input in an interval of time $[0, t]$ and the initial state $x(0)$ uniquely determine the state $x(t)$ and the output $y(t)$ at the final time t. In other words, the future evolution of the system is always guaranteed and uniquely determined. In the case of a linear system, this is clear from the Lagrange formulas (B.22) and (B.23) holding for $t \geq 0$. In some systems, the existence of the evolution is guaranteed and uniquely determined also in the past. Such systems are called *reversible*. For a linear system, *Theorem 3* holds.

THEOREM 3 *(reversibility condition)*

Continuous-time linear systems are reversible, while discrete-time systems are such if and only if their matrix A is nonsingular.

The proof of *Theorem 3* follows from the fact that in continuous-time systems the transition matrix $\Phi(t) = e^{At}$ is invertible [its inverse is, in fact, $\Phi^{-1}(t) = e^{-At}$]. On the contrary, in the case of discrete-time systems $\Phi(t) = A^t$, so that $\Phi(t)$ is invertible if and only if A^t and, therefore, A is invertible.

Theorem 3 let us envisage a strong analogy between reversible continuous and discrete -time systems. On the other hand, it is clear that discrete-time systems need more attention. The peculiarity of irreversibility is often not emphasized as it should be, mainly because the discrete-time systems that are more frequently studied are the *sampled-data systems* that, as it will be shown in Section B.9, are reversible. However, there are important classes of discrete-time systems that are irreversible, such as the *finite memory systems*. Such systems have the property that the initial state influences the systems evolution only for a finite period of time. Since

$$x(t) = \Phi(t)x(0) + \Psi(t)u_{[0,t)}(\cdot)$$

the free motion is zero from a certain time, for any initial state $x(0)$. This implies that $\det \Phi(t) = \det(A^t) = (\det A)^t = 0$, that is, the system is irreversible.

B.9 SAMPLED-DATA SYSTEMS

The input of many continuous-time systems is often changed at precise instants of time, and then kept constant for an interval of time. This may occur in a production system, in which the production rate is fixed each week, in the reactions induced

by a drug treatment delivered by perfusion each day, in the exploitation of water supplies for the production of electric power in which the power of an hydraulic turbine is scheduled to vary each hour, and in many other systems characterized by the presence of a supervisor who, for different reasons, considers it not appropriate to control the system continuously. Once a decision is taken at each instant, the input of the system (production rate, drug delivery rate, turbine power) is kept constant for a certain interval of time, to the end of which a new decision is taken. Unlike inputs, the state variables characterizing the system (stocks, concentrations, water supplies) vary (sometimes quite heavily) during such an interval of time. An analogous situation can often be found in industrial automation, where computers are used for controlling various processes: During a certain interval of time, the computer processes the information received and determines the value $\tilde{u}$ of the input to be applied to the system during the subsequent interval of time. As shown in *Fig. B.7*, an interface is needed between the computer and the system, called *hold circuit*, able to transform the digital input of the computer into a constant analog signal (input of the system).

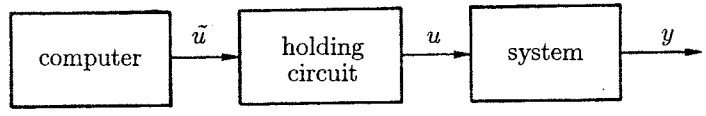

Figure B.7 Sampled-data system.

Also, for state (x) and output (y) variables, the assumption considered in order to define sampled-data systems is consistent with the modern measurement techniques that repeatedly "read" the values $\tilde{x}$ and $\tilde{y}$ at specific times, called *sampling times*. The interval of times occurring between successive sampling times is called a *sampling interval*.

The simplest sampling scheme, depicted in *Fig. B.8* is characterized by having the same sampling time for all variables (state and output) and sampling interval T constant and equal to the interval in which the input is kept constant.

More complex sampling schemes are obtained when T is not constant over time (random sampling and adaptive sampling), when the sampling interval is not the same for all variables (multirate sampling), when, though T is the same for all variables, the sampling instants are shifted over time (asynchronous sampling); or when the holding circuit, instead of keeping constant the input to the system, allows it to vary following a fixed law (*e.g.*, linearly). If we consider the simplest case, we are then dealing with a continuous-time system

$$\dot{x}(t) = Ax(t) + bu(t)$$

$$y(t) = c^T x(t) + du(t)$$

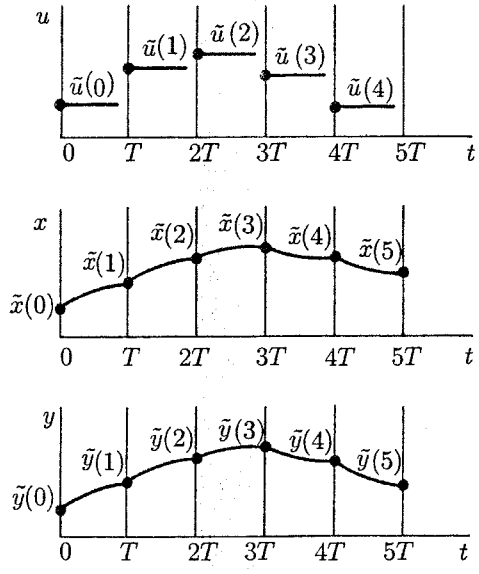

Figure B.8 Input, state, and output of the simplest sampled-data system.

with piecewise constant input. After having numbered the sampling instants with the index $k = 0, 1, 2$, and so on, we can then define input $\tilde{u}$, state $\tilde{x}$, and output $\tilde{y}$ variables of the sampled-data system, in the following way (see *Fig. B.8*):

$$\tilde{u}(k) = u(t) \qquad kT \leq t < (k+1)T$$
$$\tilde{x}(k) = x(kT)$$
$$\tilde{y}(k) = y(kT)$$

By applying the Lagrange formula (B.22) to the continuous-time system with initial time kT and final time $(k+1)T$, and by taking into account that between those two instants the input is constant and equal to $\tilde{u}(k)$, one gets

$$x((k+1)T) = e^{AT}x(kT) + \int_0^T e^{A(T-\xi)}d\xi b\tilde{u}(k)$$

By substituting in this equation $x(kT)$ and $x((k+1)T)$ with $\tilde{x}(k)$ and $\tilde{x}(k+1)$ and taking into account that

$$\int_0^T e^{A(T-\xi)}d\xi = \int_0^T e^{A\xi}d\xi$$

one obtains

$$\tilde{x}(k+1) = e^{AT}\tilde{x}(k) + \left(\int_0^T e^{A\xi}d\xi\right) b\tilde{u}(k)$$

which is the state equation of the sampled-data system, interpreted as a discrete-time system. Since, clearly, the output transformation holds also for sampled variables, we can conclude that a sampled-data system is a discrete-time system

$$\tilde{x}(k+1) = \tilde{A}\tilde{x}(k) + \tilde{b}\tilde{u}(k)$$
$$\tilde{y}(k) = \tilde{c}^T\tilde{x}(k) + \tilde{d}\tilde{u}(k)$$

with

$$\tilde{A} = e^{At} \quad \tilde{b} = \left(\int_0^T e^{A\xi}d\xi\right) b \quad \tilde{c}^T = c^T \quad \tilde{d} = d$$

EXAMPLE 5 (mechanical system)
Let us consider again the mechanical system described in *Example 1*, composed of a point mass m moving along a straight line and to which a force $u(t)$ is applied. If $x_1 = y$ and x_2 are the position and velocity of the mass and there is no friction, the system is described by the triple

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad b = \begin{pmatrix} 0 \\ 1/m \end{pmatrix}$$
$$c^T = \begin{pmatrix} 1 & 0 \end{pmatrix}$$

Since $A^2 = 0$ (and, therefore, $A^i = 0, i \geq 2$) the transition matrix results to be

$$e^{AT} = I + At = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

so that the sampled-data system is described by the triple

$$\tilde{A} = e^{AT} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \quad \tilde{b} = \left(\int_0^T e^{A\xi}d\xi\right) b = \begin{pmatrix} T & T^2/2 \\ 0 & T \end{pmatrix} \begin{pmatrix} 0 \\ 1/m \end{pmatrix} = \begin{pmatrix} T^2/2m \\ T/m \end{pmatrix}$$
$$\tilde{c}^T = c^T = \begin{pmatrix} 1 & 0 \end{pmatrix}$$

♣

It is important to note that an entire family of sampled-data systems $\tilde{\Sigma}$ is associated to each continuous-time system Σ, since, even if the notation adopted does not show this clearly, the matrix $\tilde{A}$ and the vector $\tilde{b}$ depend on the sampling interval T. It is therefore interesting to know whether a property holding for the continuous-time system Σ is maintained for the family $\tilde{\Sigma}$ or whether a property that does not hold for Σ can be "gained" by sampling the system using an appropriate sampling interval. As stated in Section B.8, we can find reversibility among the properties that are maintained under sampling. All sampled-data systems $\tilde{\Sigma}$ are in fact reversible (just as continuous-time systems), since the matrix $\tilde{A} = e^{AT}$ is nonsingular for any sampling interval (actually, e^{AT} admits as inverse the matrix e^{-AT}).

B.10 INTERNAL STABILITY: DEFINITIONS

Stability is certainly the most studied property of a dynamical system. As we will see, it allows us to characterize the asymptotic behavior ($t \to \infty$) of the system, which is a very important feature in applications.

DEFINITION 2 *(asymptotic stability, simple stability, and instability)*

A linear system is *asymptotically* stable if and only if its free motion tends to zero as ($t \to \infty$) for any initial state. If, instead, the free motion is bounded but does not tend to zero for some initial state, the system is said to be *simply stable*. Finally, if the free motion is unbounded for some initial state, the system is said to be *unstable*.

On the basis of this definition, it is immediate to see that the two systems discussed in the first two Examples (Newton's law and Fibonacci's rabbits) are both unstable. The first, however, is a *weakly unstable* system since the free motion, though unbounded, grows with time following a polynomial law, which is linear in this case. On the other hand, the second is *strongly unstable*, since the free motion grows exponentially.

From *Definition 2* it follows that a system is asymptotically stable if and only if

$$\lim_{t\to\infty} \Phi(t) = 0$$

that is, if and only if all the entries of the transition matrix tend to zero as $t \to \infty$. Finite memory systems are then asymptotically stable.

The most important property (easy to prove) of asymptotically stable systems, sometimes used as an alternative definition of asymptotic stability, is the following:

THEOREM 4 *(asymptotic stability and convergence to equilibrium)*

A system is asymptotically stable if and only if for any input $\bar{u}$ there exists a single equilibrium state $\bar{x}$ and $x(t)$ tends to $\bar{x}$ for $t \to \infty$ for any $x(0)$ when $u(t) = \bar{u}$.

It is worth noting that in the unstable system corresponding to Newton's law, we have for $\bar{u} = 0$ an infinite number of equilibria $\bar{x}^T = |\bar{x}_1 \quad 0|^T$, while in the system describing Fibonacci's rabbits, the equilibrium state $\bar{x}$ is unique but $x(t)$ does not tend to $\bar{x}$ as $t \to \infty$.

B.11 EIGENVALUES AND STABILITY

Stability of linear systems can be fully understood by making reference to the *Jordan canonical form* A_J of the matrix A. More precisely, by means of an appropriate change of the state variables

$$z = T_J x$$

it is possible to transform the given system (A, b, c^T, d) into an equivalent one $(T_J, AT_J^{-1}, T_J b, c^T T_J^{-1}, d)$ in which the matrix T_J, AT_J^{-1} is the Jordan matrix A_J. The free motion of the system is then described by the equations

$$\dot{z}(t) = A_J z(t)$$

in the case of a continuous-time system and by the equations

$$z(t+1) = A_J z(t)$$

in the case of a discrete-time system. The advantage of such a transformation is that, due to the structure of the matrix A_J (see *Appendix A*), the system is decomposed in a number of noninteracting subsystems, one for each Jordan block

$$J_i^h = \begin{pmatrix} \lambda_i & 1 & 0 & \dots & 0 \\ 0 & \lambda_i & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & \lambda_i \end{pmatrix}$$

where λ_i is the ith distinct eigenvalue of A and J_i^h has dimension $n_i^h \times n_i^h$ with n_i^h smaller than or equal to the multiplicity of the eigenvalue λ_i in the minimal polynomial $\Psi_A(\lambda)$. In the case of continuous-time systems, the transition matrix of each of these subsystems is

$$e^{J_i^h t} = e^{\lambda_i t} \begin{pmatrix} 1 & t & \frac{t^2}{2} & \frac{t^3}{3!} & \dots \\ 0 & 1 & t & \frac{t^2}{2} & \dots \\ 0 & 0 & 1 & t & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$

and contains terms of the form $t^k e^{\lambda_i t}$ with k smaller than the multiplicity of the eigenvalue λ_i in the minimal polynomial $\Psi_A(\lambda)$. Since $t^k e^{\lambda_i t}$ tends to zero, as $t \to \infty$, if and only if the real part of λ_i is negative, it follows that a continuous-time system is asymptotically stable if and only if all the eigenvalues of A have a negative real part. In contrast, if eigenvalues with a positive real part exist, some of the terms of the transition matrix are unbounded and grow exponentially ($k = 0$) or more than exponentially ($k \geq 1$). In both cases, the system is strongly unstable. In the remaining cases, that is, when there are zero or purely imaginary eigenvalues λ_{i^*} but there are no eigenvalues with a positive real part, one has simple stability if the term $t^k e^{\lambda_{i^*} t}$ is bounded ($k = 0$) and weak instability in the opposite case ($k \geq 1$). By taking into account that k is necessarily zero only in the case in which the eigenvalue λ_{i^*} with zero real part is a simple root of the minimal polynomial $\Psi_A(\lambda)$ and noting that in the case of discrete-time systems the exponential term $e^{\lambda_i t}$ is replaced by the power λ_i^t, which tends to zero if and only if $|\lambda_i| < 1$, we can summarize the previous discussion with *Theorem 5*.

THEOREM 5 *(stability conditions)*

A continuous-time [discrete-time] linear system (A, b, c^T, d) is	
a. Asymptotically stable if and only if	$\mathrm{Re}(\lambda_i) < 0$ $[\lvert\lambda_i\rvert < 1]$ $\forall i$
b. Simply stable if and only if	$\mathrm{Re}(\lambda_i) \leq 0$ $[\lvert\lambda_i\rvert < 1]$ $\forall i$
	$\exists i^* : \mathrm{Re}(\lambda_{i^*}) = 0$ $[\lvert\lambda_{i^*}\rvert = 1]$
	all λ_{i^*} are simple roots of Ψ_A
c. Weakly unstable if and only if	$\mathrm{Re}\ (\lambda_i) \leq 0$ $[\lvert\lambda_i\rvert \leq 1]$ $\forall i$
	$\exists i^* : \mathrm{Re}(\lambda_{i^*}) = 0$ $[\lvert\lambda_{i^*}\rvert = 1]$
	at least one λ_{i^*} is not a simple root
	of Ψ_A
d. Strongly unstable if and only if	$\exists i : \mathrm{Re}(\lambda_i) > 0$ $[\lvert\lambda_i\rvert > 1]$

The system representing Newton's law (*Example 1*), which has the matrix A in Jordan form, has a zero eigenvalue that is a double root of the minimal polynomial. As previously stated, it is then weakly unstable. The Fibonacci's system (*Example 2*) has, instead, two eigenvalue $\lambda_{1,2} = (1 \pm \sqrt{5})/2$ so that one of them is > 1 and the system is therefore strongly unstable.

The n eigenvalues of matrix A of a continuous-time linear system can be divided into three classes, depending on the sign of their real part: n^- eigenvalues, called *stable*, have a negative real part, n^0 have a zero real part and are called *critical*, and n^+ have a positive real part and are called unstable. Obviously, $n = n^- + n^0 + n^+$. The corresponding eigenvectors define three disjoint invariant subspaces X^-, X^0, and X^+ with dimension n^-, n^0, and n^+ respectively. Initial states in the subspace X^- give rise to free motions that tend to zero, while initial states in the subspace X^+ give rise to free motions that tend to infinity at least at an exponential rate. For this reason, these two subspaces are called, respectively, *stable manifold* and *unstable manifold*. The subspace X^0 is called *center manifold*: The free motions corresponding to initial states in X^0 remain in X^0, do not tend to zero and eventually tend to infinity at a polynomial rate. Systems without center manifold (*i.e.*, without critical eigenvalues) are called *hyperbolic* and are divided into *attractors* ($X^- = \mathbb{R}^n$), *saddles* ($X^- \oplus X^+ = \mathbb{R}^n$), and *repellors* ($X^+ = \mathbb{R}^n$). Systems possessing a center manifold are called *nonhyperbolic*. *Figure B.9* shows the trajectories corresponding to the free motion of eight different second-order continuous-time systems. Each figure also shows the two eigenvalues of the system. The first five systems (stable focus, stable node, unstable focus, unstable node, saddle) are hyperbolic and the last three are non-hyperbolic. The last system (pure imaginary eigenvalues) is called *center* and this explains the choice of the term "center manifold".

The advantage of the decomposition of the state space $\mathbb{R}^n$ into the direct sum of three subspaces X^-, X^0, and X^+ is particularly clear when visualizing the

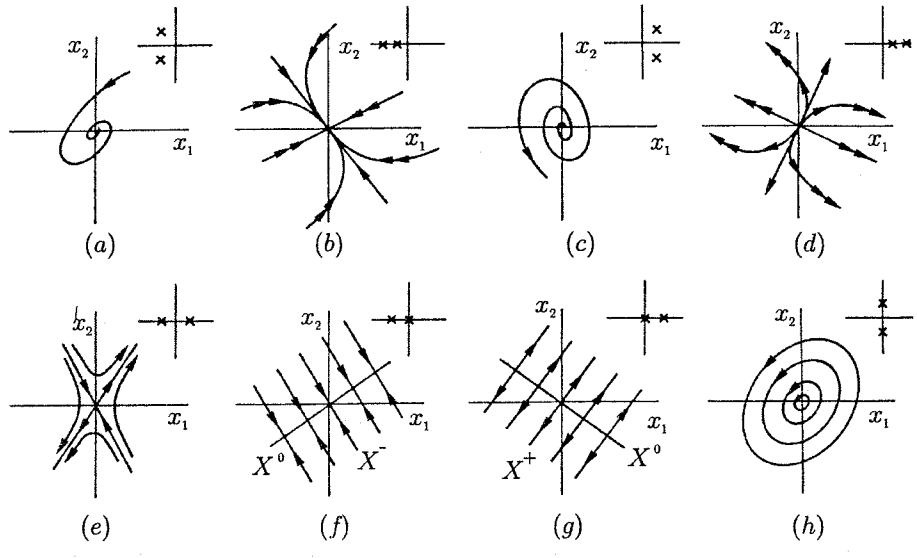

Figure B.9 Trajectories corresponding to the free motion of second-order continuous-time systems: (*a*) (stable focus) and (*b*) (stable node) are attractors; (*c*) (unstable focus) and (*d*) (unstable node) are repellors; (*e*) is a saddle; (*f*), (*g*), and (*h*) (center) are systems with center manifold X^0. The straight trajectories correspond to eigenvectors associated to real eigenvalues. The double arrow indicates parts of the trajectories where the state of the system moves more rapidly.

geometry of the free motion, in particular for third-order systems, as the two saddles shown in *Fig. B.10*.

Obviously, what has been said for continuous-time systems also holds for discrete-time systems when separately considering the cases in which we have stable eigenvalues ($|\lambda_i| < 1$), critical ($|\lambda_i| = 1$), and unstable ($|\lambda_i| > 1$).

B.12 TESTS OF ASYMPTOTIC STABILITY

In Section B.11, we showed that knowing the eigenvalues of matrix A of a linear system one can establish whether such a system is asymptotically stable (or not). Unfortunately, the computation of the eigenvalues of a matrix can be very onerous if the matrix is large, as it is often the case in real applications. For this reason, it is convenient to use some tests or methods that, avoiding the computation of the eigenvalues, allow us to infer the asymptotic stability or instability of a system.

One of the most popular of such tests, which is a sufficient condition for instability, is the *trace criterion*, which states that a continuous-time [discrete-time]

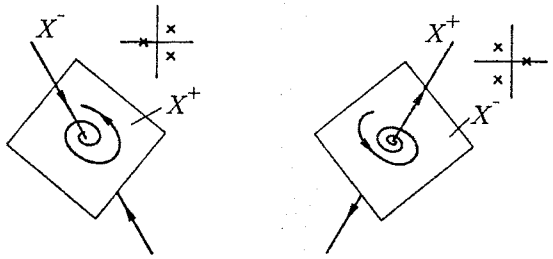

Figure B.10 Two third-order saddles: (*a*) $n^- = 1, n^+ = 2$; (*b*) $n^- = 2, n^+ = 1$.

system of dimension n, has the trace of its matrix A positive [$> n$ in modulus], then the system is unstable. For proving this theorem it suffices to remember that the trace of a matrix equals the sum of its eigenvalues.

A condition that requires a much greater computational burden (but is still more effective than the computation of the eigenvalues) is known as the *Hurwitz criterion*. Such criterion (whose proof is not reported here) is a necessary and sufficient condition for the n roots of a polynomial equation with real coefficients

$$\alpha_0\lambda^n + \alpha_1\lambda^{n-1} + \cdots + \alpha_n = 0$$

to have a negative real part. When applied to the characteristic equation $\Delta_A(\lambda) = 0$, which can be determined using the Souriau method cited in section B.3, the criterion allows one to establish whether a continuous-time system is asymptotically stable or not.

THEOREM 6 *(Hurwitz criterion)*

Let

$$\Delta_A(\lambda) = \lambda^n + \alpha_1\lambda^{n-1} + \cdots + \alpha_n$$

be the characteristic polynomial of a continuous-time linear system $\dot{x}(t) = Ax(t)$. Consider the following $n \times n$ matrix (called the *Hurwitz matrix*)

$$H = \begin{pmatrix} \alpha_1 & 1 & 0 & 0 & \dots \\ \alpha_3 & \alpha_2 & \alpha_1 & 1 & \dots \\ \alpha_5 & \alpha_4 & \alpha_3 & \alpha_2 & \dots \\ \alpha_7 & \alpha_6 & \alpha_5 & \alpha_4 & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$

in which $\alpha_i = 0$ for $i > n$. Then, a necessary and sufficient condition for asymptotic stability of the system is that all the principal minors of the Hurwitz matrix be positive. That is, setting

$$D_1 = \alpha_1 \quad D_2 = \det\begin{pmatrix} \alpha_1 & 1 \\ \alpha_3 & \alpha_2 \end{pmatrix} D_3 = \det\begin{pmatrix} \alpha_1 & 1 & 0 \\ \alpha_3 & \alpha_2 & \alpha_1 \\ \alpha_5 & \alpha_4 & \alpha_3 \end{pmatrix} \dots D_n = \det H$$

a necessary and sufficient condition for asymptotic stability of the system is that $D_i > 0, i = 1, \dots, n$.

Another important criterion for asymptotic stability, equivalent to the Hurwitz criterion, is the following:

THEOREM 7 *(Routh criterion)*

Let

$$\Delta_A(\lambda) = \lambda^n + \alpha_1\lambda^{n-1} + \cdots + \alpha_n$$

be the characteristic polynomial of a continuous-time linear system $\dot{x}(t) = Ax(t)$. Consider the $(n+1) \times (n+1)$ matrix (called the *Routh matrix*)

$$R = \begin{pmatrix} 1 & \alpha_2 & \alpha_4 & \dots \\ \alpha_1 & \alpha_3 & \alpha_5 & \dots \\ r_{21} & r_{22} & r_{23} & \dots \\ r_{31} & r_{32} & r_{33} & \dots \\ \vdots & \vdots & \vdots & \\ r_{n1} & r_{n2} & r_{n3} & \dots \end{pmatrix}$$

in which

$$\alpha_{n+i} = 0 \quad \text{for} \quad i = 1, 2, \ldots$$

$$r_{2j} = \alpha_{2j} - \frac{\alpha_{2j+1}}{\alpha_1}$$

$$r_{ij} = r_{i-2,j+1} - \frac{r_{i-2,1} r_{i-1,j+1}}{r_{i-1,1}} \quad \text{for} \quad i = 3, 4, \ldots$$

(note that r_{ij} can be computed only if $r_{i-1,1} \neq 0$). Then, a necessary and sufficient condition for asymptotic stability of the system is that all the entries of the first column of R be positive.

Clearly, there exist criteria for asymptotic stability for discrete-time systems analogous to the Hurwitz and Routh ones. It is worth noting, however, that by means of the change of variables

$$z = \frac{s+1}{s-1}$$

one can transform the problem of checking whether the zeros of a polynomial in z are within the unitary circle to that of checking whether the zeros of a polynomial in s have a negative real part. In other words, if

$$\Delta_A(z) = \det(zI - A) = z^n + \alpha_1 z^{n-1} + \cdots + \alpha_n$$

is the characteristic polynomial of a discrete-time system $x(t+1) = Ax(t)$, one can write the characteristic equation in the form

$$\left(\frac{s+1}{s-1}\right)^n + \alpha_1 \left(\frac{s+1}{s-1}\right)^{n-1} + \cdots + \alpha_n = 0$$

which yields

$$s^n + \alpha_1' s^{n-1} + \cdots + \alpha_n' = 0$$

By applying the Hurwtitz or Routh criterion to this equation, one can determine whether the discrete-time system is asymptotically stable or not.
However, we report next one of the most popular criteria for asymptotic stability of discrete-time systems.

THEOREM 8 *(Jury criterion)*

Let

$$\Delta_A(\lambda) = \lambda^n + \alpha_1 \lambda^{n-1} + \cdots + \alpha_n$$

be the characteristic polynomial of a discrete-time system $x(t+1) = Ax(t)$. Consider the following table of dimension $2n \times (n+1)$ composed of n pairs of rows:

$$\begin{pmatrix}
p_{11} & p_{12} & \cdots & p_{1n} & p_{1,n+1} \\
q_{11} & q_{12} & \cdots & q_{1n} & q_{1,n+1} \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
p_{21} & p_{22} & \cdots & p_{2n} & \\
q_{21} & q_{22} & \cdots & q_{2n} & \\
\cdots & \cdots & \cdots & \cdots & \\
\cdots & \cdots & & & \\
\cdots & \cdots & & & \\
\cdots & \cdots & \cdots & \cdots & \\
p_{n1} & p_{n2} & & & \\
q_{n1} & q_{n2} & & &
\end{pmatrix}$$

where

a. The elements of the first row are $\alpha_n, \alpha_{n-1}, \ldots, \alpha_1, 1$, that is, the coefficients of the characteristic polynomial are in reversed order.

b. Each even row coincides with the preceding row in reversed order.

c. The elements p_{ji} can be calculated as follows:

$$p_{j+1,i} = \det \begin{pmatrix} p_{j1} & q_{ji} \\ q_{j1} & p_{ji} \end{pmatrix} \quad j = 1, 2, \ldots, n-1$$

Then, a necessary and sufficient condition for asymptotic stability of the system is that the following conditions hold:

$$\Delta_A(1) > 0 \qquad (-1)^n \Delta_A(-1) > 0$$

$$p_{21} < 0 \qquad p_{j1} > 0, j = 3, 4, \ldots, n$$

In the particular case of second-order systems (matrix A of dimension 2×2), special conditions hold that often allow us to check asymptotic stability by simple inspection of the matrix A.

THEOREM 9 *(trace and determinant criterion)*

A second-order continuous-time system $\dot{x}(t) = Ax(t)$ is asymptotically stable if and only if

$$\text{tr}A < 0 \quad \det A > 0$$

Analogously, a second-order discrete-time system $x(t+1) = Ax(t)$ is asymptotically stable if and only if

$$|\text{tr}A| < 1 + \det A \quad \det A < 1$$

The reader may check the efficacy of this criterion by applying it to the systems described in the *Examples 1* and *2*.

B.13 ENERGY AND STABILITY

It is known that in some systems (*e.g.*, electrical and mechanical ones) it is possible to define a function of the state $V(x(t))$, called energy, which has the property to be quadratic and nonnegative and decreases with time (and tends to zero) whenever the system is asymptotically stable and evolves freely. This property, studied by the Russian mathematician *Alexander Liapunov* (more than a century ago), allows us to analyze the stability of any linear system in a very synthetic and elegant way.

In order to introduce this topic, we need to define positive definite matrices. First of all, we say that a function $V(x(t))$ with $x \in \mathbb{R}^n$ is *quadratic* if

$$V(x) = x^T P x$$

where P is an $n \times n$ matrix. This means that $V(x)$ is a weighted sum of all the products $x_i x_j$. Since the weight of the term $x_i x_j$ is $(p_{ij} + p_{ji})$, there is no loss of generality in assuming that the matrix P is symmetric. Moreover, a matrix P is said to be *positive definite* if the associated quadratic form $x^T P x$ is positive for all the vectors $x \neq 0$. To know whether a given matrix P is positive definite, one can apply the following criterion:

THEOREM 10 *(Sylvester criterion)*

A symmetric matrix P is positive definite if and only if all the principal minors $D_1, D_2, \ldots, D_n$ are positive, that is,

$$D_1 = p_{11} > 0 \quad D_2 = \det \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} > 0 \quad \ldots \quad D_n = \det P > 0$$

It is clear from the previous discussion that any positive definite matrix P induces a metric in the space $\mathbb{R}^n$: in other words, $V(x) = x^T P x$ can be interpreted as the distance of point x from the origin. In this metric, the points x at the same distance from the origin lie on the manifold $x^T P x =$ constant, which in $\mathbb{R}^2$ is an ellipse.

Suppose now that a positive definite matrix P is associated to an autonomous continuous-time linear system $\dot{x}(t) = Ax(t)$. The "distance" of the point $x(t)$ from the origin is $V(x(t)) = x^T(t)Px(t)$ and such a distance varies in time since x depends on t. More precisely

$$\begin{aligned} \dot{V} &= \dot{x}Px + x^T P\dot{x} = x^T A^T Px + x^T PAx \\ &= x^T(A^T P + PA)x \end{aligned}$$

so that the distance of $x(t)$ from the origin decreases continuously with time ($\dot{V} < 0$) if the matrix $-(A^T P + PA)$ is positive definite, that is, if the so-called *Liapunov equation*

$$A^T P + PA = -Q \tag{B.26}$$

is satisfied with Q positive definite. Obviously, if this equation holds, then the system is asymptotically stable since the free motion of the system asymptotically tends to zero, since $\dot{V} < 0$ for $\neq 0$.

In the case of discrete-time systems $x(t+1) = Ax(t)$, one has to determine the quantity ΔV defined as

$$\begin{aligned} \Delta V &= V(x(t+1)) - V(x(t)) = (Ax(t))^T PAx(t) - x^T(t)Px(t) \\ &= x^T(t)(A^T PA - P)x(t) \end{aligned}$$

so that the discrete-time Liapunov equation is the following:

$$A^T PA - P = -Q \tag{B.27}$$

In conclusion, if a continuous-time [discrete-time] system satisfies the Liapunov equation (B.26) [(B.27)] with P and Q positive definite matrices, the system is asymptotically stable and the function $V(x) = x^T Px$, called the *Liapunov function*, has the properties of any energy function, that is, it is positive and decreasing with time for $x \neq 0$. Clearly, this does not imply that, given an asymptotically stable autonomous system in which $x(t)$ tends toward the origin as $t \to \infty$, any quadratic function $V(x) = x^T Px$ with P positive definite systematically decreases with time (even though V tends to zero as $t \to \infty$). This fact is illustrated in *Fig. B.11* for a second-order continuous-time system (stable focus). In such a system, $V(x) = x^T Px$ is positive definite, but $\dot{V} = x^T(A^T P + PA)x$ is positive at some points and negative at others, so that V tends to zero as $t \to \infty$ but not monotonically.

The above discussion is further specified in the following theorem.

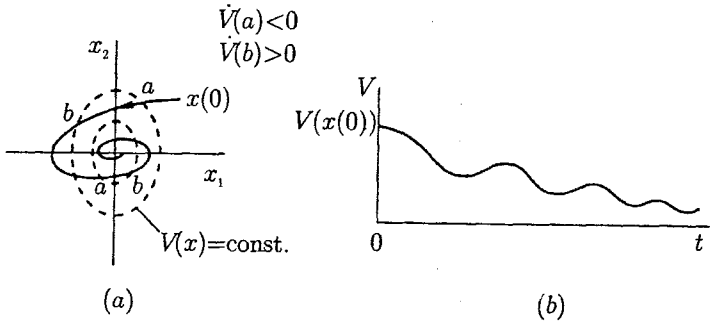

Figure B.11 Second-order asymptotically stable autonomous system (stable focus): (*a*) trajectory (−) and contour lines (...) of the function $V(x) = x^T Px$; (*b*) time evolution of V.

THEOREM 11 *(Liapunov theorem)*

A continuous-time [discrete-time] linear system $(A, -, -)$ is asymptotically stable if and only if there exists a quadratic function $V(x) = x^T Px$ with P positive definite, which is strictly decreasing in time (*i.e.*, such that $\dot{V} < 0$ [$\Delta V < 0$]) along the free motion of the system with $x \neq 0$. Moreover, such a function exists if and only if the Liapunov equation (B.26) [(B.27)] admits a solution (P, Q) with P and Q positive definite. Finally, if such a pair (P, Q) exists, then one can find an infinite number of them, one for each positive definite matrix Q.

The last statement of the Liapunov theorem allows one to derive a practical criterion for testing if a linear system is asymptotically stable. If we chose any symmetric and positive definite matrix Q in the Liapunov equation [*e.g.*, $Q = I$ (identity matrix)], one can solve the equation in the unknown P [note that if P is symmetric there are $n(n+1)/2$ linear equations in the same number of unknowns]; if the solution P exists and is positive definite (this can be checked using Sylvester criterion) the system is asymptotically stable.

It is worth noting that the Liapunov theorem does not require any particular structure for the matrix P. Often, however, the matrix P is diagonal, that is, the Liapunov function $V(x)$ does not contain mixed terms $x_i x_j$ with $i \neq j$ (as an example, consider the energy function of an electrical network). If this is the case, that is, when the Liapunov equation admits a solution (P, Q) with P and Q positive definite and P diagonal, the system, besides being asymptotically stable, remains such for any structural perturbation (*Arrow–McManus* theorem). This means that, if there exists a diagonal matrix P with $p_{ii} > 0$, $i = 1, \ldots, n$ such that the matrix $-(A^T P + PA)$ is positive definite, we can assert that the system $\dot{x} = Ax$ is asymptotically stable as well as all other systems of the kind $\dot{x} = DAx$ with D diagonal and $d_{ii} > 0$. In other words, the system $\dot{x} = Ax$ is asymptotically stable and remains such under the effect of any perturbation that results in the multiplication of any one of its state equations by a positive arbitrary constant.

B.14 DOMINANT EIGENVALUE AND EIGENVECTOR

The free motion $x(t) = \Phi(t)x(0)$ of a linear system is completely identified by the transition matrix

$$\Phi(t) = \begin{cases} e^{At} & \text{for continuous-time systems} \\ A^t & \text{for discrete-time systems} \end{cases}$$

Recalling what has been previously stated about eigenvectors, eigenvalues, and the Jordan canonical form of a matrix A, we can conclude that in the case of real and distinct eigenvalues the free motion is

$$x(t) = \begin{cases} \sum_{i=1}^{n} c_i x^{(i)} e^{\lambda_i t} & \text{for continuous-time systems} \\ \sum_{i=1}^{n} c_i x^{(i)} \lambda_i^t & \text{for discrete-time systems} \end{cases}$$

where $x^{(i)}$, $i = 1, \ldots, n$ are the eigenvectors of A [satisfying the equation $Ax^{(i)} = \lambda_i x^{(i)}$] and c_i are the components of the initial state $x(0)$ in the basis composed of the n eigenvectors $x^{(i)}$ [*i.e.*, $\sum_{i=1}^{n} c_i x^{(i)} = x(0)$]. Clearly, for particular initial conditions, some c_i may be zero, but for generic initial conditions, all the c_i's are different from zero and the free motion is the weighted sum of n exponential terms. As time passes, one of these exponential terms necessarily dominates the other since, for large values of t, $e^{\lambda_i t} >> e^{\lambda_j t}$ [$|\lambda_i^t| >> |\lambda_j^t|$] if $\lambda_i > \lambda_j$ [$|\lambda_i| > |\lambda_j|$] (notice that this occurs also in the case of simple stability or instability). The eigenvalue and eigenvector associated with this exponential term are called *dominant*. The dominant eigenvalue λ_{dom} is clearly the highest eigenvalue in continuous-time systems and the eigenvalue with maximal modulus in discrete-time systems. In *Fig. B.9*, five cases [(*b*), (*d*), (*e*), (*f*), (*g*)] deal with continuous-time systems with real eigenvalues. They show that all generic trajectories tend to align with the dominant eigenvector as time passes. This property is very important and must be accounted for when characterizing the asymptotic behavior of linear systems. Therefore, the free motion can be approximated in the long run with a single exponential term

$$x(t) \cong \begin{cases} c x_{\text{dom}} e^{\lambda_{\text{dom}} t} & \text{for continuous-time systems} \\ c x_{\text{dom}} \lambda_{\text{dom}}^t & \text{for discrete-time systems} \end{cases}$$

and if the system is asymptotically stable, this exponential term tends to zero. In these cases, the exponential term is often written in the form $e^{-t/T_{\text{dom}}}$, where $T_{\text{dom}}(> 0)$ is the so-called *dominant time constant*, linked to the dominant eigenvalue by the relationship

$$T_{\text{dom}} = \begin{cases} -\frac{1}{\lambda_{\text{dom}}} & \text{for continuous-time systems} \\ -\frac{1}{\log|\lambda_{\text{dom}}|} & \text{for continuous-time systems} \end{cases}$$

Moreover, in applications it is rather common to say that the exponential term is "practically" over after a period of time equal to five times the time constant T_{dom}. If the eigenvalues of A, though distinct, are not all real, the free motion is also characterized by terms of the form $e^{\lambda t}$ and λ^t with λ complex. In the continuous-time case, these terms correspond to an exponentially dumped sinusoidal function $e^{at}\sin(bt+\varphi)$, where a is the real part of the eigenvalue and b is the imaginary part. It follows that the dominant term of the free motion is the one associated with the (real or complex) eigenvalue with the maximal real part. In conclusion, in the long run, the free motion can be approximated by an exponential if the dominant eigenvalue is real, or by a sinusoid with an exponentially varying amplitude if the dominant eigenvalue is complex. Obviously, in the case of asymptotically stable continuous-time systems the dominant time constant is given by

$$T_{\text{dom}} = -\frac{1}{\text{Re}(\lambda_{\text{dom}})}$$

Finally, if the eigenvalues of A are not all distinct, the free motion of a continuous-time [discrete-time] system may contain terms of the form $t^k e^{\lambda t}$ $[t^k\lambda^t]$; this fact, however, does not change the previous conclusions, since the free motion is still dominated by the term associated with the (real or complex) eigenvalue with maximal real part [maximal modulus].

B.15 REACHABILITY AND CONTROL LAW

The motion of a linear system is given by

$$x(t) = \Phi(t)x(0) + \Psi(t)u_{[0,t)}(\cdot)$$

that is, the sum of the free and forced motion. The forced motion

$$\Psi(t)u_{[0,t)}(\cdot) = \begin{cases} \int_0^t e^{A(t-\xi)}bu(\xi)d\xi & \text{for continuous-time systems} \\ \sum_{i=0}^{t-1} A^{t-i-1}bu(i) & \text{for discrete-time systems} \end{cases}$$

describes the set $X_r(t)$ of all states reachable from the origin of the state space at time t. Obviously, such a set is a subspace for which the following property holds

$$X_r(t_1) \subset X_r(t_2) \quad t_1 \le t_2$$

Since $X_r(t)$ cannot grow indefinitely, there exists a time t^* such that $X_r(t) = X_r$ for $t > t^*$. Finally, if $X_r = \mathbb{R}^n$, the system is said to be *completely reachable*. *Theorem 12*, known as the *Kalman theorem*, holds.

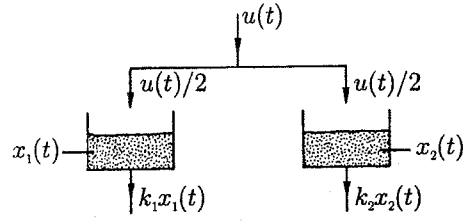

Figure B.12 Two reservoirs fed in parallel.

THEOREM 12 *(complete reachability)*

In a linear system (A, b) of order n, the reachability subspace X is spanned by the n vectors $b, Ab, \ldots, A^{n-1}b$, called reachability vectors. Thus the system is completely reachable if and only if these n vectors are linearly independent. Moreover, each state belonging to X_r is reachable in any time if the system is continuous-time and in at most n transitions if the system is discrete-time.

This theorem is often formulated by making reference to the reachability matrix (also called the *Kalman matrix*)

$$R = (\ b \quad Ab \quad \ldots \quad A^{n-1}b\ )$$

This matrix is $n \times n$ and its image is the reachability subspace, that is,

$$X_r = \mathcal{I}[R]$$

so that the complete reachability of the system is equivalent to the nonsingularity of the matrix R (*i.e.*, to the existence of R^{-1}).

EXAMPLE 6

Consider the system shown in *Fig. B.12* composed of two reservoirs $i = 1, 2$ feed in parallel with a flow $u(t)/2$ and with an output flow-rate proportional (through a coefficient k_i) to the storage $x_i(t)$.

The conservation of mass gives

$$\dot{x}_1 = -k_1x_1 + \frac{u}{2}$$
$$\dot{x}_2 = -k_2x_2 + \frac{u}{2}$$

which are the state equations of a linear system with

$$A = \begin{pmatrix} -k_1 & 0 \\ 0 & -k_2 \end{pmatrix} \qquad b = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$$

Thus, the reachability matrix is

$$R = \begin{pmatrix} 1/2 & -k_1/2 \\ 1/2 & -k_2/2 \end{pmatrix}$$

so that the system is complete reachable if and only if $k_1 \neq k_2$. The reason for this is that in the case of identical reservoirs ($k_1 = k_2$) it is impossible to unbalance the system [$x_1(t) \neq x_2(t)$] since, by assumption, the initial condition is balanced [$x_1(0) = x_2(0) = 0$].

♣

It is worth noting that all systems with $A = A_c$, $b = b_c$, where

$$A_c = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -\alpha_n & -\alpha_{n-1} & -\alpha_{n-2} & \dots & -\alpha_1 \end{pmatrix} \quad b_c = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

that is, all systems in control canonical form (see Section B.3) are completely reachable. In fact, it is immediate to check that the Kalman matrix

$$R_c = (\; b_c \quad A_c b_c \quad \dots \quad A_c^{n-1} b_c \;)$$

is nonsingular for all values of the coefficients α_i, $i = 1, 2, \dots, n$ and that

$$R_c^{-1} = \begin{pmatrix} \alpha_{n-1} & \alpha_{n-2} & \dots & \alpha_1 & 1 \\ \alpha_{n-2} & \alpha_{n-3} & \dots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix}$$

The reason for the interest in this canonical form is justified by *Theorem 13*.

THEOREM 13 *(control canonical form)*

A completely reachable system (A, b) can be put in control canonical form by means of the coordinate transformation $z = R_c R^{-1} x$, where R and R_c are the reachability matrices of (A, b) and (A_c, b_c).

This means that any system (A, b) can be assumed to be in control canonical form provided it is completely reachable. Moreover, the control canonical form (A_c, b_c) can be easily determined by evaluating (*e.g.*, with the Souriau formula) the coefficients α_i.

The importance of complete reachability emerges whenever one tries to modify the dynamics of a given system by linking its input $u(t)$ to the state $x(t)$ by means of a linear feedback rule

$$u(t) = k^T x(t) + v(t)$$

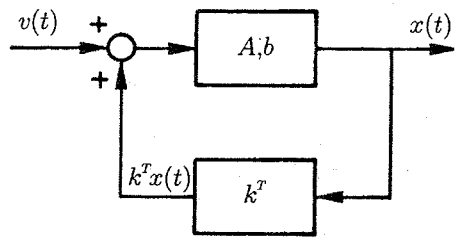

Figure B.13 Controlled system composed of a system (A, b) and its controller k^T.

known as the *control law* (algebraic and linear). The resulting system, called the *controlled system*, has $v(t)$ as input, and is shown in *Fig. B.13* through a block diagram. The feedback block, called *controller*, performs a simple weighted sum $k_1 x_1 + \dots + k_n x_n$ $(= k^T x(t))$ of the state variables.

If (A, b) is a continuous-time system, then the controlled system is described by the state equations

$$\dot{x} = Ax + b(k^T x + v) = (A + bk^T)x + bv$$

In other words, the system (A, b) is transformed, by means of the feedback controller k^T, into the controlled system $(A + bk^T, b)$. Thus, the dynamics of the system are now different since the characteristic polynomial has been modified from $\Delta_A(\lambda)$ into $\Delta_{A+bk^T}(\lambda)$. Obviously, the same holds for discrete-time systems. We can now present the main theorem of this section. It states that complete reachability is a necessary and sufficient condition for the free assignment of the eigenvalues of the controlled system.

THEOREM 14 *(eigenvalue assignment)*

The eigenvalues of the controlled system $(A + bk^T)$ can be arbitrarily assigned by means of a controller k^T, if and only if the system (A, b) is completely reachable. If α_i are the coefficients of the characteristic polynomial of A, and α_i^* those of the characteristic polynomial of $A + bk^T$, then

$$k^T = ((\alpha_n - \alpha_n^*) \; \dots \; (\alpha_1 - \alpha_1^*)) R_c R^{-1}$$

where R and R_c are the reachability matrices of (A, b) and (A_c, b_c).

Theorem 14 implies that the dynamics of a completely reachable system can be modified at will through a linear feedback. The most spectacular consequence of this theorem is the possibility of stabilizing unstable systems.

Since complete reachability is a generic property of linear systems [$\det R \neq 0$ for a generic pair (A, b)], it can be understood that the control scheme shown in *Fig. B.13* is of great practical interest.

B.16 OBSERVABILITY AND STATE RECONSTRUCTION

The observability of a dynamical system refers to the possibility of computing its initial state $x(0)$ once input and output have been recorded during time interval $[0, t)$. Analogously, reconstructability refers to the possibility of computing the final state $x(t)$. Thus, observability implies reconstructability, since once $x(0)$ and $u_{[0,t)}(\cdot)$ are known, it is possible to compute $x(t)$ (Lagrange's formula), while the inverse is possible only if the system is reversible.

In order to study observability, it is worth considering the output free motion of the system

$$c^T\Phi(t)x(0) = \begin{cases} c^T e^{At}x(0) & \text{for continuous-time systems} \\ c^T A^t x(0) & \text{for discrete-time systems} \end{cases}$$

and define the set $X_{\text{no}}(t)$ of the states indistinguishable from the origin as the set of the initial states $x(0)$ for which the output free motion is identically zero in the interval $[0, t)$. Obviously, such a set is a subspace enjoying the property

$$X_{\text{no}}(t_1) \supset X_{\text{no}}(t_2) \quad t_1 \leq t_2$$

But $X_{\text{no}}(t)$ cannot decrease indefinitely, so that a time t^* exists such that $X_{\text{no}}(t) = X_{\text{no}}$ for $t > t^*$. Finally, if $X_{\text{no}} = \{0\}$, all states $x(0) \neq 0$ are distinguishable from the zero state and can be computed from $u_{[0,t)}(\cdot)$ and $y_{[0,t)}(\cdot)$. This is the reason why a system with $X_{\text{no}} = \{0\}$ is called completely observable. It is customary to consider also the subspace $X_{\text{no}}^{\perp}$, that is, the subspace orthogonal to X_{no}. This subspace, called the *observability subspace* and denoted by X_{o}, is useful for explicitly formulating the condition of complete observability.

THEOREM 15 *(complete observability)*

In a linear system of order n, the observability subspace X_{o} is spanned by the n vectors $c, A^T c, \ldots, (A^T)^{n-1}c$ called observability vectors. Thus, the system is completely observable if and only if these vectors are linearly independent. Moreover, in a completely observable system the initial state can be computed if input and output have been recorded for a time period of any length in the case of continuous-time systems and of length n in the case of discrete-time systems.

The above condition is often formulated with reference to the *observability matrix* (also called the Kalman matrix)

$$O = \begin{pmatrix} c^T \\ c^T A \\ \vdots \\ c^T A^{n-1} \end{pmatrix}$$

Then, we have

$$X_{\text{no}} = \mathcal{N}[O] \quad X_{\text{o}} = \mathcal{I}[O^T]$$

and a system is completely observable if its observability matrix is nonsingular (*i.e.*, if the matrix O^{-1} exists). The fact that the initial state of a completely observable system can be computed from input and output records can be verified by explicitly writing the first n output values as a function of the initial state and of the input value, that is,

$$\begin{aligned} y(0) &= c^T x(0) + du(0) \\ y(1) &= c^T Ax(0) + c^T bu(0) + du(1) \\ y(2) &= c^T A^2 x(0) + c^T Abu(0) + c^T bu(1) + du(2) \\ &\vdots \\ y(n-1) &= c^T A^{n-1}x(0) + c^T A^{n-2}bu(0) + \cdots + c^T bu(n-2) + du(n-1) \end{aligned}$$

This is a system of n linear equations with n unknowns [the components of the vector $x(0)$] which admits a unique solution if and only if the observability matrix O is nonsingular.

EXAMPLE 7

Suppose that 10 pairs of adult rabbits have been captured at the beginning of the season in a population described by the Fibonacci model (see *Example* 2) and assume that at the beginning and at the end of the same year 50 and 60 pairs of rabbits (young and adult) where present. This means that

$$u(0) = 10 \quad y(0) = 50 \quad y(1) = 60$$

Since the system is described by the triple

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \quad b = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$
$$c^T = \begin{pmatrix} 1 & 1 \end{pmatrix}$$

we have

$$O = \begin{pmatrix} c^T \\ c^T A \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

so that the system is completely observable and

$$O^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

Thus, the system of 2 equations and 2 unknowns

$$y(0) = c^T x(0)$$
$$y(1) = c^T A x(0) + c^T b u(0)$$

can be solved with respect to $x(0)$

$$x(0) = O^{-1} \begin{pmatrix} y(0) \\ y(1) - c^T b u(0) \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 50 \\ 70 \end{pmatrix} = \begin{pmatrix} 30 \\ 20 \end{pmatrix}$$

We can, therefore, conclude that at the beginning of the year the population was composed of 30 pairs of young rabbits and 20 pairs of adult rabbits.

♣

The comparison of *Theorems 12* and *15* allows one to note a strong analogy between reachability and observability, which can be formalized in the following duality principle:

THEOREM 16 *(duality principle)*

A system Σ is completely reachable [observable] if and only if its dual $\Sigma^* = (A^T, c, b^T, d)$ is completely observable [reachable]. Moreover, the reachability matrix of the system is the transposed of the observability matrix of the dual system.

The duality principle enables us to obtain from *Theorem 13* and from the properties of the control canonical form, the following result:

THEOREM 17 *(reconstruction canonical form)*

A system in reconstruction canonical form

$$A_r = \begin{pmatrix} 0 & 0 & \dots & 0 & -\alpha_n \\ 1 & 0 & \dots & 0 & -\alpha_{n-1} \\ 0 & 1 & \dots & 0 & -\alpha_{n-2} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -\alpha_1 \end{pmatrix}$$

$$c_r^T = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

is completely observable. Conversely, a completely observable system (A, c^T) can be put into reconstruction canonical form by means of a suitable coordinate transformation.

Also, *Theorem 14* on eigenvalues assignment can be dualized. For this, we first introduce the notion of state *reconstructor*, illustrated in *Fig. B.14*.

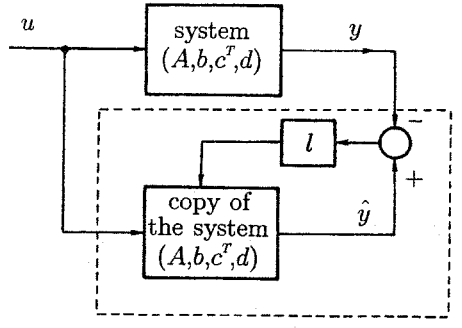

Figure B.14 A system and its state reconstructor.

The reconstructor is a copy of the system [with state $\hat{x}(t)$] with two inputs: the input $u(t)$ of the system and the difference $[\hat{y}(t) - y(t)]$ between the reconstructed output $\hat{y}(t)$ and the output of the system. The vector l identifies the reconstructor uniquely, and will be assumed to be constant in time. Thus, if the system is continuous-time, that is,

$$\dot{x}(t) = Ax(t) + bu(t)$$
$$y(t) = c^T x(t) + du(t)$$

the time invariant reconstructor is described by

$$\dot{\hat{x}}(t) = A\hat{x}(t) + bu(t) + l(\hat{y}(t) - y(t))$$
$$\hat{y}(t) = c^T \hat{x}(t) + du(t)$$

If the reconstruction error is the difference between the state $\hat{x}(t)$ of the reconstructor and the state $x(t)$ of the system

$$e(t) = \hat{x}(t) - x(t)$$

it is straightforward to check that

$$\dot{e}(t) = (A + lc^T)e(t) \tag{B.28}$$

and this means that the dynamics of the reconstruction error [Eq. (B.28) with initial state $e(0) = \hat{x}(0) - x(0)$], are independent upon the input applied to the system. In the case of discrete-time systems, Eq. (B.28) simply becomes

$$e(t+1) = (A + lc^T)e(t) \tag{B.29}$$

so that, we can conclude that the reconstructed state $\hat{x}(t)$ tends toward the state of the system $x(t)$, for any initial error $\hat{x}(0) - x(0)$, if and only if the system (B.28)

or (B.29) is asymptotically stable. If this is the case, we say that the reconstructor l is an asymptotic state reconstructor. Clearly, the rate of convergence of the reconstructed state $\hat{x}(t)$ toward $x(t)$ is determinated by the dominant eigenvalue of the matrix $A + lc^T$. In this context, the following result, which is the dual of those concerning the assignment of the eigenvalues of the controlled system, if of interest.

THEOREM 18 *(eigenvalues of the reconstructor)*

The eigenvalues of the matrix $A+lc^T$, which describes the reconstruction error dynamics (B.28) or (B.29), can be arbitrarily fixed by means of a suitable choice of the vector l, if and only if the system (A, c^T) is completely observable.

This result states that it is possible to rapidly and precisely reconstruct the state of a completely observable linear system by elaborating its inputs and outputs in real time. Since complete observability is a property that holds generically for a linear system, one can argue that the scheme of *Fig. B.14* is of great interest in applications.

B.17 DECOMPOSITION THEOREM

The notions of reachability and observability enable us to interpret any linear system as the interconnection of four subsystems called, respectively,

a. Reachable and unobservable part (r,no).

b. Reachable and observable part (r,o).

c. Unreachable and unobservable part (nr,no).

d. Unreachable and observable part (nr,o).

If the dimension of the system is n and n_a, n_b, n_c, and n_d are the dimensions of the four parts, obviously

$$n = n_a + n_b + n_c + n_d$$

A system is rarely composed of the four parts. In contrast, very often a system is composed only of part b: This happens when the system is completely reachable and completely observable. The interactions among the four subsystems $\Sigma_a, \Sigma_b, \Sigma_c$, and Σ_d are pointed out in *Fig. B.15* where, for simplicity, we have assumed that the system is proper $(d = 0)$.

The figure shows that the input u directly influences parts a and b but does not influence even indirectly, the parts c and d. This means that if the subsystems (c) and (d) are initially at rest $[x_c(0) = 0, x_d(0) = 0]$ they will remain at rest forever. On the contrary, the state vectors z_a and z_b of the first two parts, may

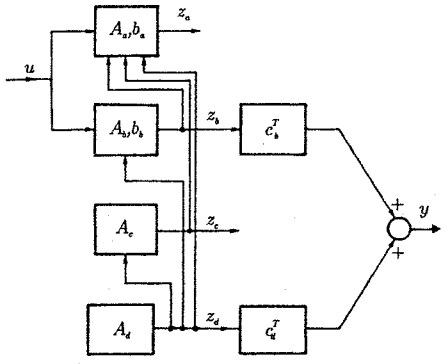

Figure B.15 A proper system decomposed in parts.

vary since they are influenced by the input. Moreover, the decomposition theorem (due to Kalman) states that the system composed of the first two parts is completely reachable. The figure shows also that the output is influenced by the input through part b, which is, therefore, the only channel through which the information flows from the input to the output of a dynamical system. The output is also influenced by part d but is completely insensitive to what is going on in parts a and c: This means that it will not be possible to compute the initial state of parts a and c, from the knowledge of the input and output functions. However, this will be possible for parts b and d, which compose a completely observable system.

THEOREM 19 *(decomposition theorem)*

Given a linear system (A, b, c^T), it is possible to perform a change of state variables $z = Tx$ in such a way that the equivalent system $(TAT^{-1}, Tb, c^T T^{-1})$ is decomposed into four parts (as shown in *Fig. B.15*), that is,

$$TAT^{-1} = \begin{pmatrix} A_a & A_{ab} & A_{ac} & A_{ad} \\ 0 & A_b & 0 & A_{bd} \\ 0 & 0 & A_c & A_{cd} \\ 0 & 0 & 0 & A_d \end{pmatrix} \quad Tb = \begin{pmatrix} b_a \\ b_b \\ 0 \\ 0 \end{pmatrix}$$

$$c^T T^{-1} = \begin{pmatrix} 0 & c_b^T & 0 & c_d^T \end{pmatrix}$$

with the following properties. System (A_b, b_b, c_b^T) is completely reachable and observable, while system (A_r, b_r, c_r^T) given by

$$A_r = \begin{pmatrix} A_a & A_{ab} \\ 0 & A_b \end{pmatrix} \quad b_r = \begin{pmatrix} b_a \\ b_b \end{pmatrix}$$

$$c_r^T = \begin{pmatrix} 0 & c_b^T \end{pmatrix}$$

is completely reachable and system (A_o, b_o, c_o^T) given by

$$A_o = \begin{pmatrix} A_b & A_{bd} \\ 0 & A_d \end{pmatrix} \quad b_o = \begin{pmatrix} b_b \\ 0 \end{pmatrix}$$

$$c_o^T = \begin{pmatrix} c_b^T & c_d^T \end{pmatrix}$$

is completely observable

The proof of *Theorem 19* is constructive. It suggest the following procedure for the computation of the coordinate transformation T.

Decomposition procedure

1. Compute the reachability and observability matrices R and O of system (A, b, c^T).
2. Determine the four subspaces

$$X_r = \mathcal{I}[R] \quad X_o = \mathcal{I}[O^T]$$

$$X_{nr} = \mathcal{N}[R^T] \quad X_{no} = \mathcal{N}[O]$$

3. Determine the four subspaces (*i.e.*, their basis)

$$X_a = X_r \cap X_{no}$$

$$X_b = X_r \cap (X_{nr} + X_o)$$

$$X_c = X_{no} \cap (X_{nr} + X_o)$$

$$X_d = X_{nr} \cap X_o$$

4. The columns of the matrix T^{-1} are the vectors of the basis of the four subspaces determined at step (3).
5. Determine $T = (T^{-1})^{-1}$ and $(TAT^{-1}, Tb, c^T T^{-1})$.

It is important to note that the four parts composing a linear system, though interconnected one to the other, do not form cycles, as pointed out by *Fig. B.15* and by the block triangular structure of the matrix TAT^{-1}. This implies that the eigenvalues of the system are the union of the eigenvalues of the four parts or, equivalently, that the characteristic polynomial of the system is the product of the characteristic polynomials of the four parts. For this reason, many properties of linear systems are linked to the stability of one or more of its parts. We now confirm this by discussing three properties of linear systems: stabilizability, detectability, and external stability.

The first property, *stabilizability*, concerns the possibility of transforming a given system (A, b, c^T, d) into an asymptotically stable system, using a linear control law. Assuming that the system is decomposed into its four parts with state vectors z_a, z_b, z_c, and z_d, the linear control law becomes

$$u(t) = k_a^T z_a(t) + k_b^T z_b(t) + k_c^T z_c(t) + k_d^T z_d(t)$$

By looking at *Fig. B.15* it is clear that such a control law does not modify the dynamics of parts c and d, whose eigenvalues remain eigenvalues of the controlled system. For the system (A, b, c^T, d) to be stabilizable, it is then necessary that its parts c and d be asymptotically stable. But, this condition is also sufficient since, parts a and b compose a completely reachable system so that from *Theorem 14* it follows that their eigenvalues can be modified at will. In conclusion, the following result (*Theorem 20*) holds:

THEOREM 20 *(stabilizability condition)*

A system is stabilizable if and only if its unreachable parts (c and d) are asymptotically stable.

A dual result holds for the so-called *detectability*, namely, for the possibility of reconstructing, at least asymptotically, the state of a system by means of a linear time-invariant reconstructor.

THEOREM 21 *(stabilizability condition)*

A system is detectable if and only if its unobservable parts (a and c) are asymptotically stable.

This result can be understood by a simple inspection of *Fig. B.15*. In fact, since parts b and d compose a completely observable system, their state variables $z_b(t)$ and $z_d(t)$ can be reconstructed from the input and output functions (see *Theorem 18*). But then, the inputs $u(t), z_b(t)$ and $z_d(t)$ of the system composed of parts a and c are known so that the forced motion of such a system can be computed. But, if the parts a and c are asymptotically stable, the forced motion tends, as time goes on, toward the state vectors $z_a(t)$ and $z_c(t)$, so that, in conclusion, the system is detectable.

We now give the definition of *external stability*, also known as *bounded input bounded output* (BIBO) stability.

DEFINITION 3 *(external stability)*

A linear system is externally stable if its forced output is bounded for any bounded input.

From *Fig. B.15*, one can readily see that external stability is a property of part b of the system, since the forced motion is characterized by $z_c(t) = 0$ and $z_d(t) = 0$, while part a gives no contribution to the output. It is not surprising, then, that the external stability of a system is equivalent to the asymptotic stability of its part b as stated in *Theorem 22*.

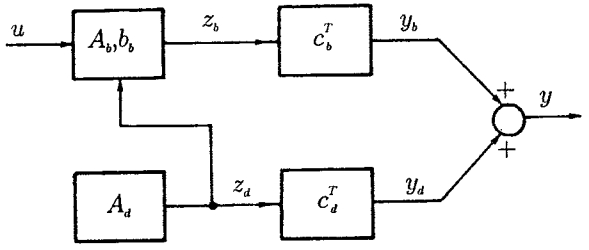

Figure B.16 Observable parts b and d of a system.

THEOREM 22 *(external stability condition)*

A system is externally stable if and only if its reachable and observable part is asymptotically stable.

This result is due to the fact that only part b is responsible for the relationship between input and output whenever the system is initially at rest. If the initial state is nonzero, the output depends also on part d, which gives, however, a bounded contribution if it is simply or asymptotically stable. In conclusion, the output of a linear system is bounded for any initial state if and only if its reachable and observable part b is asymptotically stable and its observable and unreachable part d is stable.

B.18 DETERMINATION OF THE ARMA MODELS

We can now be more precise on the problem dealt with in *Section B.3*, namely, the determination of the ARMA model of a given system (A, b, c^T, d). For this, recall that an ARMA model is the pair of polynomials $[N(p), D(p)]$ identifying the input–output equation (B.8), which is the difference equation (B.6) in the discrete-time case and the differential equation (B.7) in the continuous-time case. Recall also that the ARMA model is said to be *reduced* if the polynomials $N(p)$ and $D(p)$ are coprime. Moreover, the ratio between $N(p)$ and $D(p)$ is the *transfer function* of the system denoted by $G(p)$, that is,

$$G(p) = \frac{N(p)}{D(p)}$$

Obviously, the ARMA model $[N(p), D(p)]$ and the transfer function $G(p)$ are not equivalent unless the ARMA model is reduced. Since the ARMA model represents the relation between input and output in the general case of a nonzero initial state, it must be associated with the observable parts b and d of the system, which, for the sake of clarity, are depicted in *Fig. B.16*.

The first subsystem with output y_b is described by the (multiple inputs) ARMA model

$$\Delta_b(p)y_b(t) = N_b(p)u(t) + N_b^1(p)z_{d1}(t) + N_b^2(p)z_{d2}(t) + \cdots \tag{B.30}$$

where $\Delta_b(p)$ is the characteristic polynomial of A_b and $z_{d1}(t), z_{d2}(t)$, and so on, are the components of the state vector $z_d(t)$. The second subsystem with output y_d has no input and is therefore described by the AR model

$$\Delta_d(p)y_b(t) = 0 \tag{B.31}$$

where $\Delta_d(p)$ is the characteristic polynomial of A_d. Moreover, since each component of the state vector $z_d(t)$ can be interpreted as an output of the second subsystem, we have

$$\Delta_d(p)z_{di}(t) = 0 \quad i = 1, 2, \ldots, n_d \tag{B.32}$$

If Eqs. (B.30) and (B.31) are multiplied by $\Delta_d(p)$ and $\Delta_b(p)$, respectively, and then summed up, the resulting equation in view of Eq. (B.32) becomes

$$\Delta_b(p)\Delta_d(p)y(t) = N_b(p)\Delta_d(p)u(t)$$

where $y(t) = y_b(t) + y_d(t)$. Thus, the ARMA model of the system is not reduced, since

$$D(p) = \Delta_b(p)\Delta_d(p)$$

and

$$N(p) = N_b(p)\Delta_d(p)$$

are not coprime.

If part d is missing, that is, if the system (A, b, c^T, d) does not have the unreachable and observable part, then $\Delta_d(p) = 1$ and the ARMA model

$$\Delta_b(p)y(t) = N_b(p)u(t)$$

is reduced, since $\Delta_b(p)$ and $N_b(p)$ are coprime (part b being completely reachable and observable). We can then conclude this section stating *Theorem 23*.

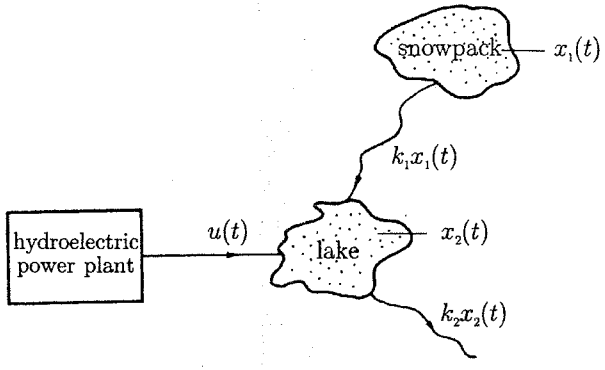

Figure B.17 A simple hydraulic system.

THEOREM 23 *(characterization of the ARMA model)*

The ARMA model $[N(p), D(p)]$ of a system (A, b, c^T, d) is the ARMA model of the system composed of its observable parts (b and d) and

$$N(p) = N_b(p)\Delta_d(p) \quad D(p) = \Delta_b(p)\Delta_d(p)$$

where $[\Delta_b(p), N_b(p)]$ is the ARMA model of the reachable and observable part and $\Delta_d(p)$ is the characteristic polynomial of the unreachable and observable part (equal to 1 if such a part is missing). Then, the ARMA model of a system is in reduced form if and only if the system does not have the unreachable and observable part. Moreover, the transfer function $G(p) = N(p)/D(p)$ of the system is the transfer function $G_b(p) = N_b(p)/\Delta_b(p)$ of the reachable and observable part.

EXAMPLE 8

Consider the hydraulic system represented in *Fig. B.17* composed of a lake with two inflows, one with flow rate $u(t)$ (discharge of a plant) and the other $k_1x_1(t)$ [melting of a snow-pack of volume $x_1(t)$].

If the flow rate of the effluent is assumed to be proportional through a coefficient k_2 to the water storage $x_2(t)$ of the lake, the mass conservation law gives

$$\begin{aligned}\dot{x}_1 &= -k_1x_1\\ \dot{x}_2 &= k_1x_1 - k_2x_2 + u\end{aligned}$$

Thus, if one considers the flow rate of the effluent as output variable $y(t)$, the system is described by the triple

$$A = \begin{pmatrix} -k_1 & 0 \\ k_1 & -k_2 \end{pmatrix} \quad b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$c^T = \begin{pmatrix} 0 & k_2 \end{pmatrix}$$

Such a system is completely observable but not completely reachable since the variable x_1 cannot be influenced. The system is therefore composed of part b (lake) and part d (snow-pack) as depicted in *Fig. B.16*. Since the eigenvalues of A are $-k_1$ (snow-pack) and $-k_2$ (lake) we have

$$\Delta_b(s) = (s + k_2) \quad N_b(s) = k_2 \quad \Delta_d(s) = (s + k_1)$$

so that

$$D(s) = (s + k_1)(s + k_2) \quad N(s) = k_2(s + k_1)$$

In conclusion, the flow rates $u(t)$ and $y(t)$ are linked by the second-order differential equation

$$\ddot{y}(t) + (k_1 + k_2)\dot{y}(t) + k_1k_2y(t) = k_2\dot{u}(t) + k_1k_2u(t)$$

In contrast, if the snow-pack is missing, the model becomes

$$\dot{y}(t) + k_2y(t) = k_2u(t)$$

which is a reduced ARMA model.

♣

In many cases of practical interest, the model of the system (A, b, c^T, d) is not known, but a pair $u(\cdot), y(\cdot)$ of input and output records of length T is available. This is, for example, the case of a river basin in which rainfall and outflow have been recorded for a period of a few months. Another example is the case of an electrical amplifier in which the input and output signals have been measured for some seconds. In these cases, it is interesting to know if it is possible to determine the model (A, b, c^T, d) of the system by processing the recorded input and output data. This problem is known as *identification* of the model and is of paramount importance in applications. Very often a solution is obtained by assuming that input and output measurements are affected by noise, and by using suitable notions of the theory of stochastic processes. However, the problem is theoretically and practically interesting even in the absence of noise. For this assume, that the system is proper and discrete-time and that the dimension $n^\circ = n_b + n_d$ of its observable part is known. Under these assumptions, the ARMA model of the system is

$$y(t) = -\alpha_1y(t-1) - \cdots - \alpha_{n^0}y(t-n^0) + \beta_1u(t-1) + \cdots + \beta_{n^0}u(t-n^0)$$

which can be written in the more compact form

$$y(t) = -(y_{t-n^0}^{t-1})^T\alpha + (u_{t-n^0}^{t-1})^T\beta \tag{B.33}$$

where

$$\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{n^0} \end{pmatrix} \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{n^0} \end{pmatrix} y_{t-n^0}^{t-1} = \begin{pmatrix} y(t-1) \\ \vdots \\ y(t-n^\circ) \end{pmatrix} u_{t-n^0}^{t-1} = \begin{pmatrix} u(t-1) \\ \vdots \\ u(t-n^\circ) \end{pmatrix}$$

Suppose, now, that a recorded time series composed of N input and output values

$$u(0), u(1), \ldots, u(N-2), u(N-1)$$
$$y(0), y(1), \ldots, y(N-2), y(N-1)$$

is known and write Eq. (B.33) in the $2n^\circ$ unknowns α_i and $\beta_i, i = 1, \ldots, n^\circ$, for $N - n^\circ$ successive values of t, that is, for $t = n^\circ, n^\circ + 1, \ldots, N-1$. This leads to the following system of $N - n^\circ$ linear equations in $2n^\circ$ unknowns:

$$\begin{pmatrix} -(y_0^{n^\circ-1})^T & (u_0^{n^\circ-1})^T \\ -(y_1^{n^\circ})^T & (u_1^{n^\circ})^T \\ \vdots & \vdots \\ -(y_{N-n^\circ-1}^{N-2})^T & (u_{N-n^\circ-1}^{N-2})^T \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} y(n^\circ) \\ y(n^\circ+1) \\ \vdots \\ y(N-1) \end{pmatrix}$$

This is an algebraic linear system of the kind

$$Fp = y_{N-1}^{n^\circ} \tag{B.34}$$

where p is the unknown vector of parameters α_i and β_i identifying the ARMA model and F is an $(N - n^\circ) \times (2n^\circ)$ matrix depending on the input and output data. By excluding special critical cases, this algebraic system can be solved if

$$N \geq 3n^\circ$$

In the case $N = 3n^\circ$, the solution is

$$\hat{p} = F^{-1} y_{N-1}^{n^\circ}$$

while in the case $N > 3n^\circ$ the solution can be given in the form

$$\hat{p} = (F^T F)^{-1} F^T y_{N-1}^{n^\circ} \tag{B.35}$$

The critical cases, of nonidentifiability are those in which the matrix F is not full rank, so that the matrix $F^T F$ is not invertible. These cases occur, for example, when the input and output data are collected during a period of time in which the system is at equilibrium (steady state). In fact, in such conditions the first [second] n° columns of the matrix F are identical because the output [input] does not vary in time. Another case of nonidentifiability occurs when the initial state of the unreachable and observable part is zero. In fact, under this circumstance, the output is not influenced by part d of the system, see *Fig. B.16*, so that the coefficients of the characteristic polynomial $\Delta_d(p)$ are not identifiable. This means that the ARMA model of the system is not identifiable since $N(p) = N_b(p)\Delta_d(p)$

and $D(p) = \Delta_b(p)\Delta_d(p)$. The previous discussion can be summarized in *Theorem 24*.

THEOREM 24 *(identifiability of ARMA models)*

The ARMA model of a proper discrete-time system with known dimension $n^\circ = n_b + n_d$ of the observable parts cannot be identified from a series of N input and output values if $N < 3n^\circ$. On the contrary, if $N \geq 3n^\circ$ the ARMA model is uniquely identified, apart from some critical cases (of nonidentifiability).

Once the ARMA model has been identified, it is possible to construct a triple (A, b, c^T, d), which realizes it. Such a triple is the reconstruction canonical form

$$A_r = \begin{pmatrix} 0 & 0 & \ldots & 0 & -\alpha_{n^\circ} \\ 1 & 0 & \ldots & 0 & -\alpha_{n^\circ-1} \\ 0 & 1 & \ldots & 0 & -\alpha_{n^\circ-2} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -\alpha_1 \end{pmatrix} \quad b_r = \begin{pmatrix} \beta_{n^\circ} \\ \beta_{n^\circ-1} \\ \beta_{n^\circ-2} \\ \vdots \\ \beta_1 \end{pmatrix}$$
$$c_r^T = (\, 0 \;\; 0 \;\; \ldots \;\; 0 \;\; 1 \,)$$

In fact, from *Theorem 17*, such a system is completely observable, so that it is composed of parts b and d, which characterize the ARMA model. If the ARMA model is in reduced form, that is, if the polynomials

$$D(p) = p^{n^\circ} + \alpha_1 p^{n^\circ-1} + \cdots + \alpha_{n^\circ}$$
$$N(p) = \beta_1 p^{n^\circ-1} + \cdots + \beta_{n^\circ}$$

are coprime, the triple (A_r, b_r, c_r^T) is also reachable, namely it is composed of part (b) only. In contrast, if the ARMA model is not in reduced form, namely, if

$$D(p) = r(p)d(p)$$
$$N(p) = r(p)n(p)$$

then, the system is composed of a reachable and observable part described by the ARMA model $[n(p), d(p)]$ and of an unreachable and observable part, described by an AR model $r(p)$, which coincides with the characteristic polynomial $\Delta_d(p)$ of such a part.

The control canonical form

$$A_c = A_r^T \quad b_c = c_r \quad c_c^T = b_r^T$$

obtained by duality from the realization in reconstruction canonical form, is not a realization of the ARMA model $[N(p), D(p)]$, if such a model is not in reduced

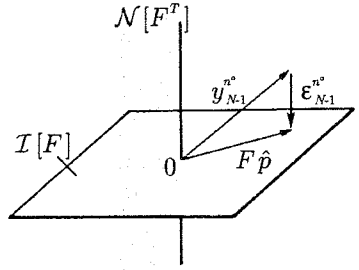

Figure B.18 Illustration of the least-square estimation principle.

form. In fact, the system (A_c, b_c, c_c^T) is completely reachable (see *Theorem 13*) so that it cannot be composed of parts b and d. Therefore, the ARMA model of the control canonical form (A_c, b_c, c_c^T) is $[n(p), d(p)]$ instead of $[N(p), D(p)]$. In other words, the control canonical form realizes the transfer function $G(p) = N(p)/D(p)$ of the system but not the ARMA model. Obviously, if the system is completely reachable and observable, the ARMA model is in reduced form and the control canonical form is one of its realizations.

If the input and output values contain errors, Eq. (B.34) must be replaced by

$$Fp - y^{n^\circ}_{N-1} = \varepsilon^{n^\circ}_{N-1}$$

where the vector ε represents the difference between the output values predicted by the ARMA model and the measured output values. In *Fig. B.18*, the measured output vector and the subspace $\mathcal{I}[F]$ of the output predicted by the ARMA model are shown in the space of dimension $N - n^\circ$. It is then natural to choose the ARMA model that is, the value $\hat{p}$ of p, that minimizes the distance between the measured and the predicted output vectors.

As illustrated in *Figure B.18*, this amounts to choosing $\hat{p}$ in such a way that the vector $\varepsilon^{n^\circ}_{N-1}$ is orthogonal to $\mathcal{I}[F]$. But, since $\mathcal{I}[F]^\perp = \mathcal{N}[F^T]$, this is equivalent to

$$F^T \varepsilon^{n^\circ}_{N-1} = 0$$

namely,

$$F^T(F\hat{p} - y^{n^\circ}_{N-1}) = 0$$

from which, assuming that $F^T F$ is invertible, it follows that:

$$\hat{p} = (F^T F)^{-1} F^T y^{n^\circ}_{N-1} \tag{B.36}$$

which coincides with Eq. (B.35).

The estimation $\hat{p}$ given by (B.36) is known as the *least-square estimation* because it minimizes the sum of the squares of the differences between predictions and actual

measurements. This estimation, here interpreted in geometrical terms, possesses a number of peculiar properties for specific statistical characteristics of the input and output measurement errors. Moreover, formula (B.35) can be fruitfully given in a recursive form, in such a way that the computation of $\hat{p}$ can be updated in real time, without the need of computing the inverse of a $2n^\circ \times 2n^\circ$ matrix each time a new pair of input–output data is available.

B.19 POLES AND ZEROS OF THE TRANSFER FUNCTION

We have already mentioned that the transfer function $G(p)$ of a linear system is, by definition, the ratio of the two polynomials $N(p)$ and $D(p)$ that identify the ARMA model of the system

$$D(p)y(t) = N(p)u(t)$$

From *Theorem 23*, we can immediately conclude that $G(p)$ coincides with the transfer function of the reachable and observable part of the system, that is,

$$G(p) = \frac{N_b(p)}{\Delta_b(p)}$$

where $N_b(p)$ and $\Delta_b(p)$ are the two coprime polynomials characterizing the ARMA model of the reachable and observable part. The zeros of the polynomials $N_b(\cdot)$ and $\Delta_b(\cdot)$ are called, respectively, *zeros* and *poles of the transfer function* (or of the system) and are denoted by z_i and p_i. The transfer function of a proper system with a reachable and observable part of dimension n can be written in the following form:

$$G(p) = \frac{\beta_r p^{n-r} + \beta_{r+1} p^{n-r-1} + \cdots + \beta_n}{p^n + \alpha_1 p^{n-1} + \cdots + \alpha_n}$$

where $r \geq 1$ is the so-called *relative degree*, or in the form

$$G(p) = \rho \frac{(p - z_1)(p - z_2) \ldots (p - z_{n-r})}{(p - p_1)(p - p_2) \ldots (p - p_n)}$$

where ρ is called *transfer constant*. Poles and zeros are of paramount importance in a number of problems in systems and control theory. We shall see in a moment that it is particularly interesting to know if a continuous [discrete] -time system has all its poles and zeros with negative real part [modulus < 1]. In other words, we are interested to know whether the poles and zeros are "stable" or not. In view of the decomposition theorem, it follows that the poles of the transfer function are the eigenvalues of the reachable and observable part, so that (see *Theorem 22*) a system is externally stable if and only if its poles are stable. Moreover, the output of a system is bounded for any bounded input if and only if its poles are stable and the unreachable and observable part d does not exist or is asymptotically or simply stable. Under these conditions, the solutions of the ARMA model

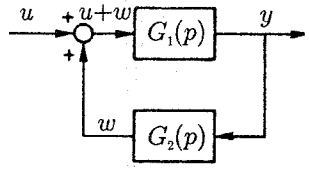

Figure B.19 Structure of a system equivalent to a completely reachable and observable system [$G_1(p)$ has no zeros].

$$\Delta_b(p)\Delta_d(p)y(t) = N_b(p)\Delta_d(p)u(t)$$

for different initial conditions and for the same bounded input $\hat{u}(t)$, are different bounded outputs $\hat{y}(t)$ which do not diverge unlimitatedly one from each other. Moreover, if the unreachable and observable part is missing or asymptotically stable, the outputs $\hat{y}(t)$ of the system tend toward the same function $y(t)$ which can be computed with great accuracy for t sufficiently large using the reduced ARMA model

$$\Delta_b(p)y(t) = N_b(p)u(t)$$

To fully understand the role of the zeros in the dynamics of a linear system, it is necessary to refer to the particular canonical form shown in *Fig. B.19*. This is always possible since a completely reachable and observable system (A, b, c^T) with n poles and $(n-r)$ zeros, can always be put, by means of an appropriate change of coordinates $z = Tx$, in the form reported in *Fig. B.19*, where the subsystem in the forward path, has dimension r and has no zeros.

If $G_1(p)$ and $G_2(p)$ are the transfer functions of the two subsystems, from the formula

$$G(p) = \frac{G_1(p)}{1 - G_1(p)G_2(p)}$$

we can conclude that the poles of $G_2(p)$ are the zeros of $G(p)$. Therefore, if (A_2, b_2, c_2^T) is the triple that defines the subsystem in the feedback path, the eigenvalues of A_2 are the zeros of the system and the free output of the feedback subsystem, obtained with $y(t)$ identically zero, is

$$w(t) = c_2^T e^{A_2 t} z_2(0)$$

If the system in the forward path is initially at rest $[z_1(0) = 0]$ and the signal $w(t)$ is compensated by the input $u(t) = -w(t)$, the system in the forward path is not excited from the outside and remains, consequently, at rest [*i.e.*, $y(t) \equiv 0$). This means that the output of the system can be identically zero even if its input

is not. This happens when the initial state is appropriately chosen $[z_1(0) = 0]$ and the input $u(t)$ is the output of an autonomous system $(A_2, -, c_2^T)$ with eigenvalues equal to the zeros of the system. In other words, the zeros of a system completely determine the dynamics of its "hidden" inputs.

Systems with no zeros or with strictly stable zeros are called *minimum phase systems*. They have no hidden inputs or hidden inputs that asymptotically tend to zero at a speed dictated by the "dominant" zero. In contrast, continuous-time [discrete-time] nonminimum phase systems have zeros with nonnegative real part [modulus not < 1] and therefore have hidden inputs not tending to zero. The knowledge of a record of the output of a completely reachable and observable linear system allows one to reconstruct an input $\hat{u}(t)$ of the system, which is the sum of the true input $u(t)$ and of a hidden input. But if the system is minimum phase, the hidden input tends to zero as $t \to \infty$ so that $\hat{u}(t)$ tends toward the true input $u(t)$. The reconstruction algorithm is still an ARMA model

$$\Delta_b(p)y(t) = N_b(p)u(t)$$

which must be solved with respect to $u(t)$. In the case of a discrete-time system, this implies the recursive solution of the equation

$$y(t)+\alpha_1 y(t-1)+\cdots+\alpha_n y(t-n) = \beta_r \hat{u}(t-r)+\beta_{r+1}\hat{u}(t-r-1)+\cdots+\beta_n \hat{u}(t-n)$$

with respect to $\hat{u}(t-r)$. Note that this operation cannot be performed in real time since the evaluation of $\hat{u}(t-r)$ requires the knowledge of $y(t)$. At best, the input can be reconstructed after r transitions.

Clearly, the problem of the hidden inputs and of the reconstruction of the inputs from the outputs is well posed also when the system has an unreachable and observable part. In this case, taking into account *Fig. B.2*, one obtains the block diagram shown in *Fig. B.20*.

Such a diagram shows that the hidden inputs can be divided into two groups: those "generated" by the zeros of the system and those "generated" by the eigenvalues of the unreachable and observable part [zeros of the polynomial $\Delta_d(p)$]. Therefore, the hidden inputs tend to zero, if the system is minimum phase and its unreachable and observable part is asymptotically stable. In this case, the input can be reconstructed by solving the nonreduced ARMA model

$$\Delta_b(p)\Delta_d(p)y(t) = N_b(p)\Delta_d(p)u(t)$$

with respect to $u(t)$ or, alternatively the reduced ARMA model

$$\Delta_b(p)y(t) = N_b(p)u(t)$$

This section can be summarized by noting that poles and zeros play roles that are "dual" in some way. In fact, in the long run, the output of a completely reachable and observable system also can be computed, given its input, with no information on the initial state provided that the poles of the system are stable (external stability).

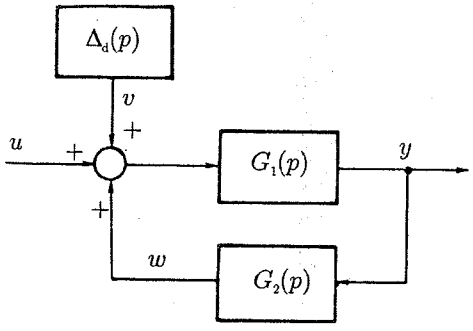

Figure B.20 Block diagram of a system with a nonreduced ARMA model: $\Delta_d(p)$ is the characteristic polynomial of the unreachable and observable part and $G_1(p)$ has no zeros.

Dually, in the long run, the input of a completely reachable and observable system can be computed, given its output, with no information on the initial state, provided that the zeros of the system are stable (minimum phase). In other words, the stability of the poles allows one to neglect the free motion in the long run, while the stability of the zeros allows one to neglect the hidden inputs. The reader can easily formulate these properties for systems with an unreachable and observable part.

B.20 POLES AND ZEROS OF INTERCONNECTED SYSTEMS

As shown in Section B4, the transfer function of two interconnected systems (*Figs. B.3, B.4,* and *B.5*) is given by

$$G(p) = G_1(p)G_2(p) \qquad \text{series connection}$$

$$G(p) = G_1(p)G_2(p) \qquad \text{parallel connection}$$

$$G(p) = \frac{G_1(p)}{1 + G_1(p)G_2(p)} \qquad \text{(negative) feedback connection}$$

where $G_1(p)$ and $G_2(p)$ are the transfer functions of the two subsystems. Apart from critical cases (related to the nonreachability and nonobservability of the resulting system), in which the denominator of the transfer function $G(p)$ turns out to be a polynomial of degree smaller than the sum of the degrees of the denominators of the two transfer functions $G_1(p)$ and $G_2(p)$, we can immediately conclude the following:

Series. The poles and the zeros of $G(p)$ are the union of those of $G_1(p)$ and $G_2(p)$.

Parallel. The poles of $G(p)$ are the union of those of $G_1(p)$ and $G_2(p)$.

Feedback. The zeros of $G(p)$ are the union of the zeros of $G_1(p)$ and of the poles of $G_2(p)$.

The computation of poles and zeros of interconnected systems is, therefore, immediate, except for the computation of the zeros of systems connected in parallel and of the poles of systems connected in feedback. These two cases however, are equivalent since the zeros of the transfer function

$$G(p) = G_1(p) + G_2(p)$$

of the system in *Fig. B.21(a)* coincide with the poles of the transfer function

$$G(p) = \frac{\dfrac{G_1(p)}{G_2(p)}}{1 + \dfrac{G_1(p)}{G_2(p)}} = \frac{G_1(p)}{G_1(p) + G_2(p)}$$

of the system in *Fig. B.21(b)*.

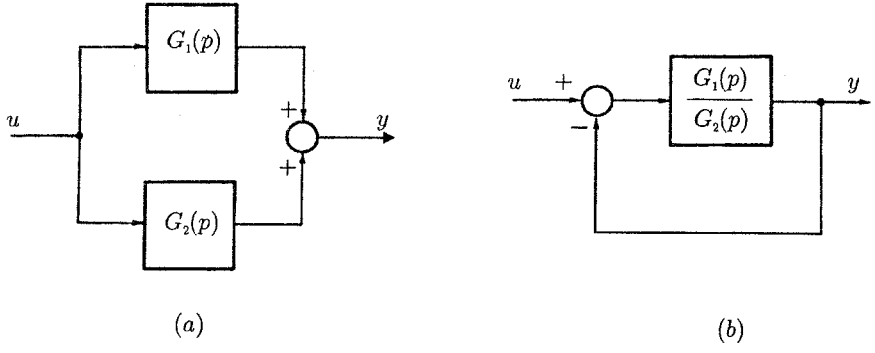

Figure B.21 The zeros of system (a) coincide with the poles of system (b).

We can then conclude that there is only one significant problem, namely, the determination of the poles of a system composed of two subsystems connected in feedback. This is the central problem of classical control theory, mainly focused on the determination of feedback systems with appropriate dynamic properties as, for example, external stability.

In applications, it is often important to determine poles and zeros when some parameter is varied (typically, a design parameter). Though today this can be simply done by means of specific software, we briefly describe a method called *root locus*, which has been often used in the past for the design of control systems, and is of great value still today for the discussion of the stability of feedback systems.

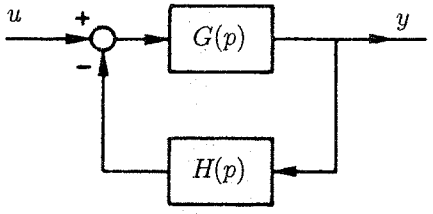

Figure B.22 Feedback system.

The root locus is by definition, the locus described in the complex plane by the poles of the feedback system shown in *Fig. B.22* when the transfer constants of the two subsystems are allowed to vary. It is, therefore, composed of n curves, one for each pole, called "branches" of the locus.

If the product of the two transfer constants is positive [negative] and assumes all values, from 0 to $\infty[-\infty]$ we obtain the direct [inverse] locus. When the transfer constants are related to a design parameter (as it is usually the case) and one wants to obtain an externally stable system, one must check whether the n segments of the branches of the locus are "stable" for the feasible values of the design parameters. This amounts to checking whether the n segments of the "locus" are in the left half-plane [unitary circle] of the complex plane if the system is continuous-time [discrete-time]. In *Fig. B.23*, six examples of direct root loci are depicted: poles and zeros of the two transfer functions $G(p)$ and $H(p)$ are represented by crosses ($\times$) and circles ($\circ$), respectively.

Obviously, there is no reason to distinguish the poles [zeros] of $G(p)$ from those of $H(p)$. In fact, if

$$G(p) = \rho_G \frac{\Pi(p - z_i^G)}{\Pi(p - p_i^G)} \qquad H(p) = \rho_H \frac{\Pi(p - z_i^H)}{\Pi(p - p_i^H)}$$

then the transfer function $F(p)$ of the system is

$$F(p) = \frac{G(p)}{1 + G(p)H(p)}$$

and its poles are the roots of the equation

$$1 + G(p)H(p) = 0$$

in which only the product $G(p)H(p)$ appears. Such equation can be fruitfully written in the form

$$k\Pi(p - z_i^G)\Pi(p - z_i^H) = -\Pi(p - p_i^G)\Pi(p - p_i^H) \qquad \text{(B.37)}$$

where the parameter k, which is positive in the direct locus and negative in the inverse one, is the product of the two transfer constants, namely, $k = \rho_G \rho_H$. The

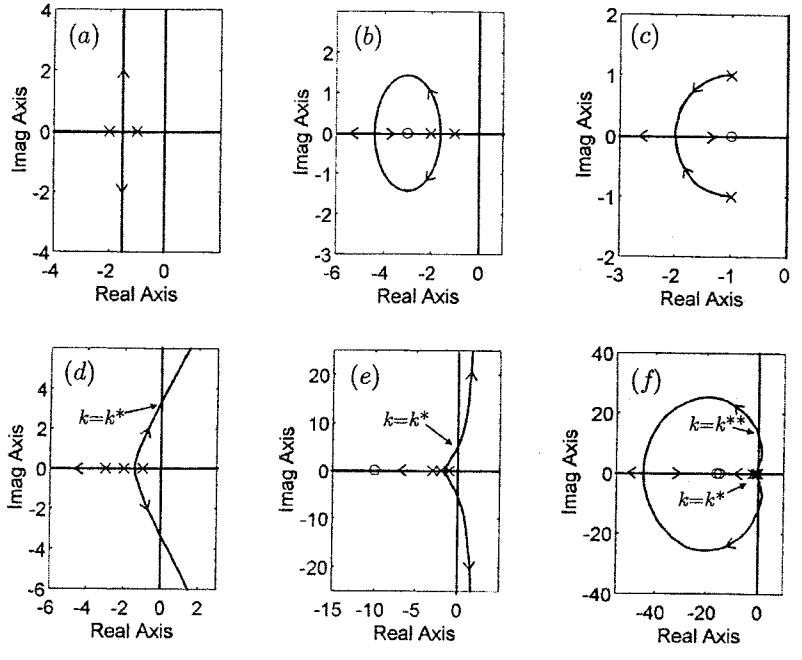

Figure B.23 Root loci of second- and third-order systems: (a) two poles; (b) and (c) two poles and one zero; (d) three poles; (e) three poles and one zero; (f) three poles and two zeros.

loci (a), (b), and (c) of *Fig. B.23* refer to a feedback system composed of two subsystems, which have only two poles, while in cases (d), (e), and (f) there are three poles. Consistently, the first three loci are composed of two branches, while the last three are composed of three branches. If we imagine that we are dealing with continuous-time systems, we can infer that the feedback system is externally stable for all values of k in cases a, b and c, only for $k < k^*$ in cases d and e and for $k < k^*$ and $k > k^{**}$ in case f. The values k^* and k^{**} are very important, since they mark the transition from stability to instability.

The six loci depicted in *Fig. B.23* show some general properties of the root locus that are worth noting. First of all, the locus is symmetrical with respect to the real axis. Moreover, each branch starts from a pole [of $F(p)$ or $G(p)$] since for $k \to 0$ the roots of (B.37) tend to p_i^G and p_i^H. On the other hand, for $k \to \infty$, $(n - r)$ branches tend to the zeros z_i^G and z_i^H [see (B.37)] while the remaining r tend to infinity forming an angle of $2\pi/r$. Finally, all the points of the real axis that have on their right an odd [even] number of singularities (poles and zeros) belong to the direct [inverse] locus. All such properties can be easily proved. In contrast,

other properties as the "rule of the center of mass" are less immediate. Such a rule states that if the relative degree r is ≥ 2, the sum of the n poles is independent of k. This can be checked by writing Eq. (B.37) in the form

$$p^n + \gamma_1 p^{n-1} + \gamma_2 p^{n-2} + \cdots = 0$$

and by noting that γ_1, which is equal to the opposite of the sum of the poles, is independent of k if $r \geq 2$. The consequences of this rule are evident in the loci a and e in *Fig. B.23*. In case a, the point in which the two branches collide when k increases is the central point of the segment connecting the two poles. In case e, since for $k = 0$ the sum of the three poles is equal to -6 and for $k \to \infty$ one of the three poles tends to the zero located at -8, the other two poles must have, for $k \to \infty$, real part equal to 1.

All such rules often allows one to discuss qualitatively, but effectively, the external stability of a feedback system when a design parameter is varied. From *Fig. B.21*, it is clear that the same rules also allow the discussion of the minimum phase of systems composed by subsystems connected in parallel.

B.21 IMPULSE RESPONSE

The impulse response of a continuous-time linear system is, as the term itself suggests, the output of the system corresponding to an impulsive input. In order to uniquely define the impulse response, one must specify the initial state which, for simplicity, is assumed to be zero. The impulse response, denoted in the following by $g(t)$, is then the output of the system

$$\begin{aligned} \dot{x} &= Ax + bu \\ y &= c^T x \end{aligned}$$

with $x(0) = 0$ and $u(t) = \text{imp } t$.

The impulse response, can often be measured directly in the field or in the laboratory. For example, *Fig. B.24* reports the impulse responses of four systems. The first concerns the position of a point mass moving along a straight line after it has been hit by another point mass (impulsive force), the second is the voltage of an $R - C$ circuit fed by an impulse of current, the third is the flow of a river after a short but intensive storm in the river basin (impulsive rainfall), and the fourth is the behavior of the wings of an airplane after an air pocket (impulsive force).

From the Lagrange formula (*Theorem 1*) it follows that:

$$g(t) = c^T \int_0^t e^{A(t-\xi)} b \text{imp}\xi d\xi = c^T e^{At} \int_0^{0+} \text{imp}\xi d\xi b$$

that is,

$$g(t) = c^T e^{At} b$$

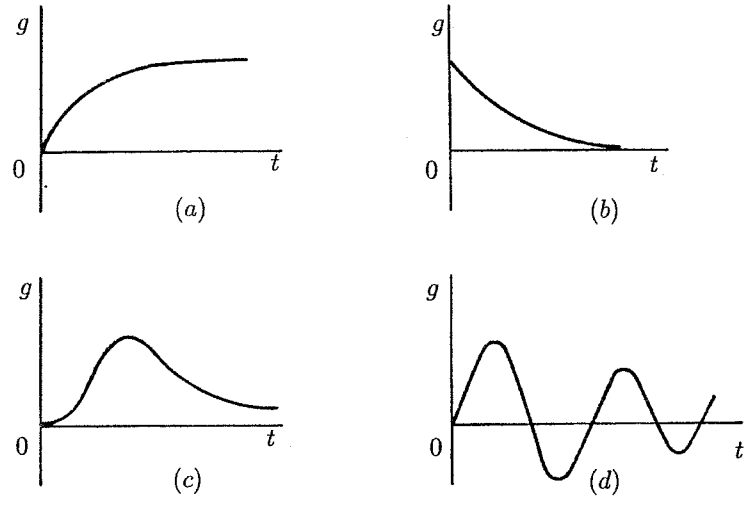

Figure B.24 Four impulse responses: (a) point mass; $R - C$ (b)electrical circuit; (c) river basin; (d) wings of an airplane.

This means that the impulse response is the free output of the system $g(t) = c^T e^{At} x(0)$ with $x(0) = b$ and this is consistent with the fact that the impulse steers the state ot the system form 0 to b in a time interval of zero measure.

By recalling that the reachable and observable part of a system (part b) is the only part determining the output in the case of a zero initial state, one can also write

$$g(t) = c_b^T e^{A_b t} b_b$$

From the Lagrange formula, it follows also that the forced evolution of a continuous-time linear system is

$$g(t) = c^T \int_0^t e^{A(t-\xi)} b u(\xi) d\xi = \int_0^t g(t-\xi) u(\xi) d\xi$$

namely, the forced output is the *convolution integral* of the impulse response and of the input.

Moreover, it is worth noting that

$$\begin{aligned} g(t) = c^T e^{At} b &= c^T \left(I + At + A^2 \frac{t^2}{2!} + \cdots \right) b \\ &= c^T b + c^T A b t + c^T A^2 b \frac{t^2}{2!} + \cdots \end{aligned}$$

so that, recalling the formula for the Taylor expansion of a function $g(t)$ in a neighborhood of the origin, we can conclude that

$$\left.\frac{d^i g(t)}{dt^i}\right|_{t=0} = c^T A^i b \quad i = 0, 1, 2, \ldots$$

The coefficients

$$g_1 = c^T b \quad g_2 = c^T A b \quad g_3 = c^T A^2 b \ldots$$

are known as *Markov coefficients*.

Other *canonical responses* of continuous-time linear systems are the *step response* and the *ramp response*, which are the output of the system with $x(0) = 0$ and

$$u(t) = \begin{cases} 1 & t \geq 0 \quad \text{step} \\ t & t \geq 0 \quad \text{ramp} \end{cases}$$

Since the step function is the integral of the impulse function and the ramp is the integral of the step, we can conclude that the step response is the integral of the impulse response and that the ramp response is the integral of the step response. By using analogous arguments one can define the impulse response of discrete-time systems, which turn out to be given by

$$g(t) = \begin{cases} 0 & t = 0 \\ c^T A^{t-1} b & t > 0 \end{cases}$$

From the above definition of Markov coefficients, one can conclude that the impulse response $g(t)$ of a discrete-time system is zero at time zero and equal to the corresponding Markov coefficients for $t > 0$, that is,

$$g(t) = g_t$$

As a useful exercise, the reader is invited to compute the impulse response of the systems considered in *Example 1* (Newton's law) and in *Example 2* (Fibonacci's rabbits).

B.22 FREQUENCY RESPONSE

For a large class of continuous-time linear systems, the periodic output corresponding to a sinusoidal input is unique and is actually a sinusoid with the same frequency as the input. This property directly leads to the definition of *frequency response* and to the possibility of determining the transfer function by means of simple experiments.

As for the existence and uniqueness of the periodic output, *Theorem 25* holds.

THEOREM 25 *(existence and uniqueness of the periodic regime)*

In a continuous-time linear system (A, b, c^T) with no eigenvalues with zero real part, one and only one periodic output of period T, say $y_T(\cdot)$, is associated to each periodic input function $u_T(\cdot)$ of period T. Moreover, if the observable part is asymptotically stable, the output $y(t)$ corresponding to the input $u_T(t)$ tends asymptotically to $y_T(t)$, for any initial state $x(0)$ of the system.

To verify that the theorem cannot be extended to systems with eigenvalues with zero real part (nonhyperbolic systems) it is sufficient to consider the case of an integrator $\dot{x} = u, y = x$. In fact, in such a system, the input

$$u_T(t) = U \cos(2\pi/T)t$$

gives rise to an infinite number of periodic outputs

$$y_T(t) = x(0) + (UT/2\pi)\sin(2\pi/T)t$$

parameterized in the initial state $x(0)$ of the system.

Among all periodic input functions $u_T(\cdot)$, of particular interest is the sinusoidal function

$$U \sin(2\pi/T)t = U \sin \omega t$$

since it is known that, under very general conditions (see Section B.23), any periodic function $u_T(\cdot)$ of period T can be expanded in Fourier series and expressed as an infinite linear combination of sinusoids and cosinusoids of angular frequency $n(2\pi/T)$, n being any nonnegative integer. In fact, a periodic function of period T can be written as

$$u_T(t) = a_0 + \sum_{n=1}^{\infty} \left[a_n \cos\left(n\frac{2\pi t}{T}\right) + b_n \sin\left(n\frac{2\pi t}{T}\right)\right]$$

where

$$a_0 = \frac{1}{T}\int_{-T/2}^{T/2} u_T(t)dt$$

$$a_n = \frac{2}{T}\int_{-T/2}^{T/2} u_T(t)\cos\left(n\frac{2\pi t}{T}\right)dt$$

$$b_n = \frac{2}{T}\int_{-T/2}^{T/2} u_T(t)\sin\left(n\frac{2\pi t}{T}\right)dt$$

Therefore, the periodic function $y_T(\cdot)$ corresponding to the periodic input $u_T(\cdot)$ can be computed by first determining the components of the Fourier series of $u_T(\cdot)$ and then by summing up all the corresponding output periodic functions (superposition principle). Moreover, the relevance of the sinusoidal regime is motivated by the following result:

THEOREM 26 *(frequency response)*

Continuous-time linear systems (A, b, c^T) with no eigenvalues with zero real part have one and only one sinusoidal output

$$y_T(t) = Y \sin\left(\frac{2\pi}{T}t + \varphi\right)$$

for any sinusoidal input

$$u_T(t) = U \sin\frac{2\pi}{T}t$$

Moreover, Y is linear in U and φ depends only on $\omega = \frac{2\pi}{T}$, that is,

$$Y = R(\omega)U \quad \varphi = \varphi(\omega)$$

Theorem 26 states that a system in sinusoidal regime has an output sinusoid of amplitude $R(\omega)U$ shifted with respect to the input sinusoid of an angle $\varphi(\omega)$. Therefore, the computation of the periodic function $y_T(\cdot)$ corresponding to a periodic input $u_T(\cdot)$ is straightforward once the two functions $R(\cdot)$ and $\varphi(\cdot)$ are known. This is why this pair of functions is called *frequency response*.

The frequency response of a system can be measured by means of simple experiments if the observable part of the system is asymptotically stable. In fact, if one applies to a system of this kind a sinusoidal input of amplitude U and angular frequency ω, after a sufficiently long time interval the output of the system is in practice a sinusoid of amplitude $R(\omega)U$ and phase $\varphi(\omega)$, no matter what the initial conditions of the system are. Thus, $R(\omega)$ is simply the ratio of the amplitudes of the output and input sinusoids while $\varphi(\omega)$ is the phase shift between the two sinusoids.

In contrast, if the observable part of the system is not asymptotically stable, it is not possible to experimentally determine the frequency response of the system since the output does not tend toward a sinusoid for a generic initial state. Nevertheless, this fact does not imply that the frequency response $[R(\omega), \varphi(\omega)]$ cannot be defined; indeed, the problem of the definition of a quantity is different from the problem of its determination.

The frequency response of a system can be graphically represented in two ways: by means of the Cartesian plots of the functions $R(\cdot)$ and $\varphi(\cdot)$ or by means of the polar plot (called *Nyquist plot*) representing the function $R(\cdot)e^{i\varphi(\cdot)}$.

Typical Cartesian plots of the function $R(\cdot)$ are depicted in *Fig. B.25*. In certain ranges of the angular frequency ω the function $R(\omega)$ is almost zero, which means that the input sinusoid is very strongly attenuated. Thus, for example, the system in *Fig. B.25(a)* attenuates all the sinusoids with angular frequency $> \omega_0$ while the sinusoids with $\omega < \omega_0$ are not attenuated. A system of this kind is called a *low-pass* filter and the interval $[0, \omega_0]$ is called a *bandwidth* [since it is not possible

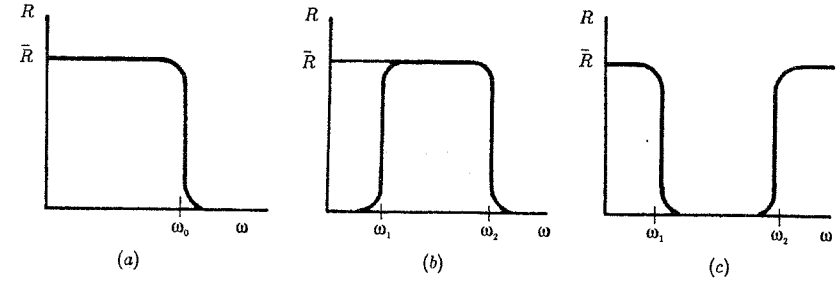

Figure B.25 Typical examples of frequency responses: (*a*) low-pass; (*b*) band-pass; (*c*) stop band.

that $R(\omega) = 1$ for $\omega \leq \omega_0$ and $R(\omega) = 0$ for $\omega > \omega_0$, the bandwidth of the system has to be defined appropriately].

On the other hand, the system with $R(\cdot)$ as in *Fig. B.25(b)*, attenuates the sinusoids with angular frequency $\omega < \omega_1$ and those with $\omega < \omega_2$, and is therefore called a *band-pass filter*. For the opposite reason, the system with $R(\cdot)$ as in *Fig. B.25(c)* is called *stop band*.

Limiting cases of particular relevance can be obtained by letting the bandwidth of a band-pass or stop-band system go to zero. By doing so one obtains systems called, respectively, *resonant filters* and *notch filters*, which are sensitive or insensitive only to a very specific angular frequency.

Obviously, not only the function $R(\cdot)$ is of interest since the function $\varphi(\cdot)$ also contributes to define the sinusoidal regime. In fact, the frequency response $[R(\omega), \varphi(\omega)]$ matters. As an example, consider a communication system which, ideally, should be able to reproduce at the output a perfect copy of the input, obviously with a certain delay τ needed to transfer the information from the input to the output. Thus, a sinusoidal input $U \sin \omega t$ must produce a sinusoid $U \sin \omega(t - \tau)$ at the output, and this must hold for any angular frequency ω. In other words, the ideal communication system is a pure delay system characterized by $R(\omega) = 1$ and $\varphi(\omega) = -\omega\tau$, as shown in *Fig. B.26*.

A pure delay system cannot be realized by a finite-dimensional linear system (A, b, c^T), so that communication systems are often designed by allowing a certain degree of distortion between input and output, that is, by approximating the shape of the functions $R(\cdot)$ and $\varphi(\cdot)$ in *Fig. B.26* in a suitable range of the angular frequency.

So far, we have shown that the frequency response $[R(\cdot), \varphi(\cdot)]$ enables one to rapidly compute the periodic regime of a linear system and its filtering properties. We have also shown how the frequency response can be experimentally measured if the observable part of the system is asymptotically stable. A third very important property is the following connection (*Theorem 27*) between frequency response and transfer function.

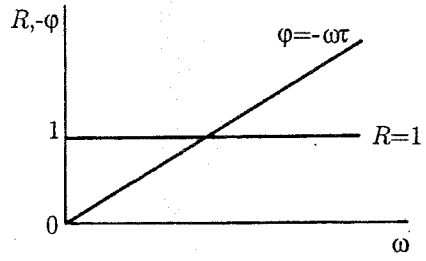

Figure B.26 Frequency response of an ideal communication system.

THEOREM 27 *(frequency response and transfer function)*

The frequency response $[R(\omega), \varphi(\omega)]$ of a continuous-time linear system is uniquely determined by its transfer function $G(s)$. More precisely, $R(\omega)$ and $\varphi(\omega)$ are, respectively, the modulus and the phase of the complex number $G(i\omega)$, that is,

$$G(i\omega) = R(\omega)e^{i\varphi(\omega)} \tag{B.38}$$

The proof of *Theorem 27* can be obtained by noting that the sinusoids $u_T(t) = U\sin\omega t$ and $y_T(t) = R(\omega)U\sin(\omega t+\varphi(\omega))$ with $R(\omega)$ and $\varphi(\omega)$ given by (B.38), satisfy the differential equation (ARMA model)

$$d(s)y(t) = n(s)u(t)$$

where $G(s) = n(s)/d(s)$. For example, if the system is the first-order system $\dot{x} = ax + bu$ with $y = x$, the transfer function is

$$G(s) = \frac{b}{s-a}$$

and the ARMA model is

$$\dot{y} - ay = bu$$

while the frequency response is

$$R(\omega) = \sqrt{\frac{b^2}{a^2+\omega^2}} \quad \varphi(\omega) = \text{arctg}\frac{\omega}{a}$$

It is therefore immediate to check (using a bit of trigonometry) that the two sinusoids

$$u_T(t) = U\sin\omega t$$

$$y_T = R(\omega)U\sin(\omega t + \varphi(\omega))$$

satisfy (B.22).

Theorem 27 is very important, since it allows one to compute the frequency response of a linear system from the triple (A, b, c^T), from the ARMA model, or from the transfer function $G(s)$. Moreover, it is the basis for a relatively simple solution of the identification problem (see *Section B.18*). In fact, if we want to model a physical system we must perform some tests on the system and, on the basis of the results, determine, for example, the triple (A, b, c^T). Such tests, must be measures of pairs of input and output functions, for example, the impulse response or the frequency response. From these functions, we must compute the transfer function of the system and, then, realize the triple (A, b, c^T) (see *Section B.3*). Among the tests that can be performed on the system, the frequency response is perhaps the most convenient. In fact, in order to measure the frequency response of an asymptotically stable system, it is not necessary that the initial state be zero, while this is necessary when dealing with the impulse or the step response of the system. Moreover, the frequency response can be measured by applying at the input of the system sinusoids of relatively small amplitude in such a way that nonlinear effects are negligible. Obviously, this is not the case when one wants to measure the impulse response. Finally, if the long-term response of an asymptotically stable system to a sinusoid is not sinusoidal, it is possible to conclude that the system is nonlinear, a conclusion, that can be hardly obtained by examining the impulse response since it is extremely difficult to check whether a function is a linear combination of exponentials or not.

B.23 FOURIER TRANSFORM

Before introducing the notions of Fourier series and Fourier transform, we give some definitions that will be used in the sequel.

DEFINITION 4 *(functions with bounded variation)*

A real function $f(\cdot)$ has a *bounded variation* in the closed interval $[a, b]$ if there exists a constant K such that for any finite set of points $t_0, t_1, t_2, \ldots, t_n$ partitioning the interval $[a, b](a = t_0 < t_1 < t_2 < \ldots < t_n = b)$ one has

$$\sum_{k=0}^{n-1} |f(t_{k+1}) - f(t_k)| \leq K$$

If a real function $f(\cdot)$ has a bounded variation in any closed interval, we say it is of bounded variation. Moreover, a complex function $f(\cdot)$ is of bounded variation if its real and imaginary parts have bounded variations.

The functions with bounded variation enjoy a number of properties that are now reported without proof.

THEOREM 28 *(properties of functions with bounded variation)*

A real function $f(\cdot)$ has a bounded variation in the interval $[a, b]$ if and only if it is the difference between two nondecreasing functions. A function $f(\cdot)$ with bounded variation in the interval $[a, b]$ is bounded in the same interval. If a function $f(\cdot)$ has bounded variation in an interval $[a, b]$, the discontinuities of the function in such interval are numerable. If a function $f(\cdot)$ has bounded variation in an interval $[a, b]$ then, for any $t \in (a, b)$ there exist the right and left limits, that is,

$$f(t^-) = \lim_{\varepsilon \to 0} f(t - \varepsilon) \quad f(t^+) = \lim_{\varepsilon \to 0} f(t + \varepsilon) \quad \varepsilon > 0$$

Moreover, for $t = a$ there exists the right limit and for $t = b$ the left one.

We can now state the first important result concerning the *Fourier series*. From an intuitive point of view, the result says that, under very general assumptions, a periodic function of period T can be represented as the linear combination of sinusoids of angular frequency equal to multiples of the angular frequency $2\pi/T$. The proof of this result is not reported because it is not easy.

THEOREM 29 *(Fourier series)*

If $f(\cdot)$ is a periodic function of period T with bounded variation, then for all t one has

$$\lim_{N \to \infty} \sum_{k=-N}^{N} f_k e^{i\frac{2\pi k}{T}t} = \frac{1}{2}(f(t^+) + f(t^-)) \tag{B.39}$$

where

$$f_k = \frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-i\frac{2\pi k}{T}t} dt \quad k = 0, \mp 1, \mp 2, \ldots \tag{B.40}$$

Obviously, if the function $f(\cdot)$ is continuous at time t, Eq. (B.39) simplifies and becomes

$$f(t) = \lim_{N \to \infty} \sum_{k=-N}^{N} f_k e^{i\frac{2\pi k}{T}t} \tag{B.41}$$

By recalling that

$$e^{i\theta} = \cos\theta + i \sin\theta$$

From (B.40) and (B.41) one easily obtains

$$f(t) = \frac{1}{2}a_0 + \lim_{N \to \infty} \sum_{k=-N}^{N} \left\{ a_k \cos\left(\frac{2\pi k}{T}t\right) + b_k \sin\left(\frac{2\pi k}{T}t\right) \right\} \tag{B.42}$$

where

$$a_k = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \cos\left(\frac{2\pi k}{T}t\right) dt$$
$$b_k = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \sin\left(\frac{2\pi k}{T}t\right) dt$$

Expression (B.42) is the most popular formulation of the Fourier series, since it shows explicitly that a periodic function $f(\cdot)$ is the linear combination of sinusoids and cosinusoids. Moreover, if

$$\int_{-T/2}^{T/2} |f(t)|^2 dt < \infty$$

it follows that

$$\lim_{N \to \infty} \int_{-T/2}^{T/2} \left| \sum_{k=-N}^{N} f_k e^{i\frac{2\pi k}{T}t} - f(t) \right|^2 dt = 0$$

where f_k is given by (B.40).

Since any function $f(\cdot)$ can be considered as a periodic function of a infinite period, from the previous results it follows that any function $f(\cdot)$ is the sum of a continuum of sinusoids and cosinusoids, since the difference $2\pi/T$ between two different angular velocities tends to zero whenever T tends to infinity. This is the basic idea of the so-called Fourier transform, which is specified below.

Let $f(\cdot)$ be a function with bounded variation over R and suppose that such a function satisfies the inequality

$$\int_{-\infty}^{\infty} |f(t)| dt < \infty$$

Denote with $f_T(\cdot)$ the periodic function of period T that coincides with $f(\cdot)$ in the interval $[-T/2, T/2)$. From the previous results, it follows that $f_T(\cdot)$ can be expanded in Fourier series, namely,

$$\lim_{N \to \infty} \sum_{k=-N}^{N} \left[\frac{1}{T} \int_{-\infty}^{\infty} f_T(t) e^{-i\frac{2\pi k}{T}t} dt \right] e^{i\frac{2\pi k}{T}t} = \frac{1}{2}(f_T(t^+) + f_T(t^-)) \tag{B.43}$$

Since, by definition, $f_T(\cdot)$ and $f(\cdot)$ coincide in the interval $[-T/2, T/2)$, the relationship (B.43) can also be written with $f(t)$ instead of $f_T(t)$ provided t belongs to the interval $[-T/2, T/2)$. By setting

$$F_T(i\omega) = \int_{-T/2}^{T/2} f(t) e^{i\omega t} dt$$

from (B.40) with $T \to \infty$ one obtains

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} F(i\omega)e^{i\omega t}d\omega = \frac{1}{2}(f(t^+) + f(t^-)) \tag{B.44}$$

and

$$F(i\omega) = \int_{-\infty}^{\infty} f(t)e^{i\omega t}dt = \lim_{T\to\infty} F_T(i\omega)$$

The function $F(\cdot)$ is called a *Fourier transform* or Fourier integral of the function $f(\cdot)$.

B.24 LAPLACE TRANSFORM

Suppose that a function $f(\cdot)$ has bounded variation in any closed interval contained in $[0, \infty)$ and that there exists a constant $\sigma < \infty$ such that

$$\int_{-\infty}^{\infty} |f(t)|e^{-\sigma t}dt < \infty$$

Then, consider, the following function $F(\cdot)$

$$F(\sigma + i\omega) = \int_0^{\infty} f(t)e^{-i\omega t}e^{-\sigma t}dt \tag{B.45}$$

and note that

$$F(\sigma + i\omega) = \int_0^{\infty} e^{-i\omega t}(\text{step}(t)e^{-\sigma t}f(t))dt$$

which means that the function $F(\cdot)$ is the Fourier transform of the function

$$\text{step}(t)e^{-\sigma t}f(t)$$

Therefore, from (B.44) it follows that

$$\frac{1}{2}f(0^+) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\sigma + i\omega)d\omega$$

and

$$\frac{1}{2}e^{-\sigma t}(f(t^+) + f(t^-)) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\sigma + i\omega)e^{-i\omega t}d\omega$$

for $t > 0$. If we denote the complex variable by s (*i.e.*, $s = \sigma + i\omega$) Eq. (B.45) becomes

$$F(s) = \int_0^{\infty} e^{-st}f(t)dt \tag{B.46}$$

The function $F(\cdot)$, often denoted by $L[f(\cdot)]$, is called a *Laplace transform* of the function $f(\cdot)$. It is a complex valued function defined on the domain $\text{Re}(s) > \sigma_0$, where σ_0 is the smallest real number such that $\sigma < \sigma_0$ implies

$$\int_{-\infty}^{\infty} |f(t)|e^{-\sigma t}dt < \infty$$

The Laplace transformation $f(\cdot) \mapsto F(\cdot)$, defined by (B.46), has a number of properties. First of all, it is a linear transformation since

$$L[\alpha f_1(\cdot) + \beta f_2(\cdot)] = \alpha L[f_1(\cdot)] + \beta L[f_2(\cdot)]$$

Moreover, the Laplace transform $F(\cdot)$ of any function $f(\cdot)$ is an analytical function in the domain $\text{Re}(s) > \sigma_0$. This implies that the function $F(\cdot)$ can often be extended to the whole complex plane [*i.e.*, there exists a unique function coinciding with $F(\cdot)$ for $\text{Re}(s) > \sigma_0$ but defined over the whole complex plane and analytical anywhere, apart from a certain number of isolated singularity points]. For example, if $f(t) = e^t$, $0 \le t < \infty$ we have

$$L[f(\cdot)] = \int_0^{\infty} e^t e^{-st}dt = \frac{1}{s-1} \quad \text{Re}(s) > 1$$

and the function $1/(s-1)$ is analytical anywhere apart from the singular point $s = 1$.

Other important properties of the Laplace transform are those concerning integration and differentiation of a function $f(\cdot)$. In fact, the following relations hold:

$$L\left[\int_0^t f(\tau)d\tau\right] = \frac{1}{s}L[f(\cdot)]$$
$$L\left[\frac{d}{dt}f(\cdot)\right] = sL[f(\cdot)] - f(0)$$

Finally, the product of two transformed functions corresponds, in the time domain, to the operation called *convolution*, that is, if $F(\cdot)$ and $G(\cdot)$ are the Laplace transforms of two functions $f(\cdot)$ and $g(\cdot)$, the inverse transform of

$$H(\cdot) = F(\cdot)G(\cdot)$$

is

$$h(t) = \int_0^t f(t-\tau)g(\tau)d\tau \quad 0 \le t < \infty$$

In the following table, we report some Laplace transforms $F(s)$ of functions $f(t)$:

$f(t)$	$F(s)$
impt	1
stept	$\frac{1}{s}$
rampt	$\frac{1}{s^2}$
$e^{\alpha t}$	$\frac{1}{s-\alpha}$
$\sin \omega t$	$\frac{\omega}{s^2+\omega^2}$
$\cos \omega t$	$\frac{s}{s^2+\omega^2}$
$f(t-\tau)$	$e^{-\tau s}F(s)$
$t^n \quad n>0$	$\frac{n!}{s^{n+1}}$

B.25 Z-TRANSFORM

Consider a function $f(\cdot)$ defined on the nonnegative integers, that is,

$$f(\cdot): t \mapsto f(t) \quad t = \text{nonnegative integer}$$

The *Zeta-transform* of such a function, denoted by

$$F(\cdot) = Z[f(\cdot)]$$

is simply given by the series

$$F(\cdot): z \mapsto F(z) = f(0) + f(1)z^{-1} + f(2)z^{-2} + \cdots \tag{B.47}$$

Obviously, this expression makes sense if the series converges in the neighborhood of the improper point $z^{-1} = 0$, where it clearly converges. Suppose now that $f(t)$ does not increase with t more rapidly than a geometric series. Then, $|f(t)|^{1/t}$ tends to one or more positive limits, the largest being denoted by R_c, that is,

$$\lim_{t\to\infty} |f(t)|^{1/t} = R_c$$

It is easy to show that the series

$$F(z) = \sum_{i=0}^{\infty} f(t)z^{-t}$$

converges absolutely for all complex z satisfying the relationship

$$|z| > R_c$$

and, for this reason, R_c is called the *convergence radius*.

The transformation operator $f(\cdot) \mapsto F(\cdot)$ is obviously linear, since

$$Z[\alpha f_1(\cdot) + \beta f_2(\cdot)] = \alpha Z[f_1(\cdot)] + \beta Z[f_2(\cdot)]$$

Theorems completely analogous to those given for the Laplace transform can be proved for the Z-transform. By denoting $f^-(\cdot)$ as the function obtained from $f(\cdot)$ after a backward time shift, that is,

$$f^-(t) = \begin{cases} 0 & \text{for } t = 0 \\ f(t-1) & \text{for } t \geq 1 \end{cases}$$

the following holds:

$$Z[f^-(\cdot)] = z^{-1}Z[f(\cdot)]$$

while, if $f^+(\cdot)$ is the function obtained from $f(\cdot)$ after a forward time shift, that is,

$$f^+(t) = f(t+1) \quad \text{for } t \geq 0$$

the following holds:

$$Z[f^+(\cdot)] = zZ[f(\cdot)] - zf(0)$$

The simplest way to find analytical expressions of the Z-transform is to determine the sum of the series (B.47). Thus, for example, if

$$f(t) = a^t \quad t \geq 0$$

we have

$$F(z) = \sum_{t=0}^{\infty} a^t z^{-t} = \sum_{t=0}^{\infty} (az^{-1})^t = (1 - az^{-1})^{-1}$$

for $|z| > |a|$ and the same formula holds if a is replaced by a square matrix A and 1 by the identity matrix I. Other Z-transforms are reported in the table below.

$f(t)$	$F(z)$
a^t	$\frac{z}{z-a}$
1	$\frac{z}{z-1}$
t	$\frac{z}{(z-1)^2}$
t^2	$\frac{z(z+1)}{(z-1)^3}$
t^3	$\frac{z(z^2+4z+1)}{(z-1)^4}$

B.26 LAPLACE AND Z–TRANSFORMS AND TRANSFER FUNCTIONS

Recall that the transfer function $G(p)$ of a linear system described by an ARMA model

$$D(p)y(t) = N(p)u(t)$$

has been defined (see Section B.2) as the ratio of the two polynomials $N(p)$ and $D(p)$, that is,

$$G(p) = \frac{N(p)}{D(p)}$$

Moreover, if the system is proper, the transfer function can be computed using the formula [see (B.13)]

$$G(p) = c^T(pI - A)^{-1}b$$

Thus, the transfer function $G(p)$ is the (Laplace and Z-) transform of the impulse response. In fact, in a continuous-time system, the Laplace transform of the impulse response is

$$L[g(t)] = L[c^T e^{At}b] = c^T L[e^{At}]b = c^T(sI - A)^{-1}b$$

and therefore coincides with (B.13). An analogous check is possible for discrete-time systems.

If we take the above discussion into account, it is straightforward to see that the transfer function can also be written in the form

$$G(p) = \sum_{i=1}^{\infty} \frac{g_i}{p^i}$$

where $g_i = c^T A^{i-1}b$ are the Markov coefficients. In fact, for discrete-time systems (see Section B.21) we have

$$g(t) = \begin{cases} 0 & \text{for } t = 0 \\ g_t & \text{for } t \geq 1 \end{cases}$$

so that the Z-transform of $g(t)$ is (see Section B.25)

$$G(z) = g_1 z^{-1} + g_2 z^{-2} + g_3 z^{-3} + \cdots$$

In an analogous way, for continuous-time systems

$$g(t) = g_1 + g_2 t + g_3 \frac{t^2}{2!} + \cdots$$

so that, recalling that the Laplace transform of t^n is $(n!/s^{n+1})$, one obtains

$$G(s) = L[g(t)] = \frac{g_1}{s} + \frac{g_2}{s^2} + \frac{g_3}{s^3} + \cdots$$