13.0 INTRODUCTION

In this chapter we review several important concepts from the theory of multi-input multi-output linear time invariant systems (abbreviated MIMO LTI systems). MIMO systems are sometimes also called multivariable systems, and should not be confused with multidimensional systems discussed in Chap. 12. There are excellent texts on this topic, for example, Kailath [1980] and Chen [1970 and 1984]. Other references for this chapter include Desoer and Schulman [1974], Gantmacher [1959], Rosenbrock [1970], and Vidyasagar [1985]. While many of these references emphasize continuous-time systems, our presentation here is entirely for the discrete-time case. Some of the results, for example, those in Sec. 13.9 and 13.10 have not appeared in text books before.

Because of the "reference" nature of this chapter, the writing style is unlike any other chapter in this text. We have chosen to present various facts and results in the form of lemmas and theorems. This has enabled us to review a large number of advanced results in an economic manner, while at the same time ensuring completeness. In this chapter we will frequently use matrix theory results, particularly concepts such as rank, nonsingularity,
orthogonal complements, and so on (summarized in Appendix A).

13.1 MULTI-INPUT MULTI-OUTPUT SYSTEMS

In Sec. 2.2 we introduced $p \times r$ LTI systems, that is, LTI systems with $r$ inputs and $p$ outputs. The $k$th output in response to all the inputs is given by (2.2.1), where $H_{km}(z)$ is the transfer function from the input $u_m(n)$ to the output $y_k(n)$. The MIMO system is characterized by the $p \times r$ transfer matrix $\mathbf{H}(z)$ or, equivalently, by the impulse response matrix $\mathbf{h}(n)$. In Fig. 2.2-2 we indicated two ways of representing the system. The output sequence $y(n)$ is related to the input $u(n)$ in terms of the matrix convolution (2.2.7) (equivalently (2.2.4) in the $z$-domain). We request the reader to review Sec. 2.2, because the same notations will be used here.

A system with $p = r = 1$ is said to be a single input single output (SISO) system, or a scalar system. Many of the concepts which follow easily in the SISO case turn out to be complicated for MIMO systems because matrices are involved. Development of basic ideas such as irreducible rational functions, minimal realizations, and transmission zeros require more careful thought in the MIMO case. In this chapter we will review many of these concepts.

Outline

Section 13.2 introduces matrix polynomials and their properties. In Sec. 13.3 we study the matrix fraction description which is useful for describing MIMO transfer functions. Section 13.4 gives a detailed exposure to state space description of MIMO LTI systems. The Smith-McMillan form is introduced in Sec. 13.5. Sections 13.6 and 13.7 deal with poles and zeros of MIMO LTI systems. The degree or McMillan degree of a system is studied in Sec. 13.8. Finally Sec. 13.9 and 13.10 present some results for the case of FIR MIMO systems, which are useful in filter bank research.

Polynomials and integers. While many of the results of this chapter are developed for polynomials and polynomial matrices, they hold equally well for integers and integer matrices. This is because, the set of polynomials as well as the set of integers belong to a common algebraic structure called the principal ideal domain [Forney, 1970], [Vidyasagar, 1985]. In Sec. 12.10.1 we have already applied one of these results (e.g., Smith decomposition of integer matrices) for the implementation of multidimensional decimators. Further applications (to multidimensional multirate systems) of the integer counterpart of the results of this chapter can be found in Chen and Vaidyanathan [1992a,b], [1993], [Gopinath and Burrus, 1991], [Evans, et al., 1992], and [Kalker, 1992].

13.2 MATRIX POLYNOMIALS

Matrix polynomials (or polynomial matrices) play a very crucial role in the description and understanding of multivariable systems. Both FIR and IIR
systems can be represented in terms of matrix polynomials. In this section we study their properties.

A \( p \times r \) polynomial matrix \( P(z) \) in the variable \( z \) is simply a \( p \times r \) matrix whose entries are polynomials in \( z \). The matrix can be expressed as

\[
P(z) = \sum_{n=0}^{K} p(n)z^n. \tag{13.2.1a}
\]

If \( p(K) \neq 0 \) the quantity \( K \) is called the order of the polynomial (not to be confused with degree which has a subtler meaning; Sec. 13.3, 13.8). When there is no room for confusion, we use the term ‘polynomial’ rather than ‘polynomial matrix’.

A polynomial in \( z^{-1} \) with order \( K \), that is, a system of the form

\[
H(z) = \sum_{n=0}^{K} h(n)z^{-n}, \tag{13.2.1b}
\]

is said to be a causal FIR system for obvious reasons. The system in (2.2.9) is such an example.

**Rank and “Normal Rank” of a Polynomial Matrix**

Given a polynomial matrix \( P(z) \), it is clear that the rank usually depends on the value of \( z \). In any case the rank \( \rho(z) \) cannot exceed \( \min(p, r) \). As an example, for the system given by

\[
\begin{bmatrix}
1 + z & 1 + 2z \\
2 + 2z & 2 + 4z
\end{bmatrix}, \tag{13.2.2}
\]

it is clear that \( \rho(z) < 2 \) for all \( z \) because the second row is proportional to the first for all \( z \). On the other hand, the matrix

\[
\begin{bmatrix}
2 + z & 1 + z \\
1 + z & 2 + z
\end{bmatrix} \tag{13.2.3}
\]

has \( \rho(z) = 2 \) for all \( z \) except \( z = -3/2 \). To see this note that the determinant is equal to \( 3 + 2z \) which is nonzero (so that the matrix has full rank) unless \( z = -3/2 \). For the example in (13.2.2), on the other hand, the determinant is identically zero so that \( \rho(z) < 2 \) for all \( z \).

**Normal rank.** Let \( \rho(z) \) denote the rank of a \( p \times r \) polynomial at \( z \). Then the normal rank is defined to be the maximum value of \( \rho(z) \) in the entire \( z \) plane. Matrices for which the normal rank is the maximum possible \( \leq \min(p, r) \) are said to have full normal rank.

Thus (13.2.3) has full normal rank whereas (13.2.2) does not. For the special case of a \( M \times M \) (i.e., square) matrix \( P(z) \), the normal rank is full
(i.e., equal to $M$) if and only if the determinant (which is a polynomial in $z$) is not identically zero for all $z$. In this case the rank is less than $M$ only at a finite number of points in the $z$ plane (viz., at the locations of the zeros of $|\det P(z)|$).

**Unimodular Matrices**

A unimodular matrix $U(z)$ in the variable $z$ is a square polynomial matrix in $z$ with constant nonzero determinant. Here are some examples:

\[
\begin{bmatrix}
1 & 0 \\
1 + z^3 & 1 \\
\end{bmatrix}, \quad \begin{bmatrix}
1 + z^2 & z \\
2z & 2 \\
\end{bmatrix}.
\]

It is readily verified that the product of unimodular matrices is unimodular. Using this, one can generate more complicated examples.

For $M \times M$ unimodular $U(z)$ the normal rank is $M$. In fact the rank $\rho(z) = M$ for all $z$. Moreover, since

\[U^{-1}(z) = \frac{\text{Adj } U(z)}{\det U(z)} \tag{13.2.4}\]

(where 'Adj' denotes the adjugate; Appendix A) the inverse is also a polynomial in $z$. Since $U(z)U^{-1}(z) = I$, the determinant of the inverse is a (nonzero) constant. So the inverse of unimodular $U(z)$ is again a unimodular polynomial matrix.

Conversely, if $U(z)$ is a polynomial matrix with polynomial inverse $V(z)$ then $U(z)V(z) = I$ so that $[\det U(z)][\det V(z)] = 1$. Since $[\det U(z)]$ and $[\det V(z)]$ are polynomials, this means that these determinants are constants so that $U(z)$ is unimodular. Summarizing, a polynomial matrix $U(z)$ is unimodular if and only if its inverse is a polynomial matrix.

**Common Factors, Glcd, and Coprimeness**

Two scalar polynomials $P(x)$ and $Q(x)$ are said to be coprime (or relatively prime) if they do not have common factors (i.e., common divisors) with order $\geq 1$. For example $(1 + x)$ and $(2 + x)(3 + x)$ are coprime, whereas $(2 + x)$ and $(2 + x)(3 + x)$ are not. A scalar transfer function $H(z) = P(z)/Q(z)$ is said to be in irreducible form if the polynomials $P(z)$ and $Q(z)$ are relatively prime.

In the next section we will express rational transfer matrices in irreducible form. This is algebraically more complicated because matrices are involved. We will require the concept of coprimeness of matrix polynomials; and we have to distinguish between left divisors and right divisors. Given a $p \times r$ polynomial $Q(z)$ we say that the $p \times p$ polynomial $L(z)$ is a left divisor (or factor) of $Q(z)$ if

\[Q(z) = L(z)Q_1(z) \tag{13.2.5}\]

for some polynomial $Q_1(z)$. Suppose $L(z)$ is a left divisor of two polynomials $P(z)$ and $Q(z)$ (with the same number of rows $p$). That is

\[Q(z) = L(z)Q_1(z), \quad P(z) = L(z)P_1(z), \tag{13.2.6}\]
for some polynomials $Q_1(z)$ and $P_1(z)$. Then $L(z)$ is said to be a left common divisor (abbreviated lcd) of $Q(z)$ and $P(z)$. For example, let

$$P(z) = \begin{bmatrix} \frac{1}{z + z^2} & -1 + z + z^2 \\ \frac{1}{z + z^2} & \frac{2 + z}{z} \end{bmatrix}, \quad \text{and} \quad Q(z) = \begin{bmatrix} \frac{z}{2z} & 2 + z \\ \frac{0}{2z} & 0 \end{bmatrix}. \quad (13.2.7)$$

We can rewrite these as

$$P(z) = \begin{bmatrix} 1 + z & -1 \\ z & 1 \end{bmatrix}^{-1}_L \begin{bmatrix} z & 1 \\ z & 1 \end{bmatrix}, \quad Q(z) = \begin{bmatrix} 1 + z & -1 \\ z & 1 \end{bmatrix}^{-1}_L \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad (13.2.8)$$

so that $L(z)$ is an lcd.

Notice that any $p \times p$ unimodular matrix $U(z)$ is an lcd of $P(z)$ and $Q(z)$ because $U^{-1}(z)$ is polynomial and we can write

$$P(z) = U(z)U^{-1}(z)P(z), \quad Q(z) = U(z)U^{-1}(z)Q(z). \quad (13.2.9)$$

**The greatest left common divisor (glcd).** Let the polynomial $L(z)$ be an lcd of the two polynomials $P(z)$ and $Q(z)$. From (13.2.6) we see that every left factor of $L(z)$ is also an lcd. We say that $L(z)$ is a greatest lcd (abbreviated glcd) if every other lcd is a left factor of $L(z)$, that is, if $L_1(z)$ is any lcd then

$$L(z) = L_1(z)R_1(z) \quad (13.2.10)$$

for some polynomial $R_1(z)$. So every lcd is a left factor of the glcd, and conversely every left factor of the glcd is an lcd, as one can readily verify. (In a similar way one can define right common divisors (rcd), and grcd. We will skip details, as we do not intend to use them.)

**Glcd is not unique.** The glcd of two polynomial matrices is not unique. For example if $L(z)$ is a glcd of $Q(z)$ and $P(z)$, then so is $L(z)W(z)$ for unimodular $W(z)$ (Problem 13.2).

**Coprimeness.** Two polynomial matrices $Q(z)$ and $P(z)$ are said to be left coprime if every lcd (hence the glcd) is unimodular. For the scalar case this is equivalent to saying that there are no common factors of order $> 0$. The matrices $P(z)$ and $Q(z)$ in the example (13.2.7) are not left coprime because there is an lcd $L(z)$ [shown in (13.2.8)], which is not unimodular.

From the above definitions and the properties of unimodular matrices, one can obtain the following two results which we use later:

**Fact 13.2.1.** Two polynomial matrices $P(z)$ and $Q(z)$ are left coprime if and only if $P(z)W(z)$ and $Q(z)V(z)$ are left coprime for every pair of unimodular $W(z)$ and $V(z)$.

664 Chap. 13. Review of MIMO LTI systems
\textbf{Proof.} See Problem 13.2.

\begin{fact}
\textbf{13.2.2. Inverses of square polynomial matrices.} Let \(P(z)\) be a \(p \times p\) polynomial matrix in \(z\) with normal rank \(p\). Let \(\alpha\) be a zero of its determinant, i.e., \((z - \alpha)\) a factor of \(|\det P(z)|\). Then \(P^{-1}(z)\) has a ‘pole’ at \(z = \alpha\). In other words, at least one element of \(P^{-1}(z)\) has the factor \((z - \alpha)\) in its denominator.
\end{fact}

\textbf{Proof.} We know

\[
P^{-1}(z) = \frac{\text{Adj} P(z)}{|P(z)|} = \frac{\text{Adj} P(z)}{(z - \alpha)^L B(z)},
\]

where \(L\) is a positive integer and \(B(z)\) is a polynomial with \(B(\alpha) \neq 0\). It is sufficient to prove that the factor \((z - \alpha)^L\) in the denominator is not completely canceled by the adjugate. We can rewrite (13.2.11) as

\[
P(z) \text{Adj} P(z) = (z - \alpha)^L B(z) I.
\]

Suppose the adjugate completely cancels the factor \((z - \alpha)^L\). After such cancelation, the determinant of the left hand side above is still zero for \(z = \alpha\) [because of \(P(z)\)], but that of the right hand side is nonzero. This is evidently a contradiction!

\begin{section}{13.3 Matrix fraction descriptions}

We know that any scalar rational transfer function \(H(z)\) can be written as \(P(z)/Q(z)\), where \(P(z)\) and \(Q(z)\) are polynomials in \(z\). Similarly, a causal \(p \times r\) transfer matrix \(H(z)\) with rational entries \(H_{km}(z)\) (i.e., a causal rational system) can be written as

\[
\underbrace{H(z)}_{p \times r} = \underbrace{Q^{-1}(z)}_{p \times p} \underbrace{P(z)}_{p \times r},
\]

where \(Q(z)\) and \(P(z)\) are matrix polynomials in \(z\) with indicated sizes, and \(Q(z)\) has full normal rank \((= p)\). This is called a left matrix fraction description (abbreviated as \textit{left MFD}). Such a description helps to study the properties of the system more compactly, and also gives rise to implementations which are more efficient than directly implementing each element \(H_{km}(z)\) of \(H(z)\).

The simplest way to obtain the MFD would be as follows: (a) write each element \(H_{km}(z)\) in the form \(H_{km}(z) = A_{km}(z)/B_{km}(z)\) where \(A_{km}(z)\) and \(B_{km}(z)\) are polynomials in \(z\) (and \textit{not} in \(z^{-1}\), which is more standard in digital filtering theory), (b) compute the least common multiple (LCM) \(D(z)\) of the denominators \(B_{km}(z)\), (c) re-express \(H_{km}(z)\) as \(H_{km}(z) = P_{km}(z)/D(z)\) (which in general is not in irreducible form), and (d) define \(P(z) = [P_{km}(z)]\), \(Q(z) = D(z) I\).

\end{section}
In a similar way, a right MFD is defined to be of the form $P_1(z)Q_i^{-1}(z)$. In this chapter we will not have much occasion to use this form. Unless mentioned otherwise, the term MFD stands for left MFD.

**Example 13.3.1**

Suppose

$$H(z) = \begin{bmatrix} 1 \\ z + a \\ \frac{z}{z + a} \end{bmatrix}.$$  \hspace{1cm} (13.3.2)

This can be rewritten as

$$H(z) = \frac{1}{z + a} \begin{bmatrix} z + a \\ z \end{bmatrix} = \frac{1}{z + a} \begin{bmatrix} z + a \\ Q^{-1}(z) \end{bmatrix} = \frac{1}{Q^{-1}(z)} \begin{bmatrix} z + a \\ P(z) \end{bmatrix}. \hspace{1cm} (13.3.3)$$

which is a valid MFD. It can be verified that the following choice:

$$Q(z) = \frac{1}{a} \begin{bmatrix} z + a & -(z + a) \\ -z & z + a \end{bmatrix}, \quad P(z) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \hspace{1cm} (13.3.4)$$

results in a second possible MFD for this system.

The matrices $Q(z)$ and $P(z)$ need not be unique (i.e., the MFD is not unique) as the above example shows. For every $H(z)$ we can always find a so-called irreducible MFD, as we will describe later. Broadly speaking, this is analogous to scalar systems expressed in the form $P(z)/Q(z)$ where $P(z)$ and $Q(z)$ have no common factors. Irreducible MFDs, again, are not unique.

**Degree or McMillan Degree of a System**

A very fundamental concept in the study of MIMO systems is the degree $\mu$ of a system (also called the *McMillan degree*). The degree $\mu$ of a $p \times r$ causal rational system $H(z)$ is the minimum number of delay units (i.e., $z^{-1}$ elements) required to implement $H(z)$. We often use “deg” as an abbreviation for degree (as in $[\deg H(z)]$). The degree is not defined for noncausal systems because they cannot be implemented with delays alone.

For scalar (i.e., SISO) systems the degree is an easily understood concept. Thus consider an $N$th order causal FIR filter $\sum_{n=0}^{N} h(n)z^{-n}$, with $h(N) \neq 0$. The degree is then $N$. Similarly an $N$th order IIR filter (Chap. 2) has degree $N$. For MIMO systems, on the other hand, the situation is more complicated. Thus, let

$$H(z) = \sum_{n=0}^{N} h(n)z^{-n},$$

666 Chap. 13. Review of MIMO LTI systems
with \( h(N) \neq 0 \). The order is \( N \) (which is merely the highest power appearing in the expression for \( H(z) \)). But the degree is in general \( \geq N \). For example consider

\[
H(z) = z^{-1}I = \begin{bmatrix} z^{-1} & 0 \\ 0 & z^{-1} \end{bmatrix}.
\] (13.3.5a)

It is clear that this requires at least two delays for its implementation (Fig. 13.3-1) so that the degree is two, whereas the order \( N = 1 \). A more advanced tool is therefore necessary to tell what the degree is. We will deal with this issue in Sec. 13.8.

**Example 13.3.2**

Consider the FIR system

\[
H(z) = z^{-1} \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}.
\] (13.3.5b)

We can rewrite \( H(z) \) as

\[
H(z) = z^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} [1 & 2] = \begin{bmatrix} 1 \\ 1 \end{bmatrix} z^{-1} [1 & 2].
\]

So the system can be implemented as in Fig. 13.3-2 with only one delay, proving that the degree is unity. More generally, consider the example \( H(z) = Rx^{-1} \) where \( R \) is \( M \times M \) with rank \( \rho \). This means that we can write

\[
R = T \frac{S}{M \times \rho \rho \times M}
\] (13.3.6)

so that \( H(z) = z^{-1}R = T[z^{-1}I]S \). This shows that we can implement the system using \( \rho \) delays (Fig. 13.3-3). So the system \( H(z) \) has degree \( \leq \rho \).

**Figure 13.3-1** Implementation of \( z^{-1}I_2 \).

**Figure 13.3-2** The system in Example 13.3.2.
It turns out that the degree of $H(z)$ above is precisely equal to the rank $\rho$ (Problem 13.8). This is readily verified to be true for the examples (13.3.5a) and (13.3.5b).

![Diagram]

**Figure 13.3-3** Pertaining to Example 13.3.2.

Irreducible MFD, and Order of $[\det Q(z)]$

For SISO systems the expression $H(z) = A(z)/B(z)$ is said to be irreducible if $A(z)$ and $B(z)$ are coprime. For the matrix case if $L(z)$ is a glcd of $Q(z)$ and $P(z)$, then we can obtain a MFD $H(z) = Q_1^{-1}(z)P_1(z)$ where $Q_1(z)$ and $P_1(z)$ are as in (13.2.6). Having done so, the matrices $Q_1(z)$ and $P_1(z)$ are left coprime, and the MFD $Q_1^{-1}(z)P_1(z)$ is said to be irreducible.

Now

$$\det Q(z) = [\det L(z)]\det Q_1(z). \quad (13.3.7)$$

If $P(z)$ and $Q(z)$ are not left coprime, then $L(z)$ is not unimodular so that the order of $[\det Q(z)]$ exceeds that of $[\det Q_1(z)]$. Thus given a reducible MFD, we can always find an irreducible MFD for which the determinant of $Q_1(z)$ has reduced order.

In Sec. 13.3.5.2 we show that all irreducible MFDs of a given system $H(z)$ have same order for $[\det Q(z)]$. From above discussions it follows that this is also the smallest order of $[\det Q(z)]$, among all possible MFDs.

**Order versus degree.** It is worth emphasizing the distinction between order and degree. The function $1 + z^{-1}$ is a polynomial in $z^{-1}$ with order $= 1$. Moreover it can also be viewed as a causal system with degree $= 1$, because it can be implemented with one delay. The function $1 + z$ is a polynomial in $z$ with order $= 1$, but we cannot say that its degree $= 1$. This is because this system cannot be implemented with one delay element! It is noncausal, and its degree is simply undefined! For this reason, we have used
the term "order" rather than "degree" when discussing the determinant of $Q(z)$ which is a polynomial in the advance operator $z$.

13.4 STATE SPACE DESCRIPTIONS

We know that a transfer function $H(z)$ can be implemented in many ways such as the direct form, cascade form and so on. Given any such implementation, the outputs of the delay elements are called state variables, and the system behavior can be completely described in the time domain by a set of equations which involve the input, state variables and output. This is called the state space description of the structure.

To demonstrate state space descriptions, consider the example of a direct form structure shown in Fig. 13.4-1. This is a causal system with transfer function $H(z) = P(z)/Q(z)$, where

$$P(z) = \sum_{n=0}^{N} p_n z^{-n}, \quad Q(z) = 1 + \sum_{n=1}^{N} q_n z^{-n}. \quad (13.4.1)$$

![Figure 13.4-1](image)

Figure 13.4-1 The direct form structure, with state variables indicated.

Since there are $N$ delays, we have $N$ state variables $x_1(n), \ldots, x_N(n)$ as shown. The input and output signals are $u(n)$ and $y(n)$, respectively. The
state space description is

\begin{align*}
x_1(n+1) &= x_2(n) \\
x_2(n+1) &= x_3(n) \\
& \\
x_N(n+1) &= -\sum_{k=1}^{N} q_{N+1-k} x_k(n) + u(n) ,
\end{align*}

(13.4.2)

\[ y(n) = \sum_{k=1}^{N} (p_{N+1-k} - p_0 q_{N+1-k}) x_k(n) + p_0 u(n) . \]

This can be expressed compactly in matrix-vector notation as

\[ x(n+1) = A x(n) + B u(n), \quad \text{(state equation),} \quad (13.4.3) \]

\[ y(n) = C x(n) + D u(n), \quad \text{(output equation),} \quad (13.4.4) \]

where

\[
A = \begin{bmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
-q_N & -q_{N-1} & \cdots & \cdots & -q_1
\end{bmatrix},
\]

(13.4.5)

\[ B = [0 \ 0 \ \cdots \ 0 \ 1]^T , \]

\[ C = [p_N - p_0 q_N \ \ p_{N-1} - p_0 q_{N-1} \ \ \cdots \ \ p_1 - p_0 q_1 ] \]

(13.4.6)

\[ D = p_0 , \]

\[ x(n) = [x_1(n) \ x_2(n) \ \cdots \ x_N(n)] , \]

(13.4.7)

\[ y(n) = y(n) , \ \text{and} \ u(n) = u(n) . \] We have used bold-faced \( u(n) \) and \( y(n) \) because we will use these equations for the MIMO case also.

We can write down the state space description in the above manner for any structure representing a \( p \times r \) causal LTI system \( \textbf{H}(z) \). (See examples below.) With \( N \) denoting the number of delays, the state equation (13.4.3) is a set of \( N \) equations. The output equation is a set of \( p \) equations. The sizes of the matrices are:

\[ A : N \times N ; \quad B : N \times r ; \quad C : p \times N ; \quad D : p \times r . \]

(13.4.8)

Figure 13.4-2 shows a schematic of the state space description.
Main Points About State Space Descriptions

1. **State variables, state vector and state transition matrix.** The outputs \( x_i(n) \) of the delay elements are the state variables, and the vector \( x(n) \) of state variables is called the state vector (or just the 'state' at time \( n \)). The matrix \( A \) is called the state transition matrix (STM). One can verify that if the state \( x(n_0) \) is given then the quantities \( x(n), n > n_0 \) and \( y(n), n \geq n_0 \) can be found by knowing \( u(m) \) for \( m \geq n_0 \). [See (13.4.11) later.]

2. **State space description.** The state equations represent the storage part (or the part of the system having the memory elements \( z^{-1} \)), and can be considered to be the recursive part, since \( x(n+1) \) is computed from \( x(n) \). (This does not necessarily imply the existence of feedback; see FIR example below). The output equation is the nonrecursive part. The quadruple \( (A, B, C, D) \) is said to be the state space description of the structure.

3. **Meaning of \( D \).** Notice from Fig. 13.4-2 that the quantity \( D \) is given by setting \( z = \infty \), that is, \( D = H(\infty) \). This means that we can find the value of \( D \) by wiping out the delay elements in the structure (i.e., set \( z^{-1} = 0 \)). Since \( H(\infty) = h(0) \) for a causal system, we conclude

\[
D = H(\infty) = h(0). \tag{13.4.9}
\]

Essentially \( D \) represents the 'direct path' (i.e., delay-free path) from the input to the output.

4. **Implicit causality.** It should be borne in mind that (13.4.3) and (13.4.4) are implicitly restricted to causal systems. The memory part [eqn. (13.4.3)] is clearly executed in a causal manner. The output at time \( n \) is completely determined by the input \( u(m), m \leq n \) (equivalently by the input \( u(m) \) for \( n_0 \leq m \leq n \) and the state vector \( x(n_0) \) for arbitrary \( n_0 \)). Unless mentioned otherwise, all results in this chapter are restricted to causal systems.
Example 13.4.1: An FIR System

Consider the FIR system of Fig. 13.4-3 with two inputs, two outputs and two state variables. The state equations are easily verified to be

\[
\begin{bmatrix}
    x_1(n + 1) \\
    x_2(n + 1)
\end{bmatrix}_{x(n+1)} =
\begin{bmatrix}
    0 & 0 \\
    1 & 0
\end{bmatrix}
\begin{bmatrix}
    x_1(n) \\
    x_2(n)
\end{bmatrix}_{x(n)} +
\begin{bmatrix}
    0 & 1 \\
    k_1 & 0
\end{bmatrix}
\begin{bmatrix}
    u_0(n) \\
    u_1(n)
\end{bmatrix}_{u(n)},
\]

and the output equations are

\[
\begin{bmatrix}
    y_0(n) \\
    y_1(n)
\end{bmatrix}_{y(n)} =
\begin{bmatrix}
    k_1 & k_2 \\
    k_1k_2 & 1
\end{bmatrix}
\begin{bmatrix}
    x_1(n) \\
    x_2(n)
\end{bmatrix}_{x(n)} +
\begin{bmatrix}
    1 & 0 \\
    k_2 & 0
\end{bmatrix}
\begin{bmatrix}
    u_0(n) \\
    u_1(n)
\end{bmatrix}_{u(n)}.
\]

![FIR example for state space description.](image)

So the state space description of the structure is

\[
A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ k_1 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} k_1 & k_2 \\ k_1k_2 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 \\ k_2 & 0 \end{bmatrix}.
\]

Notice that D could also be obtained by replacing the delay elements in the figure with zero.

Since \( A \neq 0 \), the vector \( x(n + 1) \) depends on \( x(n) \). This is represented by the feedback path in Fig. 13.4-2. However, from Fig. 13.4-3 we note that there is no feedback connection in the structure!

Example 13.4.2: An IIR System (the Coupled Form Structure)

Consider the IIR filter structure of Fig. 13.4-4 with one input, two outputs and two state variables, as labeled. This is called the coupled form structure in digital filtering literature [Oppenheim and Schafer, 1989]. The state space description is verified to be

\[
A = R \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} R \sin \theta & R \cos \theta \\ 0 & 1 \end{bmatrix}, \quad D = 0.
\]
The coupled form IIR structure.

The fact that $D = 0$ (i.e., $H(\infty) = 0$) implies that there is no direct (delay-free) path between the input and any of the outputs, as can be verified from the figure.

State space “structures,” “descriptions,” and “realizations.” It is important to distinguish between state space structures and state space descriptions. This is best explained with the above examples. For the structure of Example 13.4.2 we found that the elements of the matrices $(A, B, C, D)$ are equal to the values of the multipliers in the structure so that Fig. 13.4-4 is an implementation of (or a structure for) $(A, B, C, D)$. In Example 13.4.1 on the other hand, this is not true because of the appearance of the product $k_1k_2$ in $C$. So $(A, B, C, D)$ is merely the state space description of the structure in Fig. 13.4-3 but the figure is not a structure for $(A, B, C, D)$. Whenever we say ‘the realization $(A, B, C, D)$’ we just mean a structure whose state space description agrees with $(A, B, C, D)$.

13.4.1 Properties of State Space Descriptions

Transfer Function

The state space description $(A, B, C, D)$ completely determines the input output behavior of the system. We now proceed to substantiate this by writing down the transfer function as well as the time domain input-output relations in terms of $A, B, C,$ and $D$.

Taking $z$-transforms of both sides of (13.4.3) and (13.4.4) we obtain

\[
\begin{align*}
ZX(z) &= AX(z) + BU(z) \quad \text{(13.4.10a)} \\
Y(z) &= CX(z) + DU(z).
\end{align*}
\]

From these we obtain $Y(z) = H(z)U(z)$ where

\[
H(z) = D + C(zI - A)^{-1}B. \quad \text{(13.4.10b)}
\]
In order to express the output $y(n)$ in terms of $(A, B, C, D)$ and the input, we first write $y(n)$ in terms of the initial state vector (or 'initial condition') $x(0)$ and the input $u(m)$, $0 \leq m \leq n$. By repeated application of (13.4.3) we find

$$x(n) = A^n x(0) + \sum_{m=0}^{n-1} A^m B u(n - m - 1), \quad n \geq 1,$$

so that from (13.4.4)

$$y(n) = CA^n x(0) + \sum_{m=0}^{n-1} CA^m B u(n - m - 1) + Du(n), \quad n \geq 1. \quad (13.4.11b)$$

In particular if we have $x(0) = 0$ then

$$y(n) = Du(n) + \sum_{k=1}^{n} CA^{k-1} Bu(n - k). \quad (13.4.12)$$

**Impulse response.** By comparing the above equation with (2.2.7), we can write the the impulse response as

$$h(k) = \begin{cases} D, & k = 0 \\ CA^{k-1} B, & k > 0. \end{cases} \quad (13.4.13)$$

Thus, all coefficients of the impulse response can be calculated from the matrices $(A, B, C, D)$. A second way to obtain this based on power series expansion of (13.4.10b) is addressed in Problem 13.18.

**Poles of $H(z)$ and Eigenvalues of $A$**

We say that $z_p$ is a pole of $H(z)$ if it is a pole of at least one of the elements $H_{km}(z)$ of the matrix $H(z)$. Now from (13.4.10b) we see that

$$H(z) = D + \frac{CR(z)B}{\det (zI - A)}, \quad (13.4.14)$$

where $R(z)$ is the adjugate of $(zI - A)$. Thus, if $z_p$ is a pole then it is a zero of the determinant appearing in (13.4.14). So

$$z_p = \text{pole} \Rightarrow \det (z_p I - A) = 0$$

$$\Rightarrow (z_p I - A)v = 0 \quad \text{(for some } v \neq 0)$$

$$\Rightarrow Av = z_p v, \quad v \neq 0,$$

674 Chap. 13. Review of MIMO LTI systems
which shows that poles of $H(z)$ are eigenvalues of $A$. So the system $H(z)$ is stable if all the eigenvalues satisfy $|\lambda_i| < 1$.

But we have to be careful with the converse. In general, not all eigenvalues $\lambda$ of $A$ are poles of $H(z)$, unless $(A, B, C, D)$ is minimal (to be defined in Section 13.4.2). For the moment just note that even if the determinant in the denominator of (13.4.14) is zero at $z = \lambda$, this can, in principle, cancel with a factor $(z - \lambda)$ in every entry of $CR(z)B$, so that $\lambda$ may not in reality be a pole.

It is possible for all eigenvalues of $A$ to be zero (as in Example 13.4.1). This implies that $H(z)$ is FIR (i.e., has all poles at $z = 0$).

### Similarity Transformations

Given a state space description $(A, B, C, D)$ for a system $H(z)$, suppose we define a new set of matrices

$$ A_1 = T^{-1}AT, \quad B_1 = T^{-1}B, \quad C_1 = CT, \quad D_1 = D, \quad (13.4.16a) $$

where $T$ is any nonsingular matrix. We then see that the transfer function corresponding to this description is given by $D_1 + C_1(zI - A_1)^{-1}B_1$ which upon substitution from (13.4.16a) reduces to $H(z)$ given by (13.4.10b). In other words, $(A_1, B_1, C_1, D)$ is an equivalent state space description for the system $H(z)$. In this way we can find an infinite number of equivalent descriptions because $T$ is arbitrary. The matrix $T$ is called a **similarity transformer** (or transformation).

We can rewrite the state equation (13.4.3) as

$$ T^{-1}x(n + 1) = T^{-1}AT^{-1}x(n) + T^{-1}Bu(n), $$

which shows that the new system has the **transformed** state vector $T^{-1}x(n)$. So, the similarity transformation changes the internal state vector while retaining the same input-output relation. In the next subsection we will use this idea to generate the so-called **minimal** realizations.

Notice that the eigenvalues of $A$ are the same as those of $A_1$ (Appendix A). From Cayley-Hamilton theorem (the same Appendix) we conclude that the quantities $A^N$ and $A_1^N$ can be expressed as

$$ A^N = \sum_{k=0}^{N-1} \alpha_k A^k, \quad A_1^N = \sum_{k=0}^{N-1} \alpha_k A_1^k, \quad (13.4.16b) $$

by using the same set of scalars $\alpha_k$. This is a useful fact.

### 13.4.2 Minimal Realizations

The number $N$ of delay elements in a structure is equal to the size of the state vector $x(n)$. So $N$ is said to be the **dimension of the state space**. This integer $N$ also governs the size of $A$ (which is $N \times N$).
A structure (or implementation or realization) for a transfer function is said to be \textit{minimal} if the number of delay elements $N$ is the smallest possible, viz., the degree $\mu$ of $H(z)$. So, a structure is minimal if and only if the state space has smallest dimension (i.e., size of $A$ is smallest). In this section we introduce two properties called reachability and observability, and show that a structure is minimal if and only if it satisfies these two properties.

\textbf{Example 13.4.3}

To motivate, consider Fig. 13.4-5 which shows a nonminimal structure. The system transfer function is $H(z) = z^{-1}$, which has degree one. But there are four delays in the structure. The first is unconnected to the output $y(n)$, and its output is not observable; the second delay is unconnected to the input $u(n)$ and is not reachable, that is, its output cannot be changed by choice of input $u(n)$. The third delay represents an unreachable and unobservable component. The fourth delay is the only useful one, as it is both reachable and observable.

![Figure 13.4-5](image)

\textit{Figure 13.4-5} A simple example of a structure with unreachable and unobservable states.

![Figure 13.4-6](image)

\textit{Figure 13.4-6} A more subtle example of a structure with unreachable states.

Not all structures are as simple to analyze and understand as the above one. Thus consider Fig. 13.4-6, where $H(z) = 2z^{-1}$. Evidently
this has degree one, but the structure has two delays. Both delays are
connected to the input as well as to the output. However, this structure
is not ‘reachable’ in the sense that we cannot find an input sequence to
achieve arbitrary values for the state variables [as they are constrained
by \( x_1(n) = x_2(n) = u(n - 1) \)].

Decoupling unreachable and unobservable state variables. As
the above example shows, there are situations where we cannot identify a
particular state variable to be unreachable, even though the system has more
delays than necessary. We will show that such structures can be transformed
(using the similarity transformation) in such a manner that the transformed
state vector has two sets of state variables. One of these is the set of un-
reachable variables and can be discarded, thereby reducing the number of
delay elements. In a similar way, we can perform a transformation that
separates or “decouples” the unobservable state variables, which can then
be discarded. The result of these manipulations will be a structure with
minimum number of delays.

It is now time to make our definitions and analysis mathematically more
precise.

Reachability

Consider a structure for \( H(z) \) with \( N \) delays. Let \((A, B, C, D)\) be the
state space description so that \( A \) is \( N \times N \). The structure for \( H(z) \) is said to
be reachable (or completely reachable, often abbreviated cr) if we can reach
any specified final state \( x_f \) starting from any arbitrary initial state \( x_i \) by
application of an appropriate finite length input sequence. Because of the
shift invariant nature of the system, we will assume that the initial time is
zero, that is, \( x(0) = x_i \).

To study the conditions for reachability, recall that (13.4.11a) gives us
the general relation between any state \( x(n), n > 0 \) and the initial state \( x(0) \).
Given arbitrary initial state \( x(0) \), we can reach any specified state \( x_f \) at time
\( n \) (i.e., force \( x(n) = x_f \)) by choice of the inputs \( u(m) \) in the summation of
(13.4.11a), provided the matrix

\[
\begin{bmatrix}
B & AB & \ldots & A^{n-1}B
\end{bmatrix}
\] (13.4.17)

has full row rank \( N \) (Appendix A). If this is not the case, we can try to reach
the desired state \( x_f \) at time \( n + 1 \). Now suppose that we have not been able
to reach the desired state up until time \( N \). This means that the matrix

\[
\mathcal{R}_{A,B} \triangleq \begin{bmatrix}
B & AB & \ldots & A^{N-1}B
\end{bmatrix}_{N \times N^r}
\] (13.4.18)

does not have rank \( N \). If this is the case, then any further waiting will prove
to be unfruitful. In other words, it will not be possible to reach the state

Sec. 13.4 State space descriptions 677
\( x_f \) for any \( n > N \). The reason for this is that the additional columns of the matrix (13.4.17), which are of the form

\[
A^N B, \ A^{N+1} B, \ldots \tag{13.4.19}
\]

can be expressed as linear combinations of the columns of \( R_{A,B} \) (Cayley-Hamilton theorem) so that we do not obtain additional linearly independent columns after time \( N \). In other words, the rank of (13.4.17) does not increase as \( n \) is increased beyond \( N \). Summarizing, we have proved the following important result.

Lemma 13.4.1. Reachability. A structure with \( N \) delays (i.e., \( N \) state variables) is reachable if and only if the matrices \( A \) and \( B \) are such that the matrix \( R_{A,B} \) has full rank \( N \). \( \square \)

Remarks

1. Consider the \( N \times N \) matrix \( R_{A,B} R_{A,B}^\dagger \). We know this is positive semi definite for any \( R_{A,B} \). The full-rank condition on \( R_{A,B} \) is equivalent to the condition \( R_{A,B} R_{A,B}^\dagger > 0 \) (i.e., positive definite).

2. Since reachability is governed by the two matrices \( A \) and \( B \), we often say "\((A,B)\) is reachable" instead of "the structure is reachable."

3. Evidently reachability is a property of the structure implementing \( H(z) \), rather than a property of \( H(z) \).

4. The requirement of reachability is stronger than controllability which is a notion we will not use (see Problem 13.7).

What if a structure is not reachable? This means that we can perform a transformation on the structure so that the new state vector reveals the unreachable state variables explicitly. These variables can then be eliminated to obtain an equivalent structure for \( H(z) \) with fewer delays. More formally we will prove:

Lemma 13.4.2. Reduction to reachable form. Suppose \((A,B)\) is such that \( R_{A,B} \) does not have full rank \( N \). Then there exists a similarity transformation \( T \) such that the equivalent system (13.4.16a) has matrices \( A_1 \) and \( B_1 \) of the form

\[
A_1 = \begin{pmatrix} \rho & N - \rho \\ N - \rho & 0 \end{pmatrix}, \quad B_1 = \begin{pmatrix} B_{11} \\ 0 \end{pmatrix}. \tag{13.4.20}
\]

where \( \rho \) denotes the rank of \( R_{A,B} \). \( \square \)

Consequence of the Lemma. If we have a structure for which the matrices \( A_1 \) and \( B_1 \) have the form (13.4.20), then the state equation can be partitioned as

\[
x_1(n + 1) = A_{11} x_1(n) + A_{12} x_2(n) + B_{11} u(n),
\]

\[
x_2(n + 1) = A_{22} x_2(n). \tag{13.4.21}
\]

678 Chap. 13. Review of MIMO LTI systems
So the subset of states $x_2(n)$ is independent of the input $u(n)$. Moreover its evolution in time is independent of $x_1(n)$. This means that $x_2(n)$ is the \textit{unreachable component} of the state (as it cannot be changed in any way by changing the input). Assuming that the initial state is zero, the quantity $x_2(n)$ is zero for all future time, and does not affect the input-output behavior [and hence the transfer function $H(z)$].

The reduced system $A_{11},B_{11},C_{11},D$ (with state vector $x_1(n)$ of reduced dimension $\rho$) has the same input-output behavior as the original system. Here $C_{11}$ represents the leftmost $p \times \rho$ submatrix of $C_1 \triangleq CT$.

**Proof of Lemma 13.4.2.** Let $t_0,t_1,\ldots,t_{\rho-1}$ denote a set of $\rho$ independent columns of $\mathcal{R}_{A,B}$. Evidently all columns of $\mathcal{R}_{A,B}$ (which are $N$-vectors) are linear combinations of these. Let $t_\rho,\ldots,t_{N-1}$ be a basis for the orthogonal complement of the column space of $\mathcal{R}_{A,B}$. Defining the $N \times N$ nonsingular matrix

$$T = [t_0 \ldots t_\rho \ldots t_{N-1}], \quad (13.4.22)$$
we then have

$$\mathcal{R}_{A,B} = T \begin{bmatrix} P \\ 0 \end{bmatrix} \quad (13.4.23)$$

where the 0 on the RHS is $(N - \rho) \times Nr$. From this and from the definition (13.4.18) of $\mathcal{R}_{A,B}$ we deduce, in particular, that $B$ has the form

$$B = T \begin{bmatrix} B_{11} \\ 0 \end{bmatrix}. \quad (13.4.24)$$

Since $A^N$ can be expressed as a linear combination of $A^k$, $0 \leq k \leq N - 1$ (Cayley-Hamilton theorem), Eq. (13.4.23) also implies

$$A \begin{bmatrix} B & AB & \ldots & A^{N-1}B \end{bmatrix}_{\mathcal{R}_{A,B}} = T \begin{bmatrix} \times \\ 0 \end{bmatrix}, \quad (13.4.25)$$

where $\times$ denotes possibly nonzero entries. Since the first $\rho$ columns of $T$ span the column space of $\mathcal{R}_{A,B}$, (13.4.25) implies

$$AT = T \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \quad \text{i.e.,} \quad A = T \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}T^{-1}, \quad (13.4.26)$$

where $A_{11}$ is $\rho \times \rho$, and $A_{12}$ and $A_{22}$ are of appropriate sizes. The results (13.4.24) and (13.4.26) establish the existence of a similarity transformer $T$ such that the new state space description has the form (13.4.20).

Sec. 13.4 State space descriptions 679
Observability

A structure for $H(z)$ is said to be observable (or completely observable, often abbreviated co) if the state $x(m)$ at time $m$ can be uniquely determined by observing a finite-length segment of the output sequence starting from time $m$, and knowing the input sequence for the corresponding set of sample values. In view of shift invariance we shall set $m = 0$ in our discussions. Like reachability, observability is a property of the structure and not of the transfer function $H(z)$. Once again, if the structure is not observable, then measurement of more than $N$ output samples does not help to identify the initial state $x(0)$, where $N$ is the size of $x(n)$. So the structure is observable if the knowledge of $u(n), y(n), 0 \leq n \leq N - 1$ can be used to find $x(0)$ uniquely.

To determine the conditions for observability, recall that $y(n)$ can be expressed as in (13.4.11b). In terms of the impulse response $h(k)$ defined in (13.4.13) we therefore get

$$
\begin{bmatrix}
y(0) \\
y(1) \\
\vdots \\
y(N-1)
\end{bmatrix} =
\begin{bmatrix}
C \\
CA \\
\vdots \\
CA^{N-1}
\end{bmatrix} x(0) +
\begin{bmatrix}
h(0) & 0 & \cdots & 0 \\
h(1) & h(0) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
h(N-1) & h(N-2) & \cdots & h(0)
\end{bmatrix}
\begin{bmatrix}
u(0) \\
u(1) \\
\vdots \\
u(N-1)
\end{bmatrix}
$$

Given the quantities $y(n)$ and $u(n)$ for $0 \leq n \leq N - 1$, that is, given $Y$ and $f$, there exists $x(0)$ satisfying (13.4.27) because, by definition, $y(n)$ satisfies (13.4.27). It is also clear (Appendix A) that we can find a unique value of $x(0)$ if and only if

$$S_{C,A} \triangleq \begin{bmatrix}
C \\
CA \\
\vdots \\
CA^{N-1}
\end{bmatrix}_{Np \times N}
$$

has full column rank $N$. Under this condition, the unique solution is

$$x(0) = [S_{C,A}^\dagger S_{C,A}]^{-1} S_{C,A}^\dagger (Y - f).$$

Summarizing, we have proved:

♠Lemma 13.4.3. Observability. A structure is observable if and only if the matrices $C$ and $A$ are such that $S_{C,A}$ has full rank $N$. □
Remarks. (a) Once again, the full-rank condition on $S_{C,A}$ is equivalent to the positive definite condition $S_{C,A}^\dagger S_{C,A} > 0$. (b) We often say that $(C, A)$ is observable instead of "the structure is observable."

If $(C, A)$ is not observable, it is possible to apply a similarity transformation such that the transformed state reveals the unobservable state variables. These variables can then be eliminated, resulting in a system with fewer delays. This is the consequence of the following lemma.

 Lemma 13.4.4. Reduction to observable form. Suppose $(C, A)$ is such that $S_{C,A}$ does not have full rank $N$. Then there exists a similarity transformation $T$ such that the equivalent system (13.4.16a) has matrices $C_1$ and $A_1$ of the form

$$
C_1 = \begin{pmatrix} \rho & N - \rho \\ C_{11} & 0 \end{pmatrix}, \quad A_1 = \begin{pmatrix} \rho & N - \rho \\ A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}
$$

(13.4.30)

where $\rho$ denotes the rank of $S_{C,A}$.

Proof. Very similar to proof of Lemma 13.4.2.

If the state space description has $C_1$ and $A_1$ of the form (13.4.30), then the state vector can be partitioned into $\begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix}$ where $x_1(n)$ has $\rho$ elements. The part $x_2(n)$ does not affect $x_1(n)$ or the output and is therefore the nonobservable part. The reduced system $A_{11}, B_{11}, C_{11}, D$ (with state vector $x_1(n)$ of dimension $\rho$) has same input-output behavior as the original system. Here $B_{11}$ represents the first $\rho$ rows of $B_1 = T^{-1}B$.

A Beautiful Significance of $S_{C,A}$ and $R_{A,B}$

Suppose we form the product $S_{C,A} R_{A,B}$. The result is

$$
S_{C,A} R_{A,B} = \begin{bmatrix} h(1) & h(2) & \cdots & h(N) \\ h(2) & h(3) & \cdots & h(N + 1) \\ \vdots & \vdots & \cdots & \vdots \\ h(N) & h(N + 1) & \cdots & h(2N - 1) \end{bmatrix}
$$

(13.4.31)

This is an $Np \times Nr$ matrix. It can also be considered as an $N \times N$ block matrix, with each block having size $p \times r$. So the product of the matrices which arise in the reachability and observability conditions is merely a matrix of the impulse response coefficients $h(n)$ for $1 \leq n \leq 2N - 1$.

Reachability and Observability Imply Minimality

It is clear from the above discussions that if a structure is minimal (i.e., $N$ is as small as possible, viz., $N = \text{McMillan degree } \mu$) then it has to be reachable and observable. For, otherwise, we can find a smaller matrix
A_{11} and a corresponding set of matrices B_{11} and C_{11} resulting in the same transfer function H(z). The converse of this result is provided by:

Lemma 13.4.5. Suppose a realization (A, B, C, D) is reachable and observable. Then it is minimal, that is, there does not exist an equivalent structure for the same transfer function with fewer delays.

Proof. We prove this by contradiction. Let N denote the state space dimension for (A, B, C, D) and let (a, b, c, D) be an equivalent realization with smaller dimension \( \rho < N \) (i.e., a is \( \rho \times \rho \), and so on). Since the structures represent the same transfer function, the impulse response coefficients \( h(n) \) are the same for both. From (13.4.13) we therefore have

\[
h(n) = CA^{n-1}B = ca^{n-1}b, \quad n > 0. \tag{13.4.32}
\]

By using the result (13.4.31) we immediately arrive at

\[
\begin{bmatrix}
C \\
CA \\
\vdots \\
CA^{N-1}
\end{bmatrix}
= 
\begin{bmatrix}
B & AB & \cdots & A^{N-1}B
\end{bmatrix}
\begin{bmatrix}
c \\
a \\
\vdots \\
1
\end{bmatrix}
\begin{bmatrix}
b \\
ab & \cdots & a^{N-1}
\end{bmatrix}
\]

from which we obtain

\[
S_{C,A} \begin{bmatrix}
S_{C,A} \\
\mathcal{R}_{A,B} \\
\mathcal{R}_{A,B} \end{bmatrix} = S_{C,A} \begin{bmatrix}
\mathcal{S} \\
\mathcal{S} \mathcal{R} \mathcal{R}_{A,B} \end{bmatrix}.
\tag{13.4.34}
\]

As (A, B, C, D) is reachable and observable, the \( N \times N \) matrices \( S_{C,A} \) and \( \mathcal{R}_{A,B} \) are nonsingular. So, the LHS of (13.4.34) has rank \( N \). But the rank of the RHS is at most \( \rho < N \). This is a contradiction.

The results of the above three Lemmas can be summarized as follows.

Theorem 13.4.1. Minimality, reachability, and observability. A realization (A, B, C, D) of a transfer function \( H(z) \) is minimal if and only if it is reachable and observable (i.e., (A, B) reachable and (C, A) observable).

In all future discussions, the word “minimal” is therefore synonymous to the condition “reachable as well as observable.” This is also abbreviated as \textit{crco} (i.e., completely reachable and completely observable).

Minimal Realizations are Related by Similarity Transforms

A nice property of minimal realizations is that all minimal realizations of a system \( H(z) \) are related by similarity transformations:

\[
682 \quad \text{Chap. 13. Review of MIMO LTI systems}
\]
Lemma 13.4.6. Let \((A, B, C, D)\) and \((a, b, c, D)\) be two minimal realizations of a \(p \times r\) system \(H(z)\). Then there exists a unique similarity transformation \(T\) which transforms \((A, B, C, D)\) to \((a, b, c, D)\).

Proof. Let \(N\) denote the state space dimension so that \(A\) and \(a\) are both \(N \times N\). Our aim is to prove the existence of a unique \(N \times N\) nonsingular \(T\) such that

\[
a = T^{-1}AT, \quad b = T^{-1}B, \quad c = CT.
\]

Assuming \(T\) exists, uniqueness is established as follows: Eq. (13.4.35) implies \(c a^n = C A^n T\) for all integers \(n \geq 0\) so that \(S_{c,a} = S_{C,A} T\). This implies \(S^\dagger_{C,A} S_{c,a} = S^\dagger_{C,A} S_{C,A} T\). Since minimality implies observability, \(S^\dagger_{C,A} S_{C,A}\) is \(N \times N\) nonsingular. So \(T\) is uniquely determined as

\[
T = [S^\dagger_{C,A} S_{C,A}]^{-1} S^\dagger_{C,A} S_{c,a}.
\]  

(13.4.36)

Using (13.4.33) one can verify that \(T\) can also be expressed as

\[
T = \mathcal{R}_{A,B} \mathcal{R}_{a,b} \mathcal{R}_{a,b}^{-1} \mathcal{R}_{a,b}^{-1}
\]  

(13.4.37)

We now prove existence of \(T\). We know that a relation similar to (13.4.35) holds, from which we have

\[
\mathcal{R}_{A,B} = [S^\dagger_{C,A} S_{C,A}]^{-1} S^\dagger_{C,A} S_{c,a} \mathcal{R}_{a,b}.
\]  

(13.4.38a)

This step has been possible because the required inverse exists (by observability). For clarity, let us rewrite this more explicitly:

\[
\begin{bmatrix}
B & A^N & B & A^N B \\
A & B & A & B \\
\vdots & \vdots & \ddots & \vdots \\
A^N & B & A & B \\
\end{bmatrix} = T \begin{bmatrix}
b & ab & \ldots & a^N b \\
ab & \ldots & \vdots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
a b & \ldots & \ldots & a b \\
\end{bmatrix},
\]  

(13.4.38b)

where \(T\) is the \(N \times N\) matrix indicated in (13.4.38a). Note that this choice of \(T\) agrees with (13.4.36) as expected. From this we obtain \(B = T b\), which is one of the three relations in (13.4.35). Now from (13.4.32) we have

\[
S_{C,A} A^N B = S_{c,a} a^N b.
\]  

(13.4.39)

So if we append the columns \(A^N B\) and \(a^N b\) to the matrices \(\mathcal{R}_{A,B}\) and \(\mathcal{R}_{a,b}\) in (13.4.33) respectively, the equality continues to hold. In a way similar to (13.4.38b) we then obtain

\[
A \begin{bmatrix}
B & A^N & B & A^N B \\
A & B & A & B \\
\vdots & \vdots & \ddots & \vdots \\
A^N & B & A & B \\
\end{bmatrix} = T \begin{bmatrix}
b & ab & \ldots & a^N b \\
ab & \ldots & \vdots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
a b & \ldots & \ldots & a b \\
\end{bmatrix}.
\]  

(13.4.40)

Sec. 13.4 State space descriptions 683
Postmultiplying both sides of (13.4.40) by $R_{a,b}^\dagger$ and rearranging we obtain $AT = Ta$, which again is one of the relations in (13.4.35). Finally, from (13.4.33) we have

$$CR_{A,B} = cR_{a,b}. \tag{13.4.41}$$

Postmultiplying with $R_{a,b}^\dagger$ and rearranging, we obtain $CT = c$. Summarizing, we have shown that if $T$ is defined as in (13.4.36) [equivalently (13.4.37)], then all the three relations in (13.4.35) are satisfied. This establishes the existence of $T$ with the advertised properties. Notice that the minimality properties are crucial in the proof because we use the inverses of $R_{A,B}R_{A,B}^\dagger$ and $S_{C,A}S_{C,A}$.

Real coefficient systems. Suppose $H(z)$ is a real transfer matrix (i.e., all elements $H_{km}(z)$ have real coefficients). It is clear that there exists a structure with real valued multipliers (because we can realize each $H_{km}(z)$ individually in direct form!). What is perhaps not so obvious is the fact that there also exist minimal realizations with real multiplier values. A proof is requested in Problem 13.9.

The PBH Test for Reachability and Observability

The rank conditions on the matrices $R_{A,B}$ and $S_{C,A}$ can be expressed in an elegant form, in terms of the eigenvectors of $A$. This result is of considerable theoretical value (as it simplifies many proofs). It is called the PBH condition (or test) because it was invented by Popov, Belevitch, and Hautus.

To motivate the basic idea, suppose $v$ is an eigenvector of $A$ which is at the same time orthogonal to all rows of $C$, i.e.,

$$Av = \lambda v, \quad \text{and} \quad Cv = 0, \quad v \neq 0. \tag{13.4.42}$$

We then have $CAv = \lambda Cv = 0$. More generally, we can verify $CA^n v = 0$ for any integer $n \geq 0$ so that $S_{C,A}v = 0$. This implies that the rank of $S_{C,A}$ is less than $N$, so that the system is not observable. The PBH test asserts an even stronger result, summarized as follows:

\begin{itemize}
  \item Theorem 13.4.2. The PBH test. This can be stated in two parts.
    \begin{enumerate}
      \item The pair $(C, A)$ is observable if and only if there does not exist a nonzero vector $v$ such that $Av = \lambda v$ and $Cv = 0$.
      \item The pair $(A, B)$ is reachable if and only if there does not exist a nonzero vector $w$ such that $w^T A = \lambda w^T$ and $w^T B = 0$.
    \end{enumerate}
\end{itemize}

\textbf{Proof}. We prove only part 1, as the other part is similar. We already showed that (13.4.42) implies that $(C, A)$ is not observable. Conversely, suppose $(C, A)$ is not observable. Then we can apply a similarity transform $T$ and obtain the equivalent form (13.4.30). Let $w$ be an eigenvector of $A_{22}$.
We then have

\[ A_1 \begin{bmatrix} 0 \\ w \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} 0 \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ A_{22}w \end{bmatrix} = \begin{bmatrix} 0 \\ \lambda w \end{bmatrix} = \lambda \begin{bmatrix} 0 \\ w \end{bmatrix}, \]

so that \( u^{\Delta} \begin{bmatrix} 0 \\ w \end{bmatrix} \) is an eigenvector of \( A_1 \). Furthermore

\[ C_1 u = [C_{11} \ 0] \begin{bmatrix} 0 \\ w \end{bmatrix} = 0. \]

This implies that the nonzero vector \( v^{\Delta} = Tu \) satisfies (13.4.42).

We conclude this subsection with the following easily proved result.

\[ \blacktriangle \text{Fact 13.4.1. Observability and reachability preserved under similarity transform.} \text{ Let } (C, A) \text{ be observable. Then the pair } (C_1, A_1) \text{ obtained by a similarity transformation is also observable. Same holds for reachability.} \blacktriangledown \]

### 13.4.3 Stability and Lyapunov Lemma

Recall from Sec. 2.2 that \( H(z) \) is said to be stable if each element \( H_{km}(z) \) is stable (i.e., has all poles inside the unit circle). Given any minimal realization \( (A, B, C, D) \) for \( H(z) \), it turns out (see Theorem 13.6.1 later) that \( \lambda \) is a pole of \( H(z) \) if and only if it is an eigenvalue of \( A \).

So stability is equivalent to the condition that all eigenvalues \( \lambda_i \) of \( A \) satisfy \( |\lambda_i| < 1 \). If \( A \) satisfies this, we say that \( A \) is stable. The following property of stable matrices is very valuable in system theoretic work.

\[ \blacktriangle \text{Lemma 13.4.7. Let } A \text{ be stable. Then } A^n \to 0 \text{ as } n \to \infty. \blacktriangledown \]

**Proof.** If \( A \) has distinct eigenvalues, we can write \( A = T \Lambda T^{-1} \) where \( \Lambda \) is a diagonal matrix with diagonal elements representing the eigenvalues of \( A \), and the columns of \( T \) are the eigenvectors. So \( A^n = T \Lambda^n T^{-1} \). Since \( |\lambda_i| < 1 \), the quantity \( \lambda_i^n \) goes to zero as \( n \to \infty \), so that \( A^n \to 0 \).

For the case where eigenvalues are not distinct, we cannot in general diagonalize \( A \). In this case we can still apply the unitary triangularization result (Sec. A.7, Appendix A). According to this, any square matrix \( A \) can be written as \( A = U \Delta U^\dagger \), where \( U \) and \( \Delta \) are square matrices with \( U^\dagger U = I \), and \( \Delta = \) lower triangular. The diagonal elements of \( \Delta \) are equal to the eigenvalues of \( A \).

We have \( A^n = U \Delta^n U^\dagger \) (using \( U^\dagger U = I \)) so that proving \( A^n \to 0 \) is equivalent to proving \( \Delta^n \to 0 \). To prove this we adopt the following simple trick [Franklin, 1968]: define a lower triangular matrix \( \hat{\Delta} \) as follows:
1. The diagonal elements $[\Delta]_{ii}$ (i.e., eigenvalues of $\Delta$) are distinct, with $|[\Delta]_{ii}| \leq [\Delta]_{ii} < 1$. This is always possible since stability of $A$ assures us that $|[\Delta]_{ii}| < 1$ for all $i$.

2. $[\Delta]_{ij} \geq |[\Delta]_{ij}|$ for all $(i,j)$.

As the eigenvalues of $\Delta$ are distinct and bounded by unity, we conclude $\Delta^n \to 0$. Since all the elements of $\Delta$ are nonnegative and satisfy $[\Delta]_{ij} \geq |[\Delta]_{ij}|$, it is clear that $[\Delta^n]_{ij} \geq [\Delta^n]_{ij}$ for all integers $n > 0$ so that $\Delta^n \to 0$ indeed!

The stability of $A$ can be expressed elegantly in terms of a simple algebraic equation. This is called the discrete-time Lyapunov Lemma, even though it is a blending of results due to many authors [Anderson and Moore, 1979].

Lemma 13.4.8. Discrete-time Lyapunov lemma. We will state this in two parts for convenience.

1. Let $A$ be $N \times N$, $C$ be $p \times N$ and let $(C, A)$ be observable. Let the equation

$$A^\dagger PA + C^\dagger C = P,$$  

be satisfied for some Hermitian positive definite $P$. Then $A$ is stable.

2. Let $Q$ be some $N \times N$ Hermitian positive semidefinite matrix so that it can be written as $Q = C^\dagger C$ for some (possibly rectangular) $C$. Let $A$ be $N \times N$ stable with $(C, A)$ observable. Then the algebraic equation (13.4.43) has a unique solution $P$, and the solution is Hermitian positive definite.

\[ \text{Proof.} \] First consider part 1. Since $P$ is Hermitian and positive definite, there exists nonsingular $T$ such that $P = [TT^\dagger]^{-1}$. So we can rearrange (13.4.43) as

$$T^\dagger A^\dagger T^{-\dagger} T^{-1} AT + T^\dagger C^\dagger CT = I.$$  

(13.4.44)

Define $A_1 = T^{-1} AT$ and $C_1 = CT$. Then (13.4.44) becomes

$$A_1^\dagger A_1 + C_1^\dagger C_1 = I.$$  

(13.4.45)

Now suppose $\lambda$ is an eigenvalue of $A$ (hence that of $A_1$) so that $A_1 v = \lambda v$ for some $v \neq 0$. From (13.4.45) we have $v^\dagger A_1^\dagger A_1 v + v^\dagger C_1^\dagger C_1 v = v^\dagger v$, that is, $(1 - |\lambda|^2)v^\dagger v = [C_1 v]^\dagger [C_1 v]$. Since $[C_1 v]^\dagger [C_1 v] \geq 0$, this proves that $|\lambda| < 1$ unless $C_1 v = 0$. But by PBH test (Theorem 13.4.2) this cannot happen because observability of $(C, A)$ implies that of $(C_1, A_1)$ (by Fact 13.4.1). This proves Part 1.

Now consider part 2. Define

$$P = \sum_{n=0}^{\infty} [A^\dagger]^n C^\dagger CA^n.$$  

(13.4.46)
Since \( A \) is stable, this summation converges [p. 64, Anderson and Moore, 1979]. Evidently \( P \) is Hermitian and positive semidefinite. Moreover, the sum of the first \( N \) terms is precisely equal to \( S_{\text{c}, A}^\dagger S_{\text{c}, A} \), which is positive definite (by observability), so that \( P \) is positive definite. It is readily verified by substitution that this \( P \) satisfies (13.4.43).

To prove that \( P \) is the only solution, suppose \( P_1 \) and \( P_2 \) are solutions to (13.4.43). Then \( A^\dagger R A = R \), where \( R = P_1 - P_2 \). By repeated use of this, one obtains \( [A^\dagger]^n R A^n = R \), for all \( n \geq 0 \). By Lemma 13.4.7, \( A^n \) goes to zero as \( n \to \infty \) so that this implies \( R = 0 \). So \( P \) is unique.

\[ \nabla \nabla \nabla \]

13.5 **THE SMITH-McMILLAN FORM**

For transfer matrices \( H(z) \) and their structures, a study of advanced concepts such as transmission zeros, minimality, degree, and related things is greatly facilitated by a diagonalization result known as the Smith-McMillan decomposition. In this section we study these results.

13.5.1 **The Smith Form of a Polynomial Matrix**

Given a \( p \times r \) polynomial matrix \( P(z) \), it is possible to obtain simpler forms such as triangular and diagonal forms [Smith, 1861], by performing certain operations called elementary operations on the matrix.

**Elementary Operations on Polynomial Matrices**

An elementary row operation on \( P(z) \) is defined to be any one of the following:

*Type 1.* Interchange two rows.

*Type 2.* Multiply a row with a nonzero constant \( c \).

*Type 3.* Add a polynomial multiple of a row to another row.

Elementary column operations are defined in a similar way. The above row operations can be performed by premultiplying \( P(z) \) with an appropriate square matrix, called an elementary matrix. There are three types of elementary matrices corresponding to the above three operations. Examples of \( 3 \times 3 \) elementary matrices are shown below.

\[
\begin{bmatrix}
0 & 0 & 1 \\
0 & 1 & 0 \\
1 & 0 & 0
\end{bmatrix},
\begin{bmatrix}
1 & 0 & 0 \\
0 & c & 0 \\
0 & 0 & 1
\end{bmatrix},
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
\alpha(z) & 0 & 1
\end{bmatrix}.
\]

(13.5.1)

The first matrix interchanges row 0 with row 2. (Remember that rows and columns are numbered starting from zero.) The second matrix multiplies all elements of row 1 with \( c \neq 0 \). The third matrix replaces row 2 with

\[
(\text{row 2}) + (\alpha(z) \times \text{row 0}),
\]

(13.5.2)
where \( \alpha(z) \) is a polynomial in \( z \).

Notice that all the three types of elementary matrices are unimodular. Repeated application of various row operations amounts to premultiplication of \( P(z) \) with a \( p \times p \) unimodular matrix. (We will see later that any unimodular matrix is a product of the three types of elementary matrices.)

Elementary column operations are equivalent to post multiplication of \( P(z) \) with one of the three types of elementary matrices. Repeated elementary column operations amounts to post multiplication by an \( r \times r \) unimodular matrix.

Elementary operations can perform wonders. For example, by repeated row and column operations, one can reduce \( P(z) \) into a diagonal matrix, whose diagonal entries are polynomials. This result is theoretically extremely powerful, and helps us to obtain many conclusions which would otherwise be difficult to derive.

The division theorem. The key property which enables us to obtain the diagonalization is the division theorem for polynomials. This states that if \( P_1(z) \) and \( P_2(z) \) are scalar polynomials in \( z \) with order of \( P_1(z) \geq \) order of \( P_2(z) \), then we can find unique polynomials \( Q(z) \) (quotient polynomial) and \( R(z) \) (reminder polynomial) such that

\[
P_1(z) = Q(z)P_2(z) + R(z), \tag{13.5.3}
\]

and such that the order of \( R(z) < \) order of \( P_2(z) \). (This includes the case \( R(z) = 0 \).)

Example 13.5.1

Let

\[
P(z) = \begin{pmatrix} 0 & 1 \\ z + 1 & z \\ 2z^2 + 3 & 2(z + 1)^2 \end{pmatrix}. \tag{13.5.4}
\]

If we divide the \((1, 0)\) element \(2z^2 + 3\) by the \((0, 0)\) element \(z + 1\) we obtain

\[
2z^2 + 3 = \underbrace{2(z - 1)(z + 1)}_{Q(z)} + \underbrace{5}_{R(z)}. \tag{13.5.5}
\]

By using this fact we can perform an elementary row operation of Type 3 to reduce the \((1, 0)\) element to a constant \((= R(z) = 5)\). Thus

\[
\begin{pmatrix} 1 & 0 \\ -2(z - 1) & 1 \end{pmatrix} \begin{pmatrix} z + 1 & z \\ 2z^2 + 3 & 2(z + 1)^2 \end{pmatrix} = \begin{pmatrix} z + 1 & z \\ 5 & 2(3z + 1) \end{pmatrix}. \tag{13.5.6}
\]
We can now reduce the $(0,0)$ element to a constant by replacing row 0 with

$$(\text{row } 0) - z/5 \times (\text{row } 1) \quad \text{(Type 3 operation)}.$$

The result is

$$\begin{bmatrix} 1 & -z/5 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z + 1 & z \\ 5 & 2(3z + 1) \end{bmatrix} = \begin{bmatrix} 1 & 3z(1 - 2z)/5 \\ 5 & 2(3z + 1) \end{bmatrix}. \quad (13.5.7)$$

Next the $(0,1)$ element can be forced to zero by an elementary column operation as follows:

$$\begin{bmatrix} 1 & 3z(1 - 2z)/5 \\ 5 & 2(3z + 1) \end{bmatrix} \begin{bmatrix} 1 & -3z(1 - 2z)/5 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 6z^2 + 3z + 2 \end{bmatrix}. \quad (13.5.8)$$

Finally the $(1,0)$ element can be forced to be zero by an elementary row operation:

$$\begin{bmatrix} 1 & 0 \\ -5 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 5 & 6z^2 + 3z + 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 6z^2 + 3z + 2 \end{bmatrix} = \Gamma(z), \quad (13.5.9)$$

eventually resulting in a diagonal matrix. The sequence of row operations can be represented as

$$\begin{bmatrix} 1 & 0 \\ -5 & 1 \end{bmatrix} \begin{bmatrix} 1 & -z/5 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2(z - 1) & 1 \end{bmatrix} = \begin{bmatrix} (2z^2 - 2z + 5)/5 & -z/5 \\ -2z^2 - 3 & z + 1 \end{bmatrix},$$

which we denote $W_1(z)$. The only column operation is represented by

$$V_1(z) = \begin{bmatrix} 1 & 3z(2z - 1)/5 \\ 0 & 1 \end{bmatrix}, \quad (13.5.10)$$

so that

$$W_1(z)P(z)V_1(z) = \Gamma(z). \quad (13.5.11)$$

Since $W_1(z)$ and $V_1(z)$ are unimodular, their inverses are also unimodular polynomials in $z$. So we have the decomposition

$$P(z) = W(z)\Gamma(z)V(z), \quad (13.5.12)$$

where $\Gamma(z)$ is diagonal and $W(z), V(z)$ are unimodular. If desired, the diagonal entries of $\Gamma(z)$ can further be forced to be monic polynomials (i.e., highest power has coefficient unity), by use of Type 2 operations.

The above example is a demonstration of a general diagonalization theorem which can be stated as follows.
\textbf{Theorem 13.5.1. The Smith form} [Smith, 1861]. Let \( P(z) \) be a \( p \times r \) matrix polynomial in \( z \). Then there exists finite number of elementary row and column operations which reduce \( P(z) \) into a diagonal polynomial matrix. So we can write

\[
P(z) = \begin{pmatrix} \mathbf{W}(z) & \Gamma(z) & \mathbf{V}(z) \end{pmatrix} \begin{pmatrix} \mathbf{P}(z) \end{pmatrix}
\]

(Smith form decomposition). \hspace{1cm} (13.5.13)

where \( \mathbf{W}(z) \) and \( \mathbf{V}(z) \) are unimodular matrix polynomials in the variable \( z \) and

\[
\Gamma(z) = \begin{bmatrix}
\gamma_0(z) & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \gamma_1(z) & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & \cdots & \gamma_{p-1}(z) & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0
\end{bmatrix}
\]

Here \( \rho \) is the normal rank of \( P(z) \). Moreover the unimodular matrices can be so chosen that the polynomials \( \gamma_i(z) \) are monic (i.e., highest power has coefficient unity), and \( \gamma_i(z) \) is a factor of \( \gamma_{i+1}(z) \), that is,

\[
\gamma_i(z) | \gamma_{i+1}(z), \hspace{0.5cm} 0 \leq i \leq \rho - 2.
\]

Finally, for a given \( P(z) \), the matrix \( \Gamma(z) \) is unique, and the elements \( \gamma_i(z) \) are given by

\[
\gamma_i(z) = \Delta_{i+1}(z)/\Delta_i(z),
\]

where \( \Delta_i(z), i > 0, \) is the greatest common divisor of all the \( i \times i \) minors of \( P(z) \), and \( \Delta_0(z) = 1 \). \( \Gamma(z) \) is called the Smith form of \( P(z) \).

\textbf{Sketch of proof.} We will assume that \( P_{00}(z) \) is nonzero and has the smallest order among all nonzero elements. [This can be arranged by use of Type 1 row and column operations, which will be part of \( \mathbf{W}(z) \) and \( \mathbf{V}(z) \).] Suppose there is a nonzero element \( P_{0k}(z) \) in the 0th row. Let \( Q(z) \) and \( R(z) \) denote the quotient and remainder when we attempt to divide \( P_{0k}(z) \) with \( P_{00}(z) \), that is,

\[
P_{0k}(z) = Q(z)P_{00}(z) + R(z).
\]

We can now perform a Type 3 column operation so that the resulting matrix has the element \( R(z) \) in place of \( P_{0k}(z) \). (Only the \( k \)th column is affected by this operation.) Since \( R(z) \) is the remainder, its order is smaller than that of \( P_{00}(z) \). By performing another Type 1 operation we can bring \( R(z) \) to the \((0,0)\) location. In this way we can keep reducing the order of the
(0, 0) element. Since this can proceed only a finite number of times, all the remaining elements in the 0th row eventually become zero. In a similar way, by performing elementary row operations, we can convert all the elements in the 0th column (except the (0, 0) element) to zero. At this point, the polynomial matrix takes the form

$$
\begin{bmatrix}
\gamma_0(z) & 0 \\
0 & S(z)
\end{bmatrix},
$$

(13.5.16b)

where \(\gamma_0(z)\) is a scalar polynomial and \(S(z)\) is a matrix polynomial.

Suppose \(\gamma_0(z)\) is not a factor of all the elements of \(S(z)\). For example let the (1, 1) element of (13.5.16b) be one such. We can then add the 1st row to the 0th row (Type 3 operation), and create a situation whereby the order of the (0, 0) element can be reduced further. Since the order reduction cannot proceed indefinitely, we eventually obtain the form (13.5.16b) where \(\gamma_0(z)\) is a factor of all the elements in \(S(z)\).

We can repeat the entire set of operations on the smaller matrix \(S(z)\). Continuing in this way, we eventually arrive at a diagonal matrix \(\Gamma(z)\) whose elements satisfy (13.5.15a). The monic nature of the polynomials \(\gamma_i(z)\) can be ensured trivially by use of Type 2 operations.

It remains to prove (13.5.15b). For this first consider the case where \(W(z) = I_p\) and \(V(z) = I_r\), so that \(P(z) = \Gamma(z)\). In view of the divisibility condition (13.5.15a), we can write

\[
\gamma_0(z) = \alpha_0(z), \quad \gamma_1(z) = \alpha_0(z)\alpha_1(z), \quad \gamma_2(z) = \alpha_0(z)\alpha_1(z)\alpha_2(z), \ldots
\]

where \(\alpha_i(z)\) are polynomials in \(z\). The 1 \times 1 minors of \(\Gamma(z)\) are

\[
0, \ \alpha_0(z), \ \alpha_0(z)\alpha_1(z), \ldots
\]

so that their gcd \(\Delta_1(z) = \alpha_0(z)\). So \(\gamma_0(z) = \Delta_1(z)/\Delta_0(z)\), since \(\Delta_0(z) = 1\). Similarly we can verify \(\gamma_1(z) = \Delta_2(z)/\Delta_1(z)\) and so on. This proves (13.5.15b) when \(P(z) = \Gamma(z)\). Finally consider the case when \(W(z)\) and \(V(z)\) are not identity. These matrices perform elementary operations on the rows and columns of \(\Gamma(z)\). It can be shown by use of the so-called Binet-Cauchy theorem [Gantmacher, 1959] that the gcd of \(i \times i\) minors is unaffected (except for scale factors) by these operations. So (13.5.15b) continues to hold.

\[\nabla \nabla \nabla\]

Notice that (13.5.13) implies, in particular,

\[
\det P(z) = c \times \det \Gamma(z).
\]

(13.5.17)

The matrix \(\Gamma(z)\) is called the Smith form of \(P(z)\), and (13.5.13) the Smith decomposition. Here is an example of \(\Gamma(z)\) that satisfies all conditions stated in the Theorem:

\[
\begin{bmatrix}
2+z & 0 & 0 \\
0 & (2+z)(3-z^2) & 0 \\
0 & 0 & (2+z)^2(3-z^2)
\end{bmatrix}.
\]

Sec. 13.5 The Smith-McMillan form 691
Another example is \( \Gamma(z) = (a + z)I \).

**Example 13.5.2**

Even though the Smith form \( \Gamma(z) \) is unique, the matrices \( W(z) \) and \( V(z) \) are not. For example,

\[
\begin{bmatrix}
1 & 0 \\
z^2 & z
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 \\
z^2 & 1
\end{bmatrix}
\begin{bmatrix}
1 & 0 \\
0 & z
\end{bmatrix}
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
1 & 0 \\
z & 1
\end{bmatrix},
\]

which shows that there are (at least) two possible choices for \( W(z) \) and \( V(z) \).

**Application to unimodular matrices.** Now suppose that \( P(z) \) is itself a unimodular matrix, i.e., a square matrix with nonzero constant determinant. In view of (13.5.17), the Smith form has to be \( \Gamma(z) = I \) in this case. Since \( W(z) \) and \( V(z) \) are products of elementary matrices, this proves that any unimodular matrix is a product of elementary matrices. This important result can be summarized as

\[ \text{Corollary 13.5.1. Factorization of unimodular matrices. A square polynomial matrix } P(z) \text{ is unimodular if and only if it is a product of finite number of elementary matrices.} \]

**13.5.2 Application in Glcd Extraction**

Given the polynomial matrices \( P(z) \) and \( Q(z) \) in the MFD \( Q^{-1}(z)P(z) \), suppose we wish to extract a greatest left common divisor (glcd). This can be done by applying the above decomposition theory. For this let

\[
S(z) \overset{\Delta}{=} P \begin{pmatrix} Q(z) & P(z) \end{pmatrix}. \tag{13.5.18}
\]

Performing the Smith decomposition we get

\[
W_1(z)[Q(z) \quad P(z)]V_1(z) = [\times \quad 0], \tag{13.5.19}
\]

where \( \times \) is \( p \times p \) denoting possibly nonzero polynomial entries. The fact that \( \times \) is diagonal is irrelevant in this application. Since \( W_1(z) \) is unimodular, its inverse is a polynomial in \( z \) and we can rewrite (13.5.19) as

\[
[Q(z) \quad P(z)]V_1(z) = [R(z) \quad 0] \tag{13.5.20}
\]

where \( R(z) \) is a \( p \times p \) polynomial matrix. Now \( V(z)\overset{\Delta}{=}V_1^{-1}(z) \) is a polynomial in \( z \) (since \( V_1(z) \) is unimodular), so we can write (13.5.20) as

\[
[Q(z) \quad P(z)] = [R(z) \quad 0] \begin{pmatrix} V_{00}(z) & V_{01}(z) \\ V_{10}(z) & V_{11}(z) \end{pmatrix} \tag{13.5.21}
\]

692 Chap. 13. Review of MIMO LTI systems
where $V_{ij}(z)$ are polynomials of appropriate sizes. So

$$Q(z) = R(z)V_{00}(z), \quad P(z) = R(z)V_{01}(z) \quad (13.5.22)$$

which shows that $R(z)$ is an lcd of $P(z)$ and $Q(z)$.

We now show that $R(z)$ is in fact a gcd of $Q(z)$ and $P(z)$. By partitioning $V_1(z)$ in an obvious manner we can write (13.5.20) as:

$$[Q(z) \quad P(z)] \begin{bmatrix} V_q(z) \\ V_p(z) \end{bmatrix} \times = [R(z) \quad 0] \quad (13.5.23)$$

where $\times$ denotes entries whose details are irrelevant. From this equation we obtain

$$Q(z)V_q(z) + P(z)V_p(z) = R(z). \quad (13.5.24)$$

This shows that any lcd of $P(z)$ and $Q(z)$ is also a left divisor of $R(z)$. So $R(z)$ is in fact a gcd of $Q(z)$ and $P(z)$. Summarizing, we have established:

Lemma 13.5.1. Extension of Euclid’s theorem. Let $Q(z)$ and $P(z)$ be matrix polynomials in $z$ with sizes $p \times p$ and $p \times r$, respectively, and let $R(z)$ be a glcd. Then there exist matrix polynomials $V_p(z)$ and $V_q(z)$ such that $Q(z)V_q(z) + P(z)V_p(z) = R(z).$

Comments

1. When $Q(z)$ and $P(z)$ are scalar polynomials, this reduces to the well-known Euclid’s theorem [Sec. 2.3, Bose, 1985]. This says that when $Q(z)$ and $P(z)$ are polynomials in $z$ with greatest common factor $R(z)$, there exist polynomials $V_p(z)$ and $V_q(z)$ such that $Q(z)V_q(z) + P(z)V_p(z) = R(z)$.

2. When $Q(z)$ and $P(z)$ are left coprime, we can take $R(z) = I$. So there exist polynomial matrices $V_q(z)$ and $V_p(z)$ such that

$$Q(z)V_q(z) + P(z)V_p(z) = I_p. \quad (13.5.25)$$

This result is called Bezout’s identity. Conversely, (13.5.25) implies that $Q(z)$ and $P(z)$ are left coprime (why?).

Obtaining deeper results about irreducible MFDs

Using (13.5.25) we can develop deeper results on irreducible matrix fraction descriptions.

Lemma 13.5.2. Let $Q^{-1}(z)P(z)$ and $Q_2^{-1}(z)P_2(z)$ be two MFDs representing $H(z)$ and let $Q^{-1}(z)P(z)$ be irreducible. Then $Q_2(z)Q^{-1}(z)$ is a polynomial matrix.

Proof. Since $Q^{-1}(z)P(z)$ is irreducible, (13.5.25) holds for some polynomials $V_p(z)$ and $V_q(z)$. But $Q^{-1}(z)P(z) = Q_2^{-1}(z)P_2(z)$ so that

$$Q(z)V_q(z) + \underbrace{Q(z)Q_2^{-1}(z)P_2(z)}_{P(z)} V_p(z) = I.$$

Sec. 13.5 The Smith-McMillan form 693
This can be rearranged as
\[
Q_2(z)V_q(z) + P_2(z)V_p(z) = Q_2(z)Q'^{-1}(z).
\]
Since the left hand side is a polynomial, this proves the desired result. \(\nabla\nabla\nabla\)

The above lemma shows that we can write
\[
Q_2(z) = L(z)Q(z), \quad P_2(z) = L(z)P(z),
\]
where \(L(z) = Q_2(z)Q'^{-1}(z)\) is a polynomial (an \(\text{lcd}\) of \(Q_2(z)\) and \(P_2(z)\)). In the above lemma if both the MFDs are irreducible, then we can interchange their roles to argue that \(Q(z)Q'^{-1}(z)\) is a polynomial (i.e., \(L^{-1}(z)\) is a polynomial. In other words, \(L(z)\) is not only a polynomial, but is unimodular. Summarizing, we have proved:

\[\star \text{Corollary 13.5.2.} \] Let \(Q_1^{-1}(z)P_1(z)\) and \(Q_2^{-1}(z)P_2(z)\) be irreducible MFDs for the same system \(H(z)\). Then \(Q_2(z)Q'^{-1}(z)\) is a unimodular polynomial matrix. So we can relate the MFDs by
\[
Q_2(z) = V(z)Q_1(z), \quad P_2(z) = V(z)P_1(z),
\]
where \(V(z) = Q_2(z)Q'^{-1}(z)\) is a unimodular polynomial matrix. \(\diamond\)

This gives rise to a few other very valuable conclusions. First
\[
\det Q_2(z) = c[\det Q_1(z)],
\]
where \(c = \det V(z) = \text{constant}\), so that the determinant of \(Q(z)\) is the same (except for scale factor) for all irreducible MFDs of a given system. In particular the order of this determinant is same. From Sec. 13.3 we also know that this order is strictly smaller for irreducible MFDs than for reducible ones. Summarizing we have:

\[\star \text{Corollary 13.5.3.} \] The quantity \([\det Q(z)]\) is the same (except for constant scale factor) for all irreducible MFDs of a fixed system \(H(z)\). This means in particular that the order of \([\det Q(z)]\) is the same for all irreducible MFDs of \(H(z)\). Furthermore, the order of \([\det Q(z)]\) is strictly smaller for irreducible MFDs, than for any reducible MFD. \(\diamond\)

\subsection*{13.5.3 The Smith-McMillan Form for Transfer Matrices}

Let \(H(z)\) be a \(p \times r\) rational transfer matrix representing a causal LTI system. Assume each element \(H_{km}(z)\) has been expressed as \(H_{km}(z) = P_{km}(z)/D(z)\) where \(D(z)\) is the least common multiple of the denominators of \(H_{km}(z)\). Here \(P_{km}(z)\) and \(D(z)\) are polynomials in the variable \(z\). Define the \(p \times r\) matrix \(P(z) = [P_{km}(z)]\), and let (13.5.12) be its Smith decomposition. We can then write
\[
H(z) = \underbrace{W(z)\Lambda(z)}_{p \times p} \underbrace{V(z)}_{r \times r}, \quad (13.5.26)
\]
where

\[
\Lambda(z) = \begin{bmatrix}
\lambda_0(z) & 0 & \ldots & 0 & 0 & \ldots & 0 \\
0 & \lambda_1(z) & \ldots & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \lambda_{p-1}(z) & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0 & 0 & \ldots & 0
\end{bmatrix}.
\] (13.5.27)

Evidently \( \lambda_i(z) = \gamma_i(z)/D(z) \). By canceling off the common factors between \( \gamma_i(z) \) and \( D(z) \) we can write this in irreducible form as

\[
\lambda_i(z) = \frac{\alpha_i(z)}{\beta_i(z)}, \quad 0 \leq i \leq \rho - 1.
\] (13.5.28)

In view of the divisibility property (13.5.15a), the polynomials \( \alpha_i(z) \) and \( \beta_i(z) \) satisfy

\[
\alpha_i(z) | \alpha_{i+1}(z), \quad \beta_{i+1}(z) | \beta_i(z), \quad 0 \leq i \leq \rho - 2.
\] (13.5.29)

Equation (13.5.26) is called the *Smith-McMillan decomposition* of \( H(z) \), and \( \Lambda(z) \) the Smith-McMillan form of \( H(z) \). We summarize these as

\[\blackdiamond\textbf{Theorem 13.5.2. Smith-McMillan form.} \text{Let } H(z) \text{ be a } p \times r \text{ rational transfer matrix representing a causal discrete-time system. Then it can be decomposed into the form (13.5.26) where}
1. \( W(z) \) and \( V(z) \) are unimodular matrix polynomials in \( z \).
2. \( \Lambda(z) \) is a \( p \times r \) diagonal matrix as in (13.5.27) where \( \rho \) is the normal rank of \( H(z) \).
3. The diagonal elements of \( \Lambda(z) \) can be expressed as in (13.5.28) where \( \alpha_i(z) \) and \( \beta_i(z) \) are relatively prime polynomials in \( z \) satisfying (13.5.29). Furthermore the polynomials \( \alpha_i(z) \) and \( \beta_i(z) \) are unique up to a constant scale factor. \[\blacklozenge\]

It is worth emphasizing here that the Smith-form (13.5.14) is developed for a polynomial matrix \( P(z) \), whereas the Smith-McMillan form (13.5.27) is developed for a *causal* rational LTI system \( H(z) \).
Example 13.5.3

Let

\[
H(z) = \begin{bmatrix}
\frac{z^{-3}}{(1 + 2z^{-1})^2(1 + 3z^{-1})^2} & \frac{-z^{-1}}{(1 + 3z^{-1})^2} \\
\frac{z^{-1}}{(1 + 3z^{-1})^2} & \frac{-z^{-1}}{(1 + 3z^{-1})^2}
\end{bmatrix}
\]

(13.5.30)

\[
= \frac{1}{(z + 2)^2(z + 3)^2} \begin{bmatrix}
z & -z(z + 2)^2 \\
z(z + 2)^2 & -z(z + 2)^2
\end{bmatrix}
\]

\[
P(z)
\]

The following is a Smith-decomposition of \( P(z) \):

\[
\begin{bmatrix}
1 \\
(z + 2)^2
\end{bmatrix}
\begin{bmatrix}
z & 0 \\
0 & z(z + 2)^2(z^2 + 4z + 3)
\end{bmatrix}
\begin{bmatrix}
1 & -(z + 2)^2 \\
0 & 1
\end{bmatrix},
\]

(13.5.31)

so that \( H(z) \) can be written as

\[
\begin{bmatrix}
1 \\
(z + 2)^2
\end{bmatrix}
\begin{bmatrix}
z & 0 \\
(z + 2)^2 & 0
\end{bmatrix}
\begin{bmatrix}
\frac{z}{(z + 2)^2(z + 3)^2} & 0 \\
0 & \frac{z^2 + 4z + 3}{(z + 3)^2}
\end{bmatrix}
\begin{bmatrix}
1 & -(z + 2)^2 \\
0 & 1
\end{bmatrix}.
\]

(13.5.32)

\( W(z) \)

\( \Lambda(z) \)

\( V(z) \)

We can now identify \( \alpha_i(z) \) and \( \beta_i(z) \) as

\[
\alpha_0(z) = z, \quad \alpha_1(z) = z(z^2 + 4z + 3)
\]

\[
\beta_0(z) = (z + 2)^2(z + 3)^2, \quad \beta_1(z) = (z + 3)^2.
\]

(13.5.33)

It is clear that the divisibility conditions (13.5.29) are satisfied. This example demonstrates a very important point: the Smith-McMillan form \( \Lambda(z) \) is not necessarily causal even though \( H(z) \) is causal. We can rewrite \( \Lambda(z) \) in (13.5.32) as

\[
\begin{bmatrix}
\frac{z^{-3}}{(1 + 2z^{-1})^2(1 + 3z^{-1})^2} & 0 \\
0 & \frac{z(1 + 4z^{-1} + 3z^{-2})}{(1 + 3z^{-1})^2}
\end{bmatrix}
\]

(13.5.34)

So \( \lambda_0(z) \) has a causal inverse-transform but \( \lambda_1(z) \) does not!
Example 13.5.4

We now demonstrate the theory for a rectangular system with \( p = 2, r = 1 \). Let

\[
\mathbf{H}(z) = \begin{bmatrix} 1 + bz^{-1} \\ \frac{1}{1 + az^{-1}} \end{bmatrix}.
\]  

(13.5.35)

The reader can readily verify that this can be decomposed as

\[
\mathbf{H}(z) = \begin{bmatrix} (z + a)(z + b) & z + \frac{a^2 + b^2 + ab}{a + b} \\ z^2 & z - \frac{ab}{a + b} \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{W}(z) \\ \Lambda(z) \end{bmatrix}
\]  

(13.5.36)

Since \( r = 1 \), \( \mathbf{V}(z) \) is a scalar and \( \mathbf{V}(z) = 1 \) here. Notice that for \( p \times 1 \) systems, \( \Lambda(z) \) is also \( p \times 1 \) and since only its 'diagonal' elements can be nonzero, only the 0th element is nonzero. In this example, \( \lambda_0(z) = z^{-2}/(1 + az^{-1}) \) which happens to be causal.

Irreducible MFDs From Smith-McMillan Decomposition

From the decomposition (13.5.26), it is possible to obtain an irreducible MFD for \( \mathbf{H}(z) \). For this define two diagonal matrices

\[
\Lambda_\beta(z) = \text{diag} [\beta_0(z), \ldots, \beta_{p-1}(z), 1, \ldots, 1], \quad (p \times p)
\]

(13.5.37)

\[
\Lambda_\alpha(z) = \text{diag} [\alpha_0(z), \ldots, \alpha_{p-1}(z), 0, \ldots, 0], \quad (p \times r)
\]

so that

\[
\mathbf{H}(z) = \mathbf{W}(z)\Lambda(z)\mathbf{V}(z) = \underbrace{\mathbf{W}(z)\Lambda_\beta^{-1}(z)\Lambda_\alpha(z)\mathbf{V}(z)}_{\mathbf{Q}_1^{-1}(z)\mathbf{P}_1(z)}.
\]  

(13.5.38)

Since \( \alpha_i(z) \) and \( \beta_i(z) \) are relatively prime, the matrices \( \Lambda_\beta(z) \) and \( \Lambda_\alpha(z) \) are left coprime (Problem 13.13). So by Fact 13.2.1, the matrices defined by \( \mathbf{Q}_1(z) \overset{\Delta}{=} \Lambda_\beta(z)\mathbf{W}^{-1}(z) \) and \( \mathbf{P}_1(z) \overset{\Delta}{=} \Lambda_\alpha(z)\mathbf{V}(z) \) are left coprime proving that the MFD given by \( \mathbf{Q}_1^{-1}(z)\mathbf{P}_1(z) \) is irreducible!

\[\blacklozenge\textbf{Corollary 13.5.4.}\] If \( \mathbf{Q}^{-1}(z)\mathbf{P}(z) \) is an irreducible MFD of \( \mathbf{H}(z) \), then \( [\det \mathbf{Q}(z)] = c_1 \prod_{i=0}^{p-1} \beta_i(z) \) for some constant \( c_1 \). This follows by combining Corollary 13.5.2 with the fact that \( \mathbf{Q}_1^{-1}(z)\mathbf{P}_1(z) \) in (13.5.38) is irreducible.

\[\blacklozenge\]
13.5.4 Structural Interpretations of Unimodular Matrices and Smith-McMillan Forms

FIR system with FIR inverse. Let \( V(z) \) be an \( r \times r \) unimodular polynomial in \( z \). This represents a (noncausal) FIR system with FIR inverse (Fig. 13.5-1). The FIR nature of \( V(z) \) means that a finite length input \( u(n) \) produces a finite length output \( a(n) \). And since \( V^{-1}(z) \) is also FIR, every finite length output of \( V(z) \) is produced by a *unique, finite length, input*. This statement is evidently not true for arbitrary FIR \( V(z) \) (Problem 13.20). Moreover since \( V(z) \) and \( V^{-1}(z) \) have only positive powers of \( z \), it follows that if the input is zero after a certain time \( m \), then so is the output, and vice versa. That is,

\[
u(n) = 0, \forall n > m \iff a(n) = 0, \forall n > m.
\]  

(13.5.39)

![Figure 13.5-1](image)

**Figure 13.5-1** A unimodular system and its inverse.

\[
u(n) \rightarrow V(z) \rightarrow a(n) \text{ (a)}
\]

FIR

\[
a(n) \rightarrow V^{-1}(z) \rightarrow u(n) \text{ (b)}
\]

FIR

Now consider the Smith-McMillan decomposition, which is pictorially represented in Fig. 13.5-2. Here \( V(z) \) is an FIR unimodular prefilter and \( W(z) \) an FIR unimodular post filter. This representation allows us to ‘observe’ the poles and zeros of the elements \( \lambda_i(z) \) from ‘outside’ just by measuring the output \( y(n) \) in response to a cleverly chosen *finite length* input sequence \( u(n) \).†

† For the case of continuous time systems, we cannot attach such a simple and elegant interpretation for the decomposition \( W(s)\Gamma(s)V(s) \). This is due to the fact that the matrices \( W(s) \) and \( V(s) \), which are unimodular polynomials in \( s \), are unrealizable. This in turn has to do with the fact that the transfer function \( s^k \) (\( k \)th order differentiator) represents an unstable system for \( k > 0 \). In contrast, for the discrete time case, \( W(z) \) and \( V(z) \) are not
For example, suppose we wish to 'observe' the properties of $\lambda_0(z)$. This can be done by applying an input $u(n)$ such that the intermediate signal $a(n)$ has $z$-transform

$$A(z) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \quad (13.5.40)$$

The appropriate input is

$$U(z) = V^{-1}(z) \begin{bmatrix} 1 \\ 0 \end{bmatrix} = R(z), \quad (13.5.41)$$

where $R(z)$ is merely the 0th column of $V^{-1}(z)$. Since $V(z)$ is unimodular, the input (13.5.41) is a polynomial in $z$, that is, $u(n)$ is FIR. The output of the system $H(z)$ is

$$Y(z) = W(z)A(z)V(z)U(z)$$

$$= W(z)\begin{bmatrix} \lambda_0(z) \\ 0 \end{bmatrix} = W(z) \begin{bmatrix} \lambda_0(z) \\ 0 \end{bmatrix} \quad (13.5.42)$$

$$= \lambda_0(z)W_0(z) = \frac{\alpha_0(z)W_0(z)}{\beta_0(z)}$$

In (13.5.42) the quantity $W_0(z)$, which is the 0th column of $W(z)$, is a polynomial in $z$. Since $W(z)$ is unimodular, this column cannot have a factor of the form $(z - z_p)$ common to all its elements because this would imply that $[\det W(z)]$ has a factor $(z - z_p)$ violating unimodularity. So none of the factors of $\beta_0(z)$ is canceled in (13.5.42).

Summarizing, we can say that $z_p$ is a pole of $Y(z)$ if and only if it is a pole of $\lambda_0(z)$. And $z_0$ is a zero of $Y(z)$ if and only if it is a zero of $\lambda_0(z)$. So all the crucial dynamical properties of each of the elements $\lambda_i(z)$ can be communicated to the output $y(n)$ in this manner in finite time since an FIR input $u(n)$ will convey this information to the output terminal! The relation between poles, zeros and the polynomials $\alpha_i(z), \beta_i(z)$ will made more precise in the next few sections.

### 13.6 POLES OF TRANSFER MATRICES

The point $z_p$ is said to be a pole of $H(z)$ if it is a pole of some element $H_{km}(z)$ in $H(z)$. Recall from Sec. 2.2 that a causal system cannot have any pole at $z = \infty$, and that $H(\infty) = h(0)$ where $h(n)$ is the causal impulse response.

only realizable but also FIR (which is the best type of systems we can hope for!)

Sec. 13.5 The Smith-McMillan form 699
In terms of time domain significance, poles are related to the solutions of the homogeneous difference equation describing the system (Problem 13.22). In this section we will present several manifestations of a pole, both in the \( z \)-domain and time domain. The reader interested only in the main result can proceed to Theorem 13.6.1 and Lemma 13.6.4.

\[ \text{Lemma 13.6.1. Poles of } H(z) \text{ and zeros of } [\det Q(z)]. \text{ Let } H(z) \text{ be a } p \times r \text{ rational LTI system with irreducible MFD } Q^{-1}(z)P(z). \text{ Then } z_p \text{ is a pole of } H(z) \text{ if and only if } [\det Q(z_p)] = 0. \]

\[ \text{Proof. We have } \]
\[ H(z) = Q^{-1}(z)P(z) = S(z)P(z)/[\det Q(z)], \quad (13.6.1) \]
where \( S(z) \) is the adjugate of \( Q(z) \). Since \( S(z) \) and \( P(z) \) are polynomials in \( z \), we see that any factor which occurs in the denominator of any \( H_{km}(z) \) must also be a factor of \([\det Q(z)]\). So if \( z_p \) is a pole of \( H(z) \) then it is a zero of \([\det Q(z)]\).

Conversely suppose \( z_p \) is a zero of \([\det Q(z)]\). From (13.6.1) we see that, in general, this can cancel with every element in the numerator matrix \( S(z)P(z) \). So it is not obvious that \( z_p \) is a pole. The reasoning is somewhat subtler, and depends on the fact that the MFD is irreducible. Irreducibility means that \( Q(z) \) and \( P(z) \) are left coprime so that (13.5.25) holds for some polynomials \( V_q(z) \) and \( V_p(z) \), that is,
\[ V_q(z) + H(z)V_p(z) = \frac{\text{Adj } Q(z)}{[\det Q(z)]} \quad (13.6.2) \]
Now if \( z_p \) is a zero of \([\det Q(z)]\), then \([\text{Adj } Q(z)]\) cannot completely cancel this zero (Fact 13.2.2). So \( z_p \) is a pole of the RHS of (13.6.2), and hence of the LHS. But since \( V_q(z) \) and \( V_p(z) \) are polynomial matrices, this implies that \( z_p \) is a pole of \( H(z) \) indeed.

\[ \text{Lemma 13.6.2. Poles of } H(z) \text{ and zeros of } \beta_0(z). \text{ The point } z_p \text{ is a pole of } H(z) \text{ if and only if it is a zero of the polynomial } \beta_0(z) \text{ in the Smith-McMillan form.} \]

\[ \text{Proof. We know from the previous section that (13.5.38) is an irreducible MFD obtained from the Smith-McMillan decomposition. In this MFD, } \]
\[ Q_1(z) = \Lambda_\beta(z)W^{-1}(z), \quad (13.6.3) \]
where \( W(z) \) (hence \( W^{-1}(z) \)) is unimodular, that is, has constant non zero determinant. Clearly, \([\det Q_1(z_p)] = 0 \) if and only if \([\det \Lambda_\beta(z_p)] = 0 \). But \([\det \Lambda_\beta(z)]\) is the product of all \( \beta_i(z) \), and moreover, \( \beta_0(z) \) contains all other \( \beta_i(z) \) as factors [by (13.5.29)]. This shows that \([\det Q_1(z_p)] = 0 \) if and only if \( \beta_0(z_p) = 0 \). By Lemma 13.6.1 we, therefore, conclude that \( z_p \) is a pole if and only if \( \beta_0(z_p) = 0 \).
Time Domain Dynamical Interpretation of a Pole

Qualitatively speaking, $z_p$ is a pole if $H(z_p)$ "blows up." But this statement hardly provides any physical insight about the meaning of a pole.

In Problem 2.4 we saw that the pole of a scalar causal system $H(z)$ has a nice time domain interpretation in terms of causal inputs and outputs. According to this, $z_p$ is a pole if and only if there exists a finite length input $u(n)$ such that the system output takes the form $z_p^n$ for all $n$ greater than some finite integer (Fig. 13.6-1).

For MIMO systems a similar interpretation can be given, as elaborated in the next Lemma, which is a discrete-time version of the result presented in Desoer and Schulman [1974].

Lemma 13.6.3. Time domain meaning of a pole. Let $H(z)$ be a $p \times r$ rational transfer function representing a causal system. Let $z_p \neq 0$. Then $z_p$ is a pole if and only if there exists a finite length input (i.e., FIR input) $u(n)$ such that the output takes the form

$$y(n) = z_p^n v, \quad \text{for all } n > K,$$

for some vector $v \neq 0$ and for some finite integer $K$.

Comments. Why exclude $z_p = 0$? For any rational $H(z)$ we can trivially find FIR input such that output is FIR. For example, if $H(z) = P(z)/Q(z)$, just pick $U(z) = Q(z)$ so that $Y(z) = P(z) = \text{FIR}$. So there exists FIR input such that (13.6.4) holds with $z_p = 0$. This does not necessarily mean that there is a pole at the origin because the above argument holds for any $P(z)/Q(z)$.

Proof of Lemma 13.6.3. First suppose that there exists an FIR input
U(z) producing output y(n) of the above form. Then

\[ Y(z) = \frac{v}{1-z_p z^{-1}} + Y_1(z), \]  

(13.6.5)

where \( Y_1(z) \) is FIR. So \( Y(z) \) has the pole \( z_p \). Since \( Y(z) = H(z)U(z) \) and \( U(z) \) is FIR, it is clear that \( H(z) \) has a pole at \( z_p \). (This argument uses the assumption \( z_p \neq 0 \).)

Now consider the converse. If \( z_p \) is a pole then an element, say \( H_{00}(z) \) has this pole. If we apply the input \([S(z) \ 0 \ 0 \ 0]^T\) for some FIR \( S(z) \), we obtain \( Y(z) = S(z)H_0(z) \) where \( H_0(z) \) is the 0th column of \( H(z) \). We can always choose \( S(z) \) to cancel all the poles of \( H_0(z) \) except \( z_p \). We then have \( Y(z) = N(z)/(1-x^{-1}z_p) \) for FIR \( N(z) \). This can be written in the form (13.6.5) for FIR \( Y_1(z) \) and nonzero \( v \), so that (13.6.4) follows. \( \nabla \nabla \nabla \)

**Summary of Various Manifestations of a Pole**

\( \diamond \) **Theorem 13.6.1.** Let \( H(z) \) be a \( p \times r \) causal rational transfer function, with irreducible MFD \( Q^{-1}(z)P(z) \). Also let \( A \) be the state transition matrix of some minimal realization of \( H(z) \). Finally let the Smith-McMillan form of \( H(z) \) be as in (13.5.27), with \( \alpha_i(z) \) and \( \beta_i(z) \) relatively prime. Then the first four statements below are equivalent. If \( z_p \neq 0 \), then all the five statements are equivalent.

1. \( z_p \) is a pole of \( H(z) \).
2. \( z_p \) is a zero of \([\det Q(z)]\).
3. \( z_p \) is a zero of \( \beta_0(z) \).
4. \( z_p \) is an eigenvalue of \( A \).
5. There exists an FIR input such that the output takes the form \( z_p^n v \) for all \( n \) greater than a finite integer, for some \( v \neq 0 \). \( \diamond \)

**Proof.** It only remains to prove that statement 4 is equivalent to the others. From (13.4.14) we already know that if \( z_p \) is a pole it is an eigenvalue of \( A \). It only remains to prove the converse under the condition that \((A,B,C,D)\) is minimal. Now minimality implies reachability. This means that we can apply a finite length input \( u(0), \ldots, u(N-1) \) (with zero initial state) and reach the state \( x(N) = v \), where \( v \) is an eigenvector of \( A \) corresponding to eigenvalue \( \lambda \). We then have

\[ x(N+1) = Ax(N) = Av = \lambda v, \quad x(N+2) = Ax(N+1) = \lambda Av = \lambda^2 v, \]

and so on, so that \( x(N+k) = \lambda^k v, k \geq 1 \). Thus the output is

\[ y(N+k) = \lambda^k Cv = \lambda^{N+k}[Cv/\lambda^N], \quad k \geq 1, \]

(13.6.6)

for the case where \( \lambda \neq 0 \). Since \((C,A)\) is observable, \( Cv \neq 0 \) (by PBH test). So \( y(n) = \lambda^n w \) (w \( \neq 0 \)) for \( n > N \). Thus \( \lambda \) is a pole (by Lemma 13.6.3). If \( \lambda = 0 \), the result is still true. See Problem 13.23. \( \nabla \nabla \nabla \)
Order of a pole

We say that a scalar system \( H(z) \) has a pole \( z_p \) of order \( K \) if \( (z - z_p)^K \) appears in the denominator. For MIMO systems, we use the Smith-McMillan form to define pole order. Recall that the functions \( \lambda_i(z) \) have the irreducible form (13.5.28) where the \( \alpha_i(z) \) and \( \beta_i(z) \) satisfy (13.5.29). This means in particular that \( \beta_0(z) \) contains all other \( \beta_i(z) \) as factors. If \( z_p \) is a zero of \( \beta_0(z) \) with order \( K \), we say that \( H(z) \) has a pole of order \( K \) at \( z = z_p \).

The reason for this definition can be seen from the structural meaning of the Smith-McMillan decomposition: we can always find an FIR input (13.5.41) such that the output has the form (13.5.42). Now let us write \( \beta_0(z) = \beta_0'(z)(z - z_p)^K \) so that \( \beta_0'(z) \) is a polynomial in \( z \) which does not vanish at \( z = z_p \). If we replace the above FIR input with

\[
U(z) = \beta_0'(z)R(z),
\]

it is clear that the output is replaced with

\[
Y(z) = \frac{\alpha_0(z)W_0(z)}{(z - z_p)^K}
\]

And since \( \alpha_0(z)W_0(z) \) is nonzero for \( z = z_p \), there is no cancelation of factors in (13.6.8).

Conversely, suppose an FIR input \( U(z) \) produces output of the form (13.6.8). We know \( Y(z) = W(z)\Lambda(z)V(z)U(z) \). Since \( U(z) \), \( V(z) \), and \( W(z) \) are FIR, it is clear that at least one element of \( \Lambda(z) \) has the factor \( (z - z_p)^K \) in the denominator. This implies, in particular, that \( \beta_0(z) \) has the factor \( (z - z_p)^K \). We summarize these discussions as follows:

\[\blacklozenge\text{Lemma 13.6.4. Dynamical meaning of order of a pole. Let } H(z) \text{ be some } p \times r \text{ causal rational discrete-time system and let } z_p \neq 0. \text{ Then } z_p \text{ is a pole of order } \geq K \text{ if and only if there exists an FIR input } U(z) \text{ such that the output takes the form } Y(z) = Y_1(z)/(z - z_p)^K \text{ where } Y_1(z) \text{ is FIR with } Y_1(z_p) \neq 0. \]

\[\blacklozenge\]

13.7 ZEROS OF TRANSFER MATRICES

The definition of poles of a transfer matrix \( H(z) \) is really very simple; we say that \( z_p \) is a pole if it is a pole of some element \( H_{km}(z) \). This gives rise to several equivalent meanings for a pole as summarized above. But the definition of a zero is more complicated as elaborated next.

Fine Points in the Definition of Zeros

One can define a zero \( z_0 \) to be such that \( H(z_0) = 0 \), but this is too restricted. It means that every element \( H_{km}(z) \) has a zero at \( z_0 \). Many practical systems will simply fail to satisfy this definition for any \( z \). For example, the system in (13.5.35) is not zero for any \( z \).
A second possibility would be to take a hint from Theorem 13.6.1 for poles, and define $z_0$ to be a zero if the determinant of $P(z_0)$ is zero (where $Q^{-1}(z)P(z)$ is an irreducible MFD). This again is meaningless if $P(z)$ is not square (i.e., if $p \neq r$). A natural extension of this idea, however, is to define $z_0$ to be a zero if $P(z_0)v = 0$ for some vector $v \neq 0$. In terms of time domain, this means that there exists an input of the form
\[ u(n) = z_0^n v, \]
which results in zero output. To see this note that (2.2.13) implies
\[ y(n) = H(z_0)vz_0^n = Q^{-1}(z_0)P(z_0)vz_0^n = 0. \]
In other words, an exponential input aligned in the direction of $v$ produces zero output. This behavior is very similar to the scalar case (where $v$ would just be a nonzero scalar).

Upon deeper thinking this definition has some triviality associated with it. For example suppose $H(z) = [1 \ z^{-1}]$. Given any point $z = z_0$, consider the input $z_0^n v$ where $v = \begin{bmatrix} 1 \\ -z_0 \end{bmatrix}$. Since $H(z_0)v = 0$, this will result in zero output for all $n$. More generally, if the normal rank $\rho$ of $P(z)$ is less than $r$, then according to this definition any $z_0$ is a zero of the system!

An improved definition, and the most commonly accepted one, says that $z_0$ is a zero if the rank of $P(z_0)$ is less than the normal rank of $P(z)$. We will rephrase this in terms of the Smith-McMillan form, so that we can also explain the meaning of ‘order’ of a zero.

**Zeros in terms of Smith-McMillan form**

We say that $z_0$ is a zero of $H(z)$ if it is a zero of $\alpha_i(z)$ for some $i$. In view of the divisibility property (13.5.29), this is equivalent to the condition $\alpha_{\rho-1}(z_0) = 0$. We say that $z_0$ is a zero of order $K$ if $\alpha_{\rho-1}(z) = (z - z_0)^K \alpha'(z)$ where $\alpha'(z)$ is a polynomial in $z$ with $\alpha'(z_0) \neq 0$. So if $z_0$ is a zero then the rank of $P(z_0)$ falls below the normal rank $\rho$.

**Zeros at infinity?** In Sec. 2.2 we saw that a causal system cannot have a pole at $z = \infty$. However, a causal system might have a zero at $\infty$, as in $H(z) = z^{-1}$. The definitions in the above paragraph exclude this situation. This is because $\alpha_{\rho-1}(z)$, being a nonzero polynomial in $z$, cannot be zero at $\infty$. Furthermore, we do not obtain meaningful answer for the rank of the polynomial $P(z)$ if we set $z = \infty$. In order to study the behavior at $z = \infty$, one usually performs the mapping $z \to 1/z$ and studies the behavior at the origin. The properties of zeros to be presented in this section hold only for zeros at finite points.

As explained above, we cannot attach a nontrivial time domain significance to zeros, unless the normal rank is full [i.e., unless $\rho = \min(p, r)$]. Under the full normal-rank assumption, some results in this connection have been derived in Desoer and Schulman [1974], for continuous-time systems.
These are restated below in discrete-time language. We show that if \( z_0 \) is a zero then there exists an IIR input with a pole at \( z_0 \), such that the output is FIR. In other words, the zero of \( H(z) \) cancels the 'pole' of the input. More precisely we have:

\[ \textbf{Theorem 13.7.1. Time domain dynamics of zeros.} \]  
Let \( H(z) \) be a \( p \times r \) rational causal LTI system with \( p \geq r \), and let the normal rank be \( \rho = r \). Let \( z_0 \neq 0 \). Then \( z_0 \) is a zero of order \( \geq K \) (as defined above in terms of \( \alpha_{\rho-1}(z) \)), if and only if there exists an input of the form

\[
U(z) = \frac{G(z)}{(z - z_0)^K}, \quad G(z) \text{ FIR, } G(z_0) \neq 0, \quad (13.7.1)
\]

such that the output \( Y(z) \) is FIR.

\[ \textbf{Remarks.} \] If we assume \( K = 1 \) for a moment, we can express the above input in the time domain as

\[
u(n) = cz_0^\rho U(n) + f(n)
\]

where \( c \) is a constant vector, \( U(n) \) is the unit step function, and \( f(n) \) is a finite length sequence. Thus the input looks like the exponential sequence \( cz_0^\rho \) for sufficiently large \( n \). For sufficiently large \( n \) the output is, however, zero (because it is FIR according to the theorem)! This is the time domain interpretation of a zero, and is based only on one sided input and output sequences. This is unlike the traditional interpretation which depends on a two sided exponential input \( z_0^n v \). The point \( z_0 = 0 \) is excluded in the above result for reasons similar to those discussed after Lemma 13.6.3.

\[ \textbf{Proof of Theorem 13.7.1.} \] First assume that \( z_0 \) is a zero of order \( \geq K \) that is, \( \alpha_{\rho-1}(z) = (z - z_0)^K \alpha'_\rho(z) \), where \( \alpha'_\rho(z) \) is a polynomial in \( z \). We have to prove the existence of an input with above form such that the output is FIR. From the Smith-McMillan decomposition we know

\[
Y(z) = W(z)\Lambda(z)V(z)U(z). \quad (13.7.2)
\]

Since \( p \geq r = \rho \),

\[
\Lambda(z) = \begin{bmatrix}
\lambda_0(z) & 0 & 0 & \ldots & 0 \\
0 & \lambda_1(z) & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & \lambda_{\rho-1}(z) \\
0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 0
\end{bmatrix}. \quad (13.7.3)
\]
Suppose we pick \( U(z) \) such that

\[
V(z)U(z) = \begin{bmatrix}
0 \\
0 \\
\vdots \\
0 \\
\beta_{\rho-1}(z) \\
(z - z_0)^K
\end{bmatrix}.
\]  
(13.7.4)

Then

\[
Y(z) = \frac{W_{\rho-1}(z)\alpha_{\rho-1}(z)\beta_{\rho-1}(z)}{\beta_{\rho-1}(z)(z - z_0)^K} = W_{\rho-1}(z)\alpha'_{\rho-1}(z),
\]  
(13.7.5)

which is FIR since \( W(z) \) is unimodular. (Here \( W_{\rho-1}(z) \) is the \((\rho - 1)\)th column of \( W(z) \). The input satisfying (13.7.4) is of the form (13.7.1) where

\[
G(z) = T_{\rho-1}(z)\beta_{\rho-1}(z).
\]  
(13.7.6)

The quantity \( T_{\rho-1}(z) \) is the \((\rho - 1)\)th column of the inverse of \( V(z) \). Since \( V(z) \) is unimodular, its inverse and hence \( T_{\rho-1}(z) \) are FIR, and furthermore \( T_{\rho-1}(z_0) \neq 0 \). Also \( \beta_{\rho-1}(z) \) is relatively prime to \( \alpha_{\rho-1}(z) \) so that \( \beta_{\rho-1}(z_0) \neq 0 \). So \( G(z) \) is FIR with \( G(z_0) \neq 0 \), proving the desired result.

Conversely, suppose an input of the form (13.7.1) produces output \( Y(z) \). Now

\[
Y(z) \text{ is FIR } \Rightarrow W^{-1}(z)Y(z) \text{ is FIR (since } W(z) \text{ unimodular)}
\]
\[
\Rightarrow \frac{\Lambda(z)V(z)G(z)}{(z - z_0)^K} \text{ is FIR.}
\]  
(13.7.7)

Since \( G(z) \) is FIR with \( G(z_0) \neq 0 \), we have \( V(z_0)G(z_0) \neq 0 \) (because \( V(z) \) is unimodular). The FIR nature of \( Y(z) \) implies then that there is at least one diagonal element \( \lambda_i(z) \) in \( \Lambda(z) \), having the factor \((z - z_0)^K\) in the numerator. (This reasoning uses the assumption \( z_0 \neq 0 \)). This means that \( z_0 \) is a zero of order (at least) \( K \).

For the \( p < r \) case the above result does not hold because for every \( z \), there exists some nonzero vector \( v \) such that \( P(z)v = 0 \). For this case, a more useful dynamical significance is developed in Problem 13.21.

**Case of Square Matrices: the Cleanest Case**

Suppose \( H(z) \) is \( p \times p \) so that \( P(z) \) in an irreducible MFD \( Q^{-1}(z)P(z) \) is \( p \times p \). If in addition \( P(z) \) has normal rank \( \rho = p \), then \( z_0 \) is a zero if and only if \( [\det P(z_0)] = 0 \). The dynamical significance offered by Theorem 13.7.1 continues to hold in this case.

So this is the cleanest case in the sense that the meaning of zeros and poles in terms of the MFD quantities are very similar: \( z_p \) is a pole if, and
only if, \(|\det Q(z_p)| = 0\) whereas \(z_0\) is a zero if, and only if, \(|\det P(z_0)| = 0\). Also remember that \(z_p\) and \(z_0\) are zeros of \(\beta_0(z)\) and \(\alpha_{p-1}(z)\), respectively.

**Poles and zeros at same place.** For an SISO system a pole \(z_p\) and a zero \(z_0\) cannot be at the same place (i.e., we cannot have \(z_p = z_0\)) because these would simply cancel. In the MIMO case, a pole and a zero can exist at the same point without canceling. An example is

\[
H(z) = \text{diag} \left[ 1 + az^{-1}, \frac{1}{1 + az^{-1}} \right].
\]  \hspace{1cm} (13.7.8)

In such cases, \(\prod \alpha_i(z)\) and \(\prod \beta_i(z)\) are not coprime even though \(\alpha_i(z)\) and \(\beta_i(z)\) are coprime for each \(i\).

### 13.8 Degree of a Transfer Matrix

Let \(H(z)\) be a \(p \times r\) causal rational system. In Sec. 13.3 we defined the degree \(\mu\) of \(H(z)\) to be the minimum number of delay elements required for its implementation. Letting \(N\) denote the size of the state vector \(x(n)\) in any realization of \(H(z)\), we showed that \(N = \mu\) (i.e., the realization is minimal) if and only if \((C, A)\) is observable and \((A, B)\) is reachable.

There is an extremely important result in realization theory [Kalman, 1965] which expresses the degree \(\mu\) of \(H(z)\) in terms of the quantities \(\beta_i(z)\) in the Smith-McMillan decomposition. Recall that \(\beta_i(z)\) and \(\alpha_i(z)\) are polynomials in \(z\). With \(\mu_i\) denoting the order of \(\beta_i(z)\), the degree of \(H(z)\) is

\[
\mu = \sum_{i=0}^{\rho-1} \mu_i.
\]  \hspace{1cm} (13.8.1)


**Warning.** It is important to notice that this result applies only under the following conditions:

1. \(H(z)\) is causal (otherwise it cannot be realized with delay elements alone, and our definition of degree is meaningless).
2. The matrices \(W(z)\) and \(V(z)\) are polynomials in \(z\), and so are the quantities \(\alpha_i(z)\) and \(\beta_i(z)\). If these are re-expressed as polynomials in \(z^{-1}\), the above result does not hold.

**Relation to \([\det Q(z)]\).** An immediate consequence of (13.8.1) can be obtained from (13.5.38), which gives an irreducible MFD for \(H(z)\). The order of \([\det Q_i(z)]\) is equal to \(\mu\) (as the determinant of \(W(z)\) is constant). But the order of \([\det Q(z)]\) is the same in all irreducible MFDs of \(H(z)\) (Corollary 13.5.3). So we conclude that \(\mu = \text{order of } [\det Q(z)]\). We now summarize all results pertaining to degree.

**Theorem 13.8.1. On degree of \(H(z)\).** Let \(H(z)\) be a causal rational \(p \times r\) discrete-time system. Then the following are true.
1. Let (13.5.27) be the Smith-McMillan form. Then the degree $\mu$ is given by (13.8.1) where $\mu_i$ is the order of $\beta_i(z)$.

2. Let $Q^{-1}(z)P(z)$ be an irreducible MFD for $H(z)$, where $P(z)$ and $Q(z)$ are polynomials in $z$. Then, $\mu = \text{order of } [\det Q(z)]$.

3. Finally a realization $(A,B,C,D)$ for $H(z)$ is minimal (i.e., has the smallest possible size $\mu \times \mu$ for $A$) if, and only if, $(C,A)$ is observable and $(A,B)$ is reachable. ◊

**Degree of $H(z)$ Versus Degree of $[\det H(z)]$**

Suppose $H(z)$ is $p \times p$ with irreducible MFD $Q^{-1}(z)P(z)$. We know that the order of $[\det Q(z)]$ is equal to the degree $\mu$ of $H(z)$. Does the degree of $[\det H(z)]$ have any role in this connection? Can we relate it to $\mu$? We now provide an answer.

For example if $H(z) = \begin{bmatrix} 1 & 0 \\ f(z) & 1 \end{bmatrix}$ where $f(z)$ is causal, the degree of $H(z)$ is equal to that of $f(z)$ whereas the degree of $[\det H(z)]$ is zero regardless of $f(z)$. So the degree of $[\det H(z)]$ can be as small as zero. Next, how large can it be? It turns out that it cannot be greater than $\mu$. To see this note that (13.5.26) yields

$$\det H(z) = c\alpha(z)/\beta(z)$$

where $c$ is constant, $\alpha(z) = \prod \alpha_i(z)$ and $\beta(z) = \prod \beta_i(z)$. We know the order of $\beta(z)$ equals $\mu$. The order of $\alpha(z)$ is $\leq \mu$ because of causality (remember that $\alpha(z)$ and $\beta(z)$ are polynomials in $z$). So, the degree of (13.8.2) is at most $\mu$. It is less than $\mu$ if there are cancelations in the ratio (13.8.2). Summarizing, we have proved

$$0 \leq \text{deg } [\det H(z)] \leq \mu$$

where $\mu$ is the McMillan degree of $H(z)$.

### 13.9 FIR Transfer Matrices

The above results apply for FIR as well as IIR systems. In this text we frequently deal with MIMO FIR systems. These arise, for example, in multirate filter banks in the form of polyphase matrices for the analysis and synthesis banks. If the filter bank is FIR then these polyphase matrices are FIR and it is important to understand their system-theoretic properties.

**Degree of FIR Transfer Matrices**

Suppose $H(z)$ is $p \times r$ causal FIR so that

$$H(z) = h(0) + h(1)z^{-1} + \ldots + h(K)z^{-K},$$

with $h(K) \neq 0$. The quantity $K$ is called the order of $H(z)$, and $K + 1$ the length. If $\mu$ is the degree of $H(z)$ then $\mu \geq K$ because we require at least $K$ delays to realize this system.
Remarks

1. Examples of the form \( H(z) = z^{-1}I \) demonstrate that \( \mu \) can be larger than \( K \).

2. We also know from Example 13.3.2 that if \( H(z) = h(K)z^{-K} \) then the degree is equal to \( sK \) where \( s \) is the rank of the (possibly rectangular) matrix \( h(K) \).

3. For the special case of \( p \times 1 \) FIR systems the transfer matrix is a column vector. This arises, for example, when \( H(z) \) is an analysis bank with \( p \) analysis filters. In this case the degree \( \mu \) is equal to \( K \) because \( h(K) \) has rank one. \( H(z) \) can then be implemented as in Fig. 13.9-1 requiring \( K \) delays. Similarly if \( H(z) \) is a row vector (as in a synthesis bank) then \( \mu = K \).

![Figure 13.9-1 Implementation of \( p \times 1 \) FIR \( H(z) \).](image)

State Space Description of FIR Systems

Given a structure for a \( p \times r \) causal FIR transfer matrix, one can find a state space description \( (A, B, C, D) \) in the usual manner. Assuming that the realization is minimal, the eigenvalues of \( A \) are the poles, all of which lie at the origin of the \( z \) plane. So, all eigenvalues of \( A \) are zero. Some of the other properties are summarized in the next lemma.

\[ \text{Lemma 13.9.1. On FIR state space description.} \]

Consider a minimal realization \( (A, B, C, D) \) of a \( p \times r \) causal FIR system with order \( K \) [eqn. (13.9.1)]. Then, \( A \) has all eigenvalues equal to zero. Moreover, (a) \( A^KB = 0 \), (b) \( CA^K = 0 \), and (c) \( A^N = 0 \), where \( N \) is the state space dimension (i.e., \( A \) is \( N \times N \)). Conversely, if \( A^N = 0 \), then \( (A, B, C, D) \) represents an FIR system.

\[ \text{Proof. We know} \ h(n) = CA^{n-1}B \ \text{for} \ n > 0. \ \text{Since} \ h(n) = 0 \ \text{for} \ n > K, \ \text{we have} \ CA^iB = 0, \ \text{for} \ i \geq K. \ \text{So we can write, in particular,} \]

\[
\begin{bmatrix}
  C \\
  CA \\
  \vdots \\
  CA^{N-1}
\end{bmatrix} A^K B = 0.
\]  

(13.9.2)
Since \((A, B, C, D)\) is minimal, \(S_{C, A}\) has full column rank \(N\), so this implies \(A^KB = 0\). The proof of \(CA^K = 0\) is similar. Next \(A^KB = 0\) also implies \(A^NB = 0\), since \(N \geq K\). This means

\[
A^N \left[ \begin{array}{ccc}
B & AB & \ldots & A^{N-1}B \\
\end{array} \right] = 0. \tag{13.9.3}
\]

Since \(\mathcal{R}_{A, B}\) has full row rank \(N\), this implies \(A^N = 0\). A second way to prove \(A^N = 0\) is to note that the characteristic equation is \(\lambda^N = 0\), and \(A\) has to satisfy this equation (Cayley-Hamilton theorem; Appendix A).

Conversely, suppose \(A^N = 0\). We know that the eigenvalues of \(A^N\) are given by \(\lambda_i^N\), so that we have \(\lambda_i^N = 0\), that is, \(\lambda_i = 0\). So all eigenvalues of \(A\) are equal to zero [i.e., poles of \(H(z)\) are at the origin] proving that \(H(z)\) is FIR.

Smith-McMillan Form for the FIR Case

We now present some examples of FIR Smith-McMillan forms, highlighting the main features.

Example 13.9.1

Consider

\[
H(z) = \begin{bmatrix}
    z^{-1} + z^{-2} & z^{-1} \\
    2 + 3z^{-2} & 2(1 + z^{-1})^2
\end{bmatrix}
\]

which is \(2 \times 2\) causal FIR. To find the Smith-McMillan form we first express every entry as a ratio of polynomials in \(z\). Thus

\[
H(z) = \frac{1}{z^2} \left[\begin{array}{c}
1 + \frac{1}{3 + 2z^2} & \frac{z}{2(1 + z)^2} \\
\end{array}\right] \tag{13.9.4}
\]

The Smith-form for this \(P(z)\) has been worked out in Example 13.5.1. The Smith-McMillan form for \(H(z)\) is

\[
\Lambda(z) = \begin{bmatrix}
    \frac{1}{z^2} & 0 \\
    0 & \frac{2 + 3z + 6z^2}{z^2}
\end{bmatrix} \tag{13.9.5}
\]

so that

\[
\alpha_0(z) = 1, \ \alpha_1(z) = 2 + 3z + 6z^2, \tag{13.9.6}
\]

\[
\beta_0(z) = z^2, \ \beta_1(z) = z^2.
\]

From these we have \(\mu_0 = 2, \mu_1 = 2\) so that the degree is \(\mu = 2 + 2 = 4\).
Example 13.9.2.

Consider the causal FIR system

$$H(z) = \begin{bmatrix} 1 & 0 \\ z^{-1} & 1 \end{bmatrix}. \tag{13.9.7}$$

This obviously has degree one. It is easily verified that

$$H(z) = \begin{bmatrix} z & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} z^{-1} & 0 \\ 0 & z \end{bmatrix} \begin{bmatrix} 1 & z \\ 0 & 1 \end{bmatrix}, \tag{13.9.8}$$

so that

$$\alpha_0(z) = 1, \alpha_1(z) = z,$$

$$\beta_0(z) = z, \beta_1(z) = 1. \tag{13.9.9}$$

So $\mu = 1 + 0 = 1$ as expected. The main point to notice in this example is that all elements $\lambda_i(z)$ are powers of $z$. From (13.9.8) we see that the determinant of $H(z)$ is constant, which is also obvious by inspection of (13.9.7). This $H(z)$ therefore happens to be a unimodular polynomial in the variable $z^{-1}$. Note that we have both positive and negative powers of $z$ in $\lambda_i(z)$. This is fine because $\Lambda(z)$ can be noncausal for causal $H(z)$.

Since $\alpha_1(z) = \beta_0(z) = z$, we see that the system has a pole as well as a zero at $z = 0$. If we write $H(z)$ in irreducible MFD form we expect the rank of $P(z)$ to drop at $z = 0$. This is clearly verified from the following MFD for $H(z)$:

$$H(z) = \begin{bmatrix} 1 & 0 \\ z^{-1} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & z \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 1 & z \end{bmatrix}. \tag{13.9.10}$$

The reader should notice that for the FIR case the polynomials $\beta_i(z)$ are of the form $z^{\mu_i}$. This is consistent with the fact that for causal FIR systems all poles are at $z = 0$. In view of (13.5.29) we also have $\mu_0 \geq \mu_1 \geq \ldots \geq \mu_{\rho-1}$.

13.10 CAUSAL INVERSES OF CAUSAL SYSTEMS

13.10.1 Some Basic Results (FIR and IIR Systems)

Let $H(z)$ be a causal $p \times p$ system (FIR or IIR), and let there exist a causal $p \times p$ inverse $G(z)$, so that

$$G(z)H(z) = I_p. \tag{13.10.1}$$
Denoting the impulse response coefficients of $H(z)$ and $G(z)$ by $h(n)$ and $g(n)$ respectively, we have from (13.10.1),

$$g(0)h(0) = I_p$$

(13.10.2)

so that $g(0)$ and $h(0)$ are nonsingular.

**State Space Description of the Inverse System**

Let $(A, B, C, D)$ be a minimal realization of $H(z)$ so that

$$H(z) = D + C(zI - A)^{-1}B.$$  

(13.10.3)

If $D$ is nonsingular we can apply the matrix-inversion lemma (A.4.7) given in Appendix A, to obtain a causal $p \times p$ inverse system

$$G(z) = D_1 + C_1(zI - A_1)^{-1}B_1,$$  

(13.10.4)

where

$$A_1 = A - BD^{-1}C,$$

$$B_1 = BD^{-1},$$

$$C_1 = -D^{-1}C,$$

$$D_1 = D^{-1}.$$  

(13.10.5)

This result is true as long as $D$ is nonsingular. Since $h(0) = D$, we conclude that a $p \times p$ causal system has a causal inverse if and only if $h(0)$ is nonsingular.

It can be shown (Problem 13.10) that minimality of $(A, B, C, D)$ implies that of $(A_1, B_1, C_1, D)$. This means in particular that if a $p \times p$ causal system $H(z)$ has a causal inverse $G(z)$, then $H(z)$ and $G(z)$ have the same degree. The same is not true of rectangular systems. For example if

$$H(z) = \begin{bmatrix} 1 - z^{-1} \\ z^{-1} \end{bmatrix},$$

(13.10.6)

then

$$G(z) = [1 \quad 1],$$

(13.10.7)

is a valid left inverse (i.e., $G(z)H(z) = 1$). But the degrees of $H(z)$ and $G(z)$ are, respectively, one and zero.

**13.10.2 Causal unimodular systems**

In Sec. 5.6.2 we mentioned that the polyphase matrix $E(z)$ of any causal FIR perfect reconstruction QMF bank can be written as a product of a causal paraunitary system and a causal unimodular system. In Chap. 6 we elaborated on paraunitary systems (which are further studied in Chap. 7).
In this section we will present some details about causal unimodular systems.

The phrase ‘causal unimodular system’ stands for a square matrix polynomial \( H(z) \) in \( z^{-1} \), which is unimodular. This implies that \( H(z) \) is causal FIR, and has a causal FIR inverse. Conversely, suppose \( H(z) \) is a \( p \times p \) causal FIR system with a causal FIR inverse \( G(z) \). Then \( H(z) \) and \( G(z) \) are unimodular polynomials (in \( z^{-1} \)). To see this note that (13.10.1) implies that the product of determinants is unity, and since each determinant can at best be a polynomial in \( z^{-1} \), it has to be a constant. Summarizing, the phrase “causal unimodular system” is synonymous to “casual FIR system with causal FIR inverse.”

State Space Description for Causal Unimodular Systems

With \( (A, B, C, D) \) denoting a state space description of \( H(z) \), the causal unimodular inverse \( G(z) \) has a realization \( A_1, B_1, C_1, D_1 \) given by (13.10.5). Since both \( G(z) \) and \( H(z) \) are FIR, it follows that all the eigenvalues of \( A \) and \( A_1 \) are equal to zero. Conversely, suppose \( H(z) \) is some \( p \times p \) causal system with nonsingular \( h(0) \) so that the causal inverse \( G(z) \) exists. Suppose \( (A, B, C, D) \) is a realization such that \( A \) and \( A - BD^{-1}C \) have all eigenvalues equal to 0. Then, \( H(z) \) as well as its inverse \( G(z) \) are (causal) FIR. In other words \( H(z) \) is causal unimodular. We now summarize these results as follows:

\[ \text{Theorem 13.10.1.} \]

Let \( H(z) \) be some \( p \times p \) causal rational system, and let \( h(0) \) be nonsingular.

1. With \( (A, B, C, D) \) denoting a minimal realization of \( H(z) \), the matrix \( D \) is therefore nonsingular, and a \( p \times p \) causal rational inverse \( G(z) \) exists. The matrices (13.10.5) give a minimal realization of \( G(z) \), and \( G(z) \) has the same McMillan degree as \( H(z) \).
2. If in addition \( H(z) \) is unimodular then so is \( G(z) \), and both \( A \) and \( A_1 \) have all eigenvalues equal to zero.
3. Finally if \( (A, B, C, D) \) is a minimal realization such that \( A \) and \( A_1 \) have all eigenvalues equal to zero, then \( H(z) \) is causal unimodular (that is, causal FIR with a causal FIR inverse).

\[ \text{Example 13.10.1} \]

Let

\[
H(z) = \begin{bmatrix}
1 & 0 & 0 \\
z^{-1} & 1 & 0 \\
1 & z^{-1} & 1
\end{bmatrix}.
\] (13.10.8)

This happens to be a triangular matrix, and \( [\det [H(z)] = 1 \) by inspec-
tion. So this is causal unimodular. The inverse is verified to be

\[
G(z) = \begin{bmatrix}
1 & 0 & 0 \\
-z^{-1} & 1 & 0 \\
-1 + z^{-2} & -z^{-1} & 1
\end{bmatrix},
\]

(13.10.9)

which is also causal unimodular as expected.

Let us see if we can guess the degree of \( H(z) \) by inspection. From (13.10.8) it is obvious that at most two delays are required to implement \( H(z) \) (since \( z^{-1} \) occurs only twice), i.e., the degree \( \leq 2 \). From (13.10.9) we see that \( G(z) \) has degree \( \geq 2 \) (since \( z^{-2} \) appears). But the degrees of \( H(z) \) and \( G(z) \) are the same according to the above theorem, so we conclude that the degree is precisely two!

Even though the degrees of \( H(z) \) and \( G(z) \) are equal, their orders (highest power of \( z^{-1} \) appearing in the transfer function) need not be the same. In our example, the order of \( H(z) \) is unity whereas that of \( G(z) \) is two.

---

**Smith-McMillan Form for Causal FIR Unimodular Systems**

Suppose \( H(z) \) is a \( p \times p \) unimodular polynomial in \( z^{-1} \). What does the Smith-McMillan form \( \Lambda(z) \) look like? Since \( H(z) \) is FIR, \( \beta_i(z) = z^{\mu_i} \). So

\[
\det H(z) = \text{constant} \times z^{-\mu} \prod_{i=0}^{p-1} \alpha_i(z).
\]

(13.10.10)

Since the determinant is required to be constant by unimodularity, \( \alpha_i(z) \) must also have the form \( \alpha_i(z) = z^{m_i} \), with \( \sum_i m_i = \sum_i \mu_i \). This fact can be verified to be true in Example 13.9.2. Summarizing, we have

\[
\Lambda(z) = \begin{bmatrix}
z^{n_0} & 0 & 0 & \cdots & 0 \\
0 & z^{n_1} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & z^{n_{p-1}}
\end{bmatrix},
\]

(13.10.11)

with \( \sum_{i=0}^{p-1} n_i = 0 \). The degree \( \mu \) is equal to \( \sum_i |n_i| \) with the summation carried over negative \( n_i \) only.
**PROBLEMS**

13.1. Verify Parseval’s relation (2.2.12) for vector signals.

13.2. Let \( P(z) \) and \( Q(z) \) be matrix polynomials in \( z \), with the same number of rows.
   a) Let \( L(z) \) be a greatest left common divisor (glcd) of \( P(z) \) and \( Q(z) \). Show then that \( L(z)W(z) \) is also a glcd, for any unimodular \( W(z) \).
   b) Let \( L_1(z) \) and \( L_2(z) \) be two glc ds of \( P(z) \) and \( Q(z) \). Show then that \( L_1(z) = L_2(z)V(z) \) where \( V(z) \) is an appropriate unimodular matrix.
   c) Supply a proof for Fact 13.2.1.

13.3. For the following system:

\[
H(z) = \begin{bmatrix}
\frac{1+2z^{-1}}{3+z^{-1}} & \frac{1}{3+z^{-1}} \\
\frac{1+2z^{-1}}{(2+z^{-1})(3+z^{-1})} & 1
\end{bmatrix},
\]

find a left MFD.

13.4. Consider the following MFD.

\[
H(z) = \left[ \begin{array}{cc}
1 + z & z \\
z & 1 + z
\end{array} \right]^{-1} \left[ \begin{array}{cc}
2 + z & 1 + z \\
z & z
\end{array} \right]_{Q^{-1}(z)}^{P(z)}
\]

Find another MFD \( \tilde{Q}^{-1}(z)\tilde{P}(z) \) for this same system. To avoid trivial answers, make sure that \( Q(z) \neq f(z)Q(z) \) for scalar \( f(z) \).

13.5. For each of the structures shown in Fig. P13-5, write down the state space description \( (A, B, C, D) \). For each case answer the following: (a) Is \( (A, B) \) reachable? (b) Is \( (C, A) \) observable? (c) Is \( (A, B, C, D) \) minimal?

![Figure P13-5(a)]
13.6. For each structure in Fig. P13-5 find out whether A is stable.

13.7. Consider the state space equations (13.4.3), where A is $N \times N$. The pair $(A, B)$ is said to be controllable [Anderson and Moore, 1979] if we can start from any arbitrary initial state $x(0)$ and force $x(N) = 0$, by appropriate choice of $u(0), \ldots, u(N - 1)$. Evidently if $(A, B)$ is reachable (Sec. 13.4.2) it is also controllable. However, the converse is not true.

a) Show that the structure shown in Fig. P13-5(a) is controllable but not reachable.

b) More generally, show that $(A, B)$ is controllable as long as every column of $A^N$ can be expressed as a linear combination of the columns of $R_{A,B}$ defined in (13.4.18).

13.8. Consider the system $H(z) = z^{-1}R$ where $R$ is $M \times M$ with rank $\rho$. Then we can write $H(z) = TSz^{-1}$ where $T$ and $S$ are $M \times \rho$ and $\rho \times M$ respectively, with rank $\rho$. This leads to the implementation of Fig. 13.3-3.

a) Find the state space description $(A, B, C, D)$ of this structure.

b) Show that this structure for $H(z)$ is minimal, that is, we cannot find a structure with fewer than $\rho$ delays.

13.9. Let $H(z)$ be a $p \times r$ causal rational system with real impulse response $h(n)$. Show that there exists a minimal structure with real valued matrices $A, B, C$ and $D$. (Hint. If you prefer, find a nonminimal realization with real multipliers, and apply results of Sec. 13.4.2).

13.10. Let $H(z)$ represent an $M \times M$ causal system and let $(A, B, C, D)$ be a valid
state space description. Assume \( h(0) \) is nonsingular. Then (13.10.5) gives the state space description of the inverse system \( G(z) \). Show that if \( (A, B, C, D) \) is minimal, then so is the system (13.10.5). (You can use PBH test if you wish.)

13.11. For

\[
P(z) = \begin{bmatrix} 1+z & 2+z \\ 2+z & 3+z \end{bmatrix},
\]

find the Smith form \( \Gamma(z) \) as well as diagonalizing unimodular matrices \( W(z) \) and \( V(z) \). Now consider the causal FIR system

\[
H(z) = \begin{bmatrix} z^{-1}+1 & 2z^{-1}+1 \\ 2z^{-1}+1 & 3z^{-1}+1 \end{bmatrix}.
\]

a) Find the Smith-McMillan form \( \Lambda(z) \), as well as diagonalizing unimodular matrices \( W(z) \) and \( V(z) \).

b) What is the McMillan degree of this system?

c) Find an irreducible MFD \( Q^{-1}(z)P(z) \) for \( H(z) \), and evaluate the determinant of \( Q(z) \). Verify that this determinant has order equal to the McMillan degree.

13.12. Consider the causal unimodular system

\[
H(z) = \begin{bmatrix} 1+3z^{-1} & 3 \\ z^{-1} & 1 \end{bmatrix}.
\]

a) Find the Smith-McMillan form \( \Lambda(z) \).

b) What is the McMillan degree of this system?

c) Find an irreducible MFD \( Q^{-1}(z)P(z) \) for the system. Evaluate the determinant of \( Q(z) \) and verify that it has order equal to the McMillan degree.

13.13. Consider the diagonal matrices \( \Lambda_\beta(z) \) and \( \Lambda_\alpha(z) \) in (13.5.37). Assume that \( \alpha_i(z) \) and \( \beta_i(z) \) (which are polynomials in \( z \)) are relatively prime for each \( i \). Show that the matrices \( \Lambda_\beta(z) \) and \( \Lambda_\alpha(z) \) are left coprime.

13.14. Let \( H(z) = I + z^{-1}h_1 \).

a) Let \( h_1^2 = 0 \). Show that \( H(z) \) has the inverse \( I - z^{-1}h_1 \), and hence that \( H(z) \) is causal unimodular.

b) More generally let \( h_1^L = 0 \) for some integer \( L > 0 \). Show that \( H(z) \) is still unimodular, and find the causal FIR inverse!

13.15. The purpose of this problem is to get acquainted with practically all the concepts introduced in this chapter, using a simple example. Consider the following left MFD:

\[
H(z) = \left[ \begin{array}{cc}
1+z & z \\
z & 1+z
\end{array} \right]^{-1} \left[ \begin{array}{cc}
2+z & 1+z \\
z & 1+z
\end{array} \right]_{Q^{-1}(z) \ P(z)}
\]

a) What is the normal rank of \( P(z) \)?

b) Work out the four elements \( H_{km}(z), 0 \leq k, m \leq 1 \) of the transfer matrix \( H(z) \).
c) Find the causal impulse response matrix $h(n)$ corresponding to $H(z)$. Express all the entries $h_{km}(n)$ in closed form.
d) Compute the coefficients $h(0)$ and $h(1).
e) Is the causal system stable?
f) Find $\det Q(z)$. Hence argue that the system $H(z)$ has degree one.
g) Argue that the above MFD is irreducible.
h) Find a minimal structure (i.e., structure with only one delay) for $H(z)$. Write down the state space description $(A, B, C, D)$ for the structure.
i) Find the Smith-form $\mathbf{G}(z)$ for $P(z)$ by first computing all appropriate minors, and then using (13.5.15b). Also find the Smith-McMillan form $\mathbf{A}(z)$ for $H(z)$. Using the formula (13.8.1), verify again that the degree is one.
j) Using the Smith-McMillan form identify the poles and zeros of $H(z)$. These should agree with the zeros of $\det Q(z)$ and $\det P(z)$ respectively (since the MFD is irreducible). Verify that this is so.

13.16. Let $P$ be $p \times r$ with rank $r$. We know that $P^\dagger P > 0$. However, in general we cannot claim $P^TP > 0$ even if it turns out to be Hermitian (unless, of course, $P$ is real). Demonstrate this with an example.

13.17. Consider an $r$ input $p$ output system with input and output denoted as $x(n)$ and $y(n)$, respectively. Suppose the following properties are satisfied:

a) The response to a shifted input $x(n - K)$ is equal to $y(n - K)$, and this is true for all $K$ and all possible input sequences $x(n)$.
b) If the responses to $x_a(n)$ and $x_b(n)$ are equal to $y_a(n)$ and $y_b(n)$, then the response to $\alpha x_a(n) + \beta x_b(n)$ is equal to $\alpha y_a(n) + \beta y_b(n)$. And this is true for all scalars $\alpha, \beta$ and for all possible pairs of inputs $x_a(n)$ and $x_b(n)$.

Then show that the input output behavior of the system can be characterized by the convolution relation (2.2.7). Conversely, if (2.2.7) holds prove that the above two conditions are satisfied. This shows that the above two conditions can be taken as the defining properties of a MIMO LTI system.

13.18. Recall the relation between the impulse response $h(k)$ and state space descriptions, given in (13.4.13). This was derived by first obtaining the expression (13.4.12) and then comparing with the convolution sum. A second procedure would be to start from (13.4.10b) and express $(zI - A)^{-1}$ as a power series. Rederive (13.4.13) using this idea. (Hint: Note that $\sum_{k=0}^{L} A^k = (I - A)^{-1}(I - A^{L+1})$. Assume $A$ is stable so that $A^n \to 0$ as $n \to \infty$.)

13.19. Let $H(z)$ be a $M \times 1$ rational causal system (i.e., an $M$-channel analysis bank). Show that it can always be rewritten as

$$H(z) = W(z) \begin{bmatrix} \lambda_0(z) \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

where $W(z)$ is an $M \times M$ unimodular polynomial in $z$, and $\lambda_0(z)$ is a rational transfer function.
13.20. Consider the FIR filter \( H(z) = (z + a)(z + b) \). Suppose we wish the output \( y(n) \) to be such that \( Y(z) = (z + b) \), i.e., FIR. Find an appropriate input. Show that there does not exist an FIR input which would result in this output.

13.21. Let \( H(z) = Q^{-1}(z)P(z) \) be a \( p \times r \) causal rational system. In Sec. 13.7 we studied the meaning of a transmission zero, and its dynamical interpretation (Theorem 13.7.1) for the case \( p \geq r \). We now consider the case when \( p < r \). In this case the normal rank of \( P(z) < r \) so that for any value of \( z_0 \) we can find \( v \neq 0 \) such that \( P(z_0)v = 0 \). For this reason, it is tricky to find a useful time domain interpretation of zeros. In what follows, we provide a useful and nontrivial time domain interpretation.

Let \( z_0 \) be a zero. Let the normal rank of \( P(z) \) be equal to \( p \). Let \( c \) be any arbitrary \( r \times 1 \) vector. Then show that there exist finite length sequences \( s(n) \) and \( t(n) \) with \( T(z) \neq 0 \) for any \( z \), such that the input
\[
U(z) = \frac{c}{1 - z_0 z^{-1}} + S(z)
\]  
(P13.21)

produces an output \( Y(z) \) for which \( T(z)Y(z) \) is a finite length sequence. [Here \( S(z) \) and \( T(z) \) represent the z-transforms of \( s(n) \) and \( t(n) \).]

Note. If the normal rank of \( P(z) \) were less than \( p \), this would be true whether \( z_0 \) is a zero or not. Also note that in the above statement \( c \) is an arbitrary vector; this is what makes the statement nontrivial!

Hints. Use the fact that \( (z - z_0) \) is a factor of \( \alpha_{p-1}(z) \) in the Smith-McMillan form. The trick is to find a way to cancel the IIR components of the output, introduced by the first term in (P13.21), and by the quantity \( \beta_{p-1}(z) \) in the Smith-McMillan form. It helps to remember that \( \alpha_{p-1}(z) \) and \( \beta_{p-1}(z) \) are relatively prime.

13.22. Consider the MFD \( H(z) = Q^{-1}(z)P(z) \), and let the polynomials \( Q(z) \) and \( P(z) \) be given by
\[
Q(z) = \sum_{n=0}^{L} q_n z^n, \quad P(z) = \sum_{n=0}^{L} p_n z^n.
\]  
(P13.22a)

The input-output relation \( Y(z) = H(z)U(z) \) [i.e., \( Q(z)Y(z) = P(z)U(z) \)], can now be expressed in the time domain as follows:
\[
\sum_{m=0}^{L} q_m y(n + m) = \sum_{m=0}^{L} p_m u(n + m).
\]  
(P13.22b)

This difference equation completely describes the input-output behavior of the system. The difference equation with \( u(n) = 0 \), that is,
\[
\sum_{m=0}^{L} q_m y(n + m) = 0,
\]  
(P13.22c)

is said to be the homogeneous equation.
a) Let \( z_p \) be a pole of \( H(z) \). Show that the homogeneous equation has a solution of the form \( y(n) = v z_p^n \), for some constant non-zero vector \( v \).

b) Conversely, let \( v z_p^n \) be a solution to the homogeneous equation (\( v \neq 0, z_p \neq 0 \)). Assuming that the MFD is irreducible, show that \( z_p \) is a pole of \( H(z) \).

13.23. One of the results claimed by Theorem 13.6.1 is this: If \( (A,B,C,D) \) is a minimal realization of a causal rational system \( H(z) \), then \( z_p \) is a pole of \( H(z) \) if and only if it is an eigenvalue of \( A \). The proof that an eigenvalue \( \lambda \) of \( A \) is a pole of \( H(z) \) assumed that \( \lambda \neq 0 \). We now develop a proof (suggested to the author by Prof. John Doyle, Caltech), which works even when \( \lambda = 0 \). This is based on the fact that given any \( N \times N \) matrix, we can always apply a similarity transform to obtain a special form called the Jordan form [Chen, 1970, 1984]. Since the similarity transform leaves the transfer function unchanged, we will develop the proof based on this form. The Jordan form is a block-diagonal matrix, given by

\[
A = \begin{bmatrix}
A_0 & 0 & \ldots & 0 \\
0 & A_1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & A_{L-1}
\end{bmatrix}
\]  

(P13.23a)

where \( A_i \) are \( n_i \times n_i \) matrices of the special lower triangular form

\[
A_i = \begin{bmatrix}
\lambda_i & 0 & \ldots & 0 \\
\times & \lambda_i & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \times \lambda_i
\end{bmatrix}
\]  

(P13.23b)

Here \( \lambda_i \) is an eigenvalue of \( A \) of multiplicity \( n_i \), and \( \times \) denotes possibly nonzero entries (with values \( = 0 \) or \( 1 \), but that is irrelevant here). Let us assume that the minimal realization \( (A,B,C,D) \) has \( A \) in the above form.

a) By partitioning \( B \) and \( C \) as

\[
B = \begin{bmatrix}
B_0 \\
B_1 \\
\vdots \\
B_{L-1}
\end{bmatrix}, \quad \text{and} \quad C = [C_0 \ C_1 \ \ldots \ C_{L-1}]
\]  

(P13.23c)

show that we can express \( H(z) \) as

\[
H(z) = \sum_{k=0}^{L-1} \frac{C_k(zI - A_k)^{-1}B_k}{H_k(z)} + D
\]  

(P13.23d)

b) Let \( \lambda_0 = 0 \), that is, \( A \) has the eigenvalue zero, with multiplicity \( n_0 > 0 \). By definition, \( \lambda_i \neq 0 \) for \( i > 0 \). Show that, all the poles of \( H_0(z) \) are at \( z = 0 \). Show also that none of \( H_k(z), k > 0 \) has a pole at \( z = 0 \).

c) Thus, unless \( H_0(z) \) is a constant, \( H(z) \) has a pole at \( z = 0 \). We now show that \( H_0(z) \) cannot be constant, by using the minimality of \( (A,B,C,D) \).
(i) First show that minimality of \((A, B, C, D)\) implies \(C_0 \neq 0\) as well as \(B_0 \neq 0\). (ii) Then show that if \(H_0(z)\) is constant, then

\[
\begin{bmatrix}
B_0 \\
0 \\
\vdots \\
0
\end{bmatrix} = 0,
\]

thus violating complete observability, and hence minimality of the realization \((A, B, C, D)\).

Summarizing, we have proved that \(H_0(z)\) is not a constant, and therefore has pole(s) at \(z = 0\). Thus the eigenvalues \(\lambda_0 = 0\) of the matrix \(A\) imply that \(H(z)\) has pole(s) at \(z = 0\).

13.24. Let \(x(n)\) and \(y(n)\) be the input and output of an LTI system \(H(z)\), and define the blocked versions \(x_B(n)\) and \(y_B(n)\) as in (10.1.1), Chap. 10. Show that the \(M\)-input \(M\)-output system which produces \(y_B(n)\) in response to \(x_B(n)\) is an LTI system. In other words, verify that the defining property of a multi-input multi-output LTI system (given in Problem 13.17) is satisfied. (In Chap. 10 we just assumed this, and proceeded to show that the transfer matrix is pseudocirculant.)

13.25. Let \((A, B, C, D)\) be a minimal state space representation of a causal scalar transfer function \(H(z)\) with degree \(N\). Let \(H(z)\) be the blocked version of \(H(z)\) with block length \(M\) (Sec. 10.1.1). Find a state space description \(\hat{A}, \hat{B}, \hat{C}, \hat{D}\) for \(H(z)\) in terms of \((A, B, C, D)\) and \(M\). Make sure your answer is such that \(\hat{A}\) has size no larger than \(N \times N\). Can you show that the blocked version has McMillan degree \(N\)?