Approximating Properties of Linear Models for Input-Output Descriptions

by

Ernest G. Holzmann

March 1968

SCHOOL OF ENGINEERING
STANFORD UNIVERSITY · STANFORD, CALIFORNIA

APPROXIMATING PROPERTIES OF LINEAR MODELS

FOR INPUT-OUTPUT DESCRIPTIONS

by

Ernest Gunther Holzmann

March 1968

DEPARTMENT OF ELECTRICAL ENGINEERING
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Professor R. E. Kalman, who supervised this study, for his guidance and patience throughout the course of the investigation, and to Professor G. F. Franklin for his encouragement and helpful suggestions.

Major financial support was received from the National Aeronautics and Space Administration under Contract NGL- ~~NgR-~~05-020-073 and is gratefully acknowledged.

ABSTRACT

This report presents results of a research effort to solve certain approximation problems that arise in the computation of linear stationary models of dynamical systems from given input-output data.

B.L. Ho's existence theorem states necessary and sufficient conditions for strict realizability that are satisfied only in ideal situations. Mathematical proof is given in this report that good dynamical simulation is possible with linear models which represent partial, rather than minimal, realizations. The restrictions of B.L. Ho's theorem do not apply to partial realizations, and the class of partially realizable input-output descriptions is large enough for practical purposes. For any normal sequence of scalar Markov parameters, the transfer function of each partial realization is shown to lie on the diagonal of the E-array corresponding to the given sequence. The proof is based on the classical theory of the Padé approximation. Relevant parts of this theory are reviewed and developed in the report, including a new, stronger form of Padé's representation theorem.

As a by-product of this research, a sharpened, computationally more efficient version of B.L. Ho's minimal realization algorithm was derived. The new algorithm expresses every minimal realization of a given sequence of Markov parameters in terms of the pseudo-inverse matrices $(V^{\dagger}, W^{\dagger})$. The generating matrices (V, W) are familiar from the theory of complete controllability and observability. The algorithm is shown to be the sharpest possible, subject to the requirement that every minimal realization be obtainable.

TABLE OF CONTENTS

LIST OF SYMBOLS

Symbols	Description	Place of First Use
A^k	column vector	Lemma 2.1
$A_N(z)$	Nth partial sum of $f(z)$	Eq. (3.3)
B	matrix of linearly independent columns	Lemma 4.7
C	matrix of linearly independent rows	Lemma 4.7
E_n	$E_n = [I_n 0 \ldots 0]$	Eq. (4.7)
E_{uv}	$u \times v$ canonical diagonal matrix of rank u	Eq. (4.6)
$E_i^{(j)}(Z, s)$	(i, j) element of E-array for $Z(s)$	Eq. (2.60)
$f(z)$	power series	Theorem 2.2
F	$n \times n$ system matrix	Eq. (4.1)
G	$n \times m$ input distribution matrix	Eq. (4.1)
H	$p \times n$ output distribution matrix	Eq. (4.1)
I_n	$n \times n$ unit matrix	
J	canonical diagonal matrix	Eq. (B.7)
$\mathcal{K}$	column space of $S_r^{(0)}$	Prop. 4.11
M	companion matrix	Eq. (4.4)
N	companion matrix	Eq. (4.5)
$N(z)$	numerator polynomial	Eq. (2.19)
O_n	$n \times n$ null matrix	
P	premultiplier in realization algorithm	Theorem 4.4
$P(z)$	numerator polynomial	Theorem 2.5
P_{ij}	numerator polynomial of (i, j) Pade' pair	Theorem 2.4
$P_i^{(j)}$	numerator polynomial of $E_i^{(j)}(Z, s)$	Eq. (2.60)

Symbols	Description	Place of First Use
Q	postmultiplier in realization algorithm	Theorem 4.4
$Q(z)$	denominator polynomial	Theorem 2.5
Q_{ij}	denominator polynomial of (i, j) Pade' pair	Theorem 2.4
$Q_i^{(j)}$	denominator polynomial of $E_i^{(j)}(Z, s)$	Eq. (2.60)
$\mathcal{R}$	row space of $S_r^{(0)}$	Prop. 4.11
$S, S_r^{(k)}$	generalized Hankel matrix	Eq. (3.1)
T	$n \times n$ transformation matrix	Lemma 4.8
$T(z)$	z-transform transfer function	Eq. (4.21)
$T_{ij}(z)$	determinantal form of numerator polynomial	Eq. (3.4)
$u(k)$, $u(t)$	$m \times 1$ input vector	Eq. (4.19)
U	inverse of T	Lemma 4.8
$U_{ij}(z)$	determinantal form of denominator polynomial	Eq. (3.4)
V	$[H' \quad F'H' \quad \ldots \quad (F')^{r-1}H']'$	Eq. (4.10)
W	$[G \quad FG \quad \ldots \quad F^{r-1}G]$	Eq. (4.10)
$W(t)$	impulse response function	Eq. (4.23)
$x(k)$, $x(t)$	$n \times 1$ state vector	Eq. (4.19)
$\mathfrak{X}$	sequence of real numbers	Eq.(6.2)
$y(k)$, $y(t)$	$p \times 1$ output vector	Eq. (4.19)
y_k	scalar (Markov) parameter	Prop. 5.1
Y_k	matrix (Markov) parameter	Eq. (4.1)
$Z(s)$	Laplace transform transfer function	Eq. (6.26)
$Z(s)$	infinite series	Eq. (2.57)
Z_N	Nth partial sum of series $Z(s)$	Eq. (3.53)

Symbols	Description	Place of First Use
α_i	linear recursion coefficient	Eq. (4.2)
$\Delta_n^{(k)}$	det $S_n^{(k)}$	Eq. (3.2)
ΔZ_N	Nth term of series $Z(s)$	Eq. (3.54)
λ	nonnegative index	Theorem 2.4
ρ	dimension of minimal realization	Prop. 6.1
$\psi(z)$	annihilating polynomial of F	Prop. 4.1

Superscripts

$'$	transpose	
$\#$	(non-unique) pseudo inverse	Eq. (4.14)
$\dagger$	pseudo inverse	Eq. (4.26)
$*$	conjugate transpose	Eq. (B.5)
$\perp$	orthogonal complement (of a vector subspace)	Prop. 4.11
$-$	submatrix	Prop. 4.11

I. INTRODUCTION.

The problems considered in this report were inspired by certain questions raised by Dr. Bin-Lun Ho in his dissertation "On Effective Construction of Realizations from Input-Output Descriptions" [6].

The object is to transform given input-output data of a multivariate process into another description more suitable for simulation. In the case of stationary linear dynamical systems, which are the class studied by B.L. Ho, a useful description may take the form of state-variable differential (or difference) equations. From a practical point of view, the derived description (called the "model") should meet the following criteria:

a. The model should reproduce the observed external behavior patterns of the dynamical system with acceptable accuracy.

b. The construction of the model from the given input-output data should be economically carried out on a computer, using available or readily programmed routines.

c. The model itself should be amenable to economical simulation on a computer.

These three criteria determine the quality, price, and operating cost of the model.

B. L. Ho's methods meet the above requirements at least as well as the known methods of other researchers. In fact, B. L. Ho's models are called "realizations" precisely because they perfectly match given input-output data. Furthermore, in the sense used by B. L. Ho, a realization is a finite set of first-order linear differential equations (expressed in terms of the coefficient matrices), and programming of a realization for simulation therefore presents no special difficulty.

Suppose a given input-output sequence does not meet B. L. Ho's realizability conditions for a finite-dimensional model. Then two questions arise quite naturally:

(1) Does B. L. Ho's method give a model whose external behavior has approximating properties that make the model useful in simulating studies?

(2) Can B. L. Ho's method be modified to further improve the approximating properties found in (1)?

The present report covers the first phase of a continuing study aimed at answering these two questions.

Chapter II of the report reviews basic theorems in classical Padé approximation theory, i.e., the rational approximation of functions represented by power series, in a neighborhood of the origin of the argument. To prepare for application of the theory to the problem posed by question (1) above, a new and stronger form of Padé's representation theorem is presented (Theorem 2.5). In this chapter, we also draw attention to some pitfalls which must be avoided when generalizing results from normal to non-normal Padé tables. Examples to illustrate this point are discussed. Anticipating later applications to the theory of linear dynamical systems, the chapter concludes with the definition of the E-array and its relation to the Padé table.

Chapter III concentrates on determinantal expressions which play a prominent part in the Padé theory. Such expressions have long been known for normal Padé tables, but Theorem 3.7, giving the determinantal representation of the Padé approximant for the general case, does not appear to have been stated or proved in the available literature. A corollary of the theorem, restricted to the normal case, indicates a similar representation for the elements of the E-array.

Chapter IV states two results of B. L. Ho's work which are pertinent to the present research, namely the existence theorem (Proposition 4.1) and the algorithm for minimal realizations (Theorem 4.4). By appealing to the unique properties of the pseudo inverse, we are able to sharpen B. L. Ho's algorithm in Theorem 4.9. Corollary 4.10 presents the unique reciprocal relations between any minimal realization (F, G, H) and the matrices (V, W).

Chapter V, like Chapter III, deals primarily with determinantal relationships, but restricted to realizable sequences, i.e., to sequences corresponding to rational functions. As a by-product, Corollary 5.5 generalizes one of B. L. Ho's theorems. Theorem 5.8 gives four mathematical equivalents of the statement that a scalar sequence has a minimal realization.

In Chapter VI, we generalize the concept of the realization of a sequence by considering partial realizations and the associated approximation problems. Theorem 6.5 proves that the partial realization of a normal sequence $\mathcal{Y}$ (i.e., one for which $\Delta_r^{(0)} \neq 0$, all positive integers r) is closely related to the Padé approximants for the power series $\Sigma\, y_k z^k$. Corollary 6.6 identifies the transfer function, of the partial realization for normal $\mathcal{Y}$, with elements in the E-array for the power series.

Chapter VII serves as a review of the results obtained. The limitations of the work point to the need for further research, as indicated at the end of that chapter.

Two appendices are included. The first one summarizes a few definitions from algebra that are pertinent to Chapter II. The second appendix briefly states definitions and properties of the pseudo inverse of a matrix. These are used in Chapter IV.

II. THE PADÉ TABLE.

As pointed out in the Introduction, this chapter is concerned primarily with properties of the classical Padé approximation.

After the definition of standard terms, we prove a lemma which will later allow us to sharpen certain classical results.

Following the lemma, we consider the classical questions of existence (Theorem 2.2) and uniqueness (Theorem 2.3) of the entries in the Padé table for a given power series. The existence of Theorem 2.2 is proven by a constructive approach designed to pave the way for Theorem 2.5.

Theorem 2.5 has not been found in the published literature on the Padé approximation. It is a stronger form of the classical representation theorem (Theorem 2.4). Its formulation will be used in subsequent sections to link together the classical theory of the Padé approximation and the more recent theory of controlled linear dynamical systems.

Theorem 2.6 gives precise meaning to the notion that the Padé approximant is, in some sense, a "best" rational approximation to a given power series. The proof of the theorem is followed by a short discussion of its significance.

The possible existence of square blocks of equal approximants gives rise to the distinction between normal and other Padé tables. This important subject is introduced in Theorem 2.7 and its two corollaries.

The last section of the chapter deals with the E-array associated with a normal Padé table.

2.1 Definitions. [17, p. 378]

1. Let

$$f(z) = \sum_{k=0}^{\infty} a_k z^k \tag{2.1}$$

be a formal power series in one variable, with real coefficients.*
Let (i, j) be an ordered pair of nonnegative integers.

The (i, j) Padé approximant for f is a rational function

$$R_{ij}(f, z) = \frac{N_{ij}(z)}{D_{ij}(z)} \tag{2.2}$$

with the two properties (called the defining conditions of the approximant):

$$\text{I.}\quad \deg D_{ij} \leq i, \quad D_{ij} \neq 0,$$

$$\deg N_{ij} \leq j; \tag{2.3}$$

and $$\text{II.}\quad f(z)D_{ij}(z) - N_{ij}(z) = (z^{i+j+1}) \tag{2.4}$$

where (z^k) denotes a power series beginning with the term z^k or a higher power of z.**

2. The Padé table for f is the (doubly) infinite matrix $\mathcal{R}(f) = [R_{ij}]$, $i = 0, 1, \ldots;$ $j = 0, 1, \ldots;$ of Padé approximants for f.

3. The (i, j) Padé approximant for f is called normal if the quotient R_{ij} is distinct from all other quotients in the table.

A formal power series f is normal if all of the Padé approximants for f are normal, i.e. distinct. The Padé table for f is then also called normal, and so is the sequence of coefficients $(a_0, a_1, \ldots)$.

---

* For definition and brief discussion of the properties of formal power series, as well as other terminology and definitions from algebra, see Appendix A.

** In particular, we may have $(z^k) \equiv 0$.

2.2 Existence and Uniqueness of Padé Table.

Lemma 2.1

Hypotheses:

1. $\{A^k: k = 0, 1, \ldots, i\}$ is a set of vectors in the real Euclidean vector space R^i, $i \geq 1$.

2. The vectors A^k $(k = 0, 1, \ldots, i)$ span an r-dimensional subspace of R^i.

3. m is the largest index such that the vectors $A^m, A^{m+1}, \ldots, A^i$ are linearly dependent.

Conclusions:

1. $0 \leq i - r \leq m \leq i$ (2.5)

2. The linear homogeneous equation

$$\sum_{k=0}^{i} d_k A^k = 0 \tag{2.6}$$

has a unique solution $(d_0, d_1, \ldots, d_i)$ such that

$$d_k = \begin{cases} 0 & (k < m) \\ 1 & (k = m) \end{cases} \tag{2.7}$$

<u>Proof</u>:

1. By hypothesis 2, $r(\leq i)$ of the given vectors A^k are linearly independent, but any collection of $r + 1$ vectors from the given set are linearly dependent. In particular, the vectors A^k $(k = i - r, i - r + 1, \ldots, i)$ are linearly dependent. Hypothesis 3 then implies

$$i - r \leq m \leq i.$$

2. <u>Suppose</u> $A^m \neq 0$.

The defining property of m now implies

(i) $m < i$.

(ii) The vectors A^k $(k = m + 1, m + 2, \ldots, i)$ are all nonzero and linearly independent.

(iii) A^m lies in the $(i - m)$-dimensional subspace spanned by the vectors A^k $(k = m + 1, \ldots, i)$.

Therefore, the equation

$$A^m + \sum_{k=m+1}^{i} d_k A^k = 0 \tag{2.8}$$

has a unique solution $(d_{m+1}, d_{m+2}, \ldots, d_i)$.

Existence of the solution follows from (iii) above.

To prove uniqueness, suppose $(d'_{m+1}, \ldots, d'_i)$ were a second solution of (2.8). Then

$$\sum_{k=m+1}^{i} (d_k - d'_k) A^k = 0 \tag{2.9}$$

and (ii) above implies

$$d_k = d'_k \qquad (k = m + 1, \ldots, i).$$

The solution of (2.8), together with the values of $d_k (k \leq m)$ given by (2.7), satisfies the vector equation (2.6) and is unique.

<u>Suppose</u> $A^m = 0, \quad m = i.$

Then

$$d_k = \begin{cases} 0 & (k < i) \\ 1 & (k = i) \end{cases}$$

is a solution of (2.6) satisfying (2.7), and clearly it is the only solution satisfying (2.7).

<u>Suppose</u> $A^m = 0, \ m < i.$

Then, from the definition of m, we must have

$$A^k \neq 0 \quad (k = m + 1, m + 2, \ldots, i)$$

and these vectors are linearly independent.

Now substitute the values of d_k given by (2.7) into equation (2.6). Then

$$0 = A^m + \sum_{k=m+1}^{i} d_k A^k = \sum_{k=m+1}^{i} d_k A^k, \quad \text{since } A^m = 0.$$

By linear independence, we get the unique result

$$\sum_{k=m+1}^{i} d_k A^k = 0 \implies d_k = 0 \qquad (k = m + 1, m + 2, \ldots, i).$$

Therefore

$$d_k = \begin{cases} 0 & (k \neq m) \\ 1 & (k = m) \end{cases} \tag{2.10}$$

uniquely satisfies both (2.6) and (2.7).

Theorem 2.2 (Existence Theorem) (Padé) [11, p. 9]

Hypothesis: $f(z) = \sum_{k=0}^{\infty} a_k z^k, \; a_0 \neq 0.$

Conclusion: For each ordered pair (i, j) of nonnegative integers, there exists a rational function $R_{ij}(f, z)$ satisfying the conditions I and II of the (i, j) Padé approximant.

Proof: Let $D(z) = \sum_{k=0}^{i} d_k z^k$ (2.11)

be a polynomial with undetermined coefficients $(d_0, \ldots, d_i)$. Form the product

$$f(z)D(z) = \sum_{k=0}^{\infty} c_k z^k \tag{2.12}$$

where

$$c_k = \sum_{u+v=k} a_u d_v. \tag{2.13}$$

The undetermined coefficients d_k are chosen as follows:

(i) If $i = 0$, take $d_0 = 1$. (2.14)

(ii) If $i > 0$, we set

$$c_k = 0 \qquad (k = j+1,\ j+2,\ \ldots,\ j+i) \tag{2.15}$$

Written out, (2.15) is a system of i linear homogeneous equations in the $i + 1$ unknowns $d_0, d_1, \ldots, d_i$, and thus always has a non-trivial solution. In matrix form, the system (2.15) is

$$\begin{bmatrix} a_{j-i+1} & a_{j-i+2} & \cdots & a_{j+1} \\ a_{j-i+2} & a_{j-i+3} & \cdots & a_{j+2} \\ \cdot & \cdot & \cdots & \cdot \\ a_j & a_{j+1} & \cdots & a_{j+i} \end{bmatrix} \begin{bmatrix} d_i \\ d_{i-1} \\ \vdots \\ d_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.16}$$

or, equivalently,

$$\sum_{k=0}^{i} d_k A^k = 0 \tag{2.17}$$

where

$$A^k = \begin{bmatrix} a_{j-k+1} \\ a_{j-k+2} \\ \vdots \\ a_{j-k+i} \end{bmatrix} \in R^i.$$

Let r be the rank of the $[a]$ matrix in equation (2.16). Then the vectors A^k $(k = 0, 1, \ldots, i)$ and the integer r satisfy the hypotheses of Lemma 2.1. Let $m \geq 0$ be the index defined in the lemma.

By the conclusions of the lemma, (2.17) has a (unique) nontrivial solution $(d_0, d_1, \ldots, d_i)$ such that

$$d_k = \begin{cases} 0 & (k < m) \\ 1 & (k = m) \end{cases}$$

Substitution in (2.11) gives

$$D(z) = z^m + \sum_{k=m+1}^{i} d_k z^k \tag{2.18}$$

The product fD is therefore a power series of the form

$$fD = N + (z^{i+j+1})$$

where
$$N(z) = a_0 z^m + \sum_{k=m+1}^{j} c_k z^k . \tag{2.19}$$

From (2.18) and (2.19):

$$\deg D \leq i,\ D \neq 0, \quad \text{and} \quad \deg N \leq j.$$

Thus D and N satisfy the defining conditions I and II of the (i, j) Padé approximant for f.

Theorem 2.3 (Uniqueness Theorem) [17, p. 378]

Hypothesis: 1. $f(z) = \sum_{0}^{\infty} a_k z^k$.

2. (i, j) is an ordered pair of nonnegative integers.

3. Each of the two pairs of polynominals (N, D) and (N', D') satisfies the conditions I and II of the (i, j) Padé approximant for f.

Conclusion:

$$\frac{N}{D} = \frac{N'}{D'} .$$

Proof: By hypothesis,

(i) $fD - N = (z^{i+j+1}) \Longrightarrow fDD' = [N + (z^{i+j+1})]D'$; and

(ii) $fD' - N' = (z^{i+j+1}) \Longrightarrow fD'D = [N' + (z^{i+j+1})]D$.

Therefore

$$ND' - N'D = (z^{i+j+1}). \tag{2.20}$$

But $\deg[ND' - N'D] \leq i + j$,

$\Longrightarrow$ the left side of (2.20) contains no power of z with exponent higher than $i + j$,

$\Longrightarrow$ the right side of (2.20) is identically zero,

$\Longrightarrow$ $ND' - N'D = 0$

$\Longrightarrow$ $\frac{N}{D} = \frac{N'}{D'}$.

2.3 Representation Theorems.

Theorem 2.4 (Padé's Representation Theorem)[13, p. 421]

Hypothesis:

1. $f(z) = \sum_{0}^{\infty} a_k z^k$, $a_0 \neq 0$.

2. (i, j) is an ordered pair of nonnegative integers, and $R_{ij}(f, z)$ is the (i, j) Padé approximant for f.

Conclusions: There exists a unique pair of polynomials (P_{ij}, Q_{ij}) and a nonnegative integer λ such that

(i) $P_{ij}(0) = a_0$, $Q_{ij}(0) = 1$;

(ii) $\deg P_{ij} \leq j - \lambda$, $\deg Q_{ij} \leq i - \lambda$;

(iii) $z^{\lambda}[f\ Q_{ij} - P_{ij}] = (z^{i+j+1})$; and

(iv) P_{ij} and Q_{ij} are relatively prime.

Furthermore, the polynomials (P_{ij}, Q_{ij}) defined by (i) - (iv) also have the property

(v) $R_{ij}(f, z) = \dfrac{P_{ij}(z)}{Q_{ij}(z)}$.

Proof: By definition of the (i, j) Padé approximant for f, R_{ij} has a representation

$$R_{ij}(f, z) = \frac{N_{ij}(z)}{D_{ij}(z)}$$

where N_{ij} and D_{ij} are polynomials satisfying conditions I and II.

The greatest common divisor of N_{ij} and D_{ij} is of the form $z^{\lambda}B(z)$, where

$$\lambda \geq 0$$

$$B(z) = \sum_{v=0}^{r} b_v z^v, \quad b_0 \neq 0, \quad r \geq 0.$$

Now there exist relatively prime polynomials P_{ij}, Q_{ij} such that

$$N_{ij}(z) = z^{\lambda}B(z)P_{ij}(z), \tag{2.21}$$

$$D_{ij}(z) = z^{\lambda}B(z)Q_{ij}(z), \tag{2.22}$$

$$P_{ij}(0) = a_0, \quad Q_{ij}(0) = 1. \tag{2.23}$$

By property II of the Padé approximant, we have

$$[f(z)Q_{ij}(z) - P_{ij}(z)]z^{\lambda}B(z) = (z^{i+j+1})$$

$$\Longrightarrow [f(z)Q_{ij}(z) - P_{ij}(z)]z^{\lambda} = (z^{i+j+1}), \quad \text{since } b_0 \neq 0.$$

Next, $\deg N_{ij} \le j$

$\Longrightarrow \deg [z^{\lambda} P_{ij}] \le j - r \le j$, by (2.21);

$\Longrightarrow \deg P_{ij} \le j - \lambda$.

Similarly, $\deg D_{ij} \le i$

$\Longrightarrow \deg [z^{\lambda} Q_{ij}] \le i - r \le i$, by (2.22);

$\Longrightarrow \deg Q_{ij} \le i - \lambda$.

Thus, we have shown the existence of polynomials (P_{ij}, Q_{ij}) satisfying (i) - (iv). Conclusion (v) is immediate from (2.21) and (2.22).

The uniqueness of (P_{ij}, Q_{ij}) satisfying (i) - (iv) is shown as follows.

By Theorem 2.3, R_{ij} is a unique rational function, with at most i poles and j zeros. Since P_{ij} and Q_{ij} are relatively prime and thus have no zeros in common, it follows from

$$R_{ij} = \frac{P_{ij}}{Q_{ij}}$$

that the zeros of P_{ij} are exactly the same as the zeros of R_{ij}, with their multiplicities. By the Factor Theorem for Polynomials [16, p. 61] [9, p. 121], the polynomial P_{ij} is uniquely characterized (aside from a constant factor) by its zeros. Therefore, the zeros of R_{ij}, together with the condition $P_{ij}(0) = a_0$, uniquely specify P_{ij}. Similarly, Q_{ij} is uniquely given by the poles of R_{ij} and $Q_{ij}(0) = 1$.

<u>Definition</u>. The unique pair of polynomials (P_{ij}, Q_{ij}) postulated in Theorem 2.4 is called the (i, j) <u>Padé pair</u> (for f).

Theorem 2.5

Hypotheses:

1. $f(z) = \sum_{k=0}^{\infty} a_k z^k, \; a_0 \neq 0.$

2. (i, j) is an ordered pair of integers, $i \geq 1$, $j \geq 0$.

3. m is the largest index such that the column vectors $A^m, A^{m+1}, \ldots, A^i$ are linearly dependent, where

$$A^k = \begin{bmatrix} a_{j-k+1} \\ a_{j-k+2} \\ \cdot \\ \cdot \\ \cdot \\ a_{j-k+i} \end{bmatrix} \qquad (k = 0, 1, \ldots, i) \qquad (2.24)$$

$$a_v = 0 \text{ for } v < 0.$$

Conclusions:

1. There exists a unique pair of polynomials (P, Q) such that

(i) $P(0) = a_0$, $Q(0) = 1$;

(ii) $\deg P \leq j - m$, $\deg Q \leq i - m$;

(iii) $z^m[fQ - P] = (z^{i+j+1})$; and

2. The pair of polynomials (P, Q) defined by (i) - (iii) is the (i, j) Padé pair for f. That is, P and Q have the additional properties

(iv) $R_{ij}(f, z) = \frac{P(z)}{Q(z)}$.

(v) P and Q are relatively prime.

3. The index m has the additional properties:

(vi) $m = \max \lambda$, the maximum being taken over all integers λ satisfying

$$\deg P \leq j - \lambda, \quad \deg Q \leq i - \lambda \tag{2.25}$$

where (P, Q) is the (i, j) Padé pair for f.

(vii) Either $\deg P = j - m$, or $\deg Q = i - m$.

Proof: The existence of P and Q is readily shown. Using the same construction as in the proof of Theorem 2.2, we get two polynomials N and D,

$$D(z) = z^m + \sum_{k=m+1}^{i} d_k z^k \tag{2.18}$$

$$N(z) = a_0 z^m + \sum_{k=m+1}^{j} n_k z^k \tag{2.19}$$

satisfying conditions I and II of the (i, j) Padé approximant for f. Therefore

$$R_{ij}(f, z) = \frac{N(z)}{D(z)} \tag{2.26}$$

Let

$$\begin{aligned} P(z) &= z^{-m} N(z) = a_0 + \sum_{k=m+1}^{j} n_k z^{k-m} \\ Q(z) &= z^{-m} D(z) = 1 + \sum_{k=m+1}^{i} d_k z^{k-m}. \end{aligned} \tag{2.27}$$

Then (2.26) implies

$$R_{ij}(f, z) = \frac{P(z)}{Q(z)} \tag{2.28}$$

and $P(0) = a_0$, $Q(0) = 1$; $\deg P \le j - m$, $\deg Q \le i - m$.

Also, $fD - N = (z^{i+j+1})$ and (2.27) imply $z^m[fQ - P] = (z^{i+j+1})$. (2.29)

Therefore the polynomials (P, Q) defined by (2.27) have the properties (i) through (iv).

To prove the uniqueness of P and Q, we show first that P and Q are relatively prime and then apply Theorem 2.4.

Certainly z does not divide P or Q, because of their form (2.27). Suppose, now, that the polynomial

$$B(z) = 1 + b_1 z + \ldots + b_n z^n, \quad 0 \le n \le \min(i, j)-m \tag{2.30}$$

divides P and Q. Then there are polynomials P^* and Q^* such that

$$P(z) = B(z)P^*(z)$$

$$Q(z) = B(z)Q^*(z).$$

From equations (2.27) and (2.30), P^* and Q^* have the form

$$\begin{aligned} P^*(z) &= a_0 + \sum_{k=1}^{j-m-n} p_k^* z^k \\ Q^*(z) &= 1 + \sum_{k=1}^{i-m-n} q_k^* z^k. \end{aligned} \tag{2.31}$$

Consider now two polynomials N^* and D^*, defined by

$$\begin{aligned} N^*(z) &= z^{m+n} P^*(z) \\ D^*(z) &= z^{m+n} Q^*(z). \end{aligned} \tag{2.32}$$

N^*, D^* have the following properties:

$$\deg N^* \le j, \quad \deg D^* \le i, \quad D^* \ne 0;$$

and (2.29) implies

$$\begin{aligned} &z^m B[fQ^* - P^*] = (z^{i+j+1}) \\ \Longrightarrow\ &fD^* - N^* = (z^{i+j+1+n}). \end{aligned} \tag{2.33}$$

Furthermore, (2.31) and (2.32) imply that D^* may be written

$$D^*(z) = \sum_{k=0}^{i} d_k^* z^k \tag{2.34}$$

with $$d_k^* = \begin{cases} 0 & (k < m + n) \\ 1 & (k = m + n) \\ q_{k-m-n}^* & (k = m + n + 1,\ m + n + 2,\ \ldots,\ i). \end{cases}$$

The product fD^* is a power series:

$$fD^* = \sum_{k=0}^{\infty} c_k^* z^k$$

where

$$c_k^* = \sum_{u+v=k} a_u d_v^*.$$

From (2.33) and $\deg N^* \leq j$ we infer that the following coefficients c_k^* vanish:

$$c_k^* = 0 \quad (k = j + 1,\ j + 2,\ \ldots,\ j + i). \tag{2.35}$$

The i homogeneous linear equations (2.35) can be stated in the form of a linear relation between the $i + 1$ column vectors A^k, with coefficients d_k^*:

$$\sum_{k=0}^{i} d_k^* A^k = 0. \tag{2.36}$$

Substituting for d_k^* $(k = 0, 1, \ldots, m + n)$ in (2.36), we have

$$A^{m+n} + \sum_{k=m+n+1}^{i} d_k^* A^k = 0. \tag{2.37}$$

By hypothesis 3, the linear dependence relation (2.37) implies $n = 0$. Therefore $B(z) \equiv 1$, and P and Q are relatively prime, as claimed in (v).

Now f, (P, Q) and $\lambda = m$ satisfy the hypotheses of Theorem 2.4. Since P and Q are relatively prime, Theorem 2.4 ensures the uniqueness of the representation. This completes the proof of conclusions 1 and 2.

To show $m = \max \lambda$, we note that because of (ii), m satisfies the conditions (2.25) for λ. It remains to show that no larger value of λ can satisfy (2.25).

Suppose $\lambda_0 \geq m + 1$, and λ_0 satisfies (2.25). Then $\deg P \leq j - m - 1$, $\deg Q \leq i - m - 1$. Therefore, the two polynomials P, Q have the form

$$P(z) = \sum_{k=0}^{j-m-1} n_k z^k, \qquad n_0 = a_0;$$

$$Q(z) = \sum_{k=0}^{i-m-1} d_k z^k, \qquad d_0 = 1.$$

Define $c_k = \sum_{u+v=k} a_u d_v$, the coefficient of z^k in the power series fQ. From $z^m[fQ - P] = (z^{i+j+1})$ we obtain the two sets of equations

$$n_k = c_k \qquad (k = 0, 1, \ldots, j - m - 1),$$

$$c_k = 0 \qquad (k = j - m, j - m + 1, \ldots, j - m + i).$$

In matrix form, the second set reads

$$\begin{bmatrix} a_{j-i+1} & a_{j-i+2} & \cdots & a_{j-m} \\ a_{j-i+2} & a_{j-i+3} & \cdots & a_{j-m+1} \\ \cdot & \cdot & \cdots & \cdot \\ a_j & a_{j+1} & \cdots & a_{j-m+i-1} \\ a_{j+1} & a_{j+2} & \cdots & a_{j-m+i} \end{bmatrix} \begin{bmatrix} d_{i-m-1} \\ d_{i-m-2} \\ \vdots \\ \\ d_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \\ 0 \end{bmatrix} \qquad (2.38)$$

By the definition of m, the columns of the $[a]$ matrix in (2.38) are linearly independent. Hence (2.38) can only have the trivial solution, contrary to the requirement $d_0 = 1$. Thus we have proved that no value of λ greater than m can satisfy (2.25). Hence (vi) follows, and (vii) is a trivial consequence of (vi).

The proof of Theorem 2.5 is complete.

<u>Relation between Theorems 2.4 and 2.5</u>. Theorem 2.5 is a new and stronger version of the classical representation theorem 2.4. Theorem 2.5 preserves the uniqueness property of the classical Padé pair and has the added advantages that

1. it gives sharp upper bounds for the degrees of the polynomials characterizing the (i, j) Padé pair for a given power series f;

2. it eliminates the classical requirement that the candidate polynomials P, Q for the Padé pair be relatively prime. The property of being relatively prime turns out to be a result of, rather than a condition for, the choice of the pair (P, Q);

3. the index m appearing in Theorem 2.5 is uniquely determined for each triple (f, i, j), while the similar parameter λ appearing in Theorem 2.4 is not unique.

The non-uniqueness of λ is demonstrated in the following example.

Example to Show Non-Uniqueness of λ.

Let $f(z) = \dfrac{1+z-z^3}{1-z^3} = 1 + z + z^4 + z^7 + z^{10} + z^{13} + z^{16} + \dots$

Padé table for f:

$i \backslash j$	0	1	2	3	4
0	1	$1+z$	$1+z$	$1+z$	$1+z+z^4$
1	$\dfrac{1}{1+z}$	$1+z$	$1+z$	$1+z$	$1+z+z^4$
2	$\dfrac{1}{1-z+z^2}$	$1+z$	$1+z$	$1+z$	$1+z+z^4$
3	$\dfrac{1}{1-z+z^2-z^3}$	$\dfrac{1}{1-z+z^2-z^3}$	$\dfrac{1+z+z^2}{1+z^2-z^3}$	$\dfrac{1+z-z^3}{1-z^3}$	
4	$\dfrac{1}{1-z+z^2-z^3}$	$\dfrac{1}{1-z+z^2-z^3}$	$\dfrac{1+2z+z^2}{1+z-z^4}$		

By Theorem 2.4, the (i, j) Padé pair for f is (P_{ij}, Q_{ij}), where

(i) $P_{ij}(0) = a_0, \qquad Q_{ij}(0) = 1;$

(ii) $\deg P_{ij} \leq j - \lambda, \qquad \deg Q_{ij} \leq i - \lambda;$

(iii) $z^{\lambda}[f\, Q_{ij} - P_{ij}] = z^{\lambda}[z^4 + z^7 + z^{10} + \dots] = (z^{i+j+1});$

(iv) P_{ij}, Q_{ij} relatively prime.

The integer λ (see conclusion of Theorem 2.4) has the following admissible values for the given power series:

$i \backslash j$	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	0	0, 1	1	0	0, 1	1
2	0	0	1	2	0	1	2
3	0	0	0	0	0	0	0
4	0	1	0	0	0, 1	0, 1	0, 1
5	0	0	0	0	0, 1	0, 1, 2	0, 1, 2

For each pair (i, j), the index m (see Theorem 2.5) equals the maximum admissible value of λ.

2.4 Padé's Fundamental Proposition.

Theorem 2.6 (Padé) [11, p. 12]

Hypotheses:

1. $f(z) = \sum_{0}^{\infty} a_k z^k, \; a_0 \neq 0.$

2. (i, j) is an ordered pair of non-negative integers; and $R_{ij}(f, z)$ is the (i, j) Padé approximant for f.

3. (P, Q) are a pair of polynomials in z, with

$$\deg P \leq j, \quad \deg Q \leq i.$$

4. r is the largest integer such that

$$f(z) - \frac{P(z)}{Q(z)} = (z^r);$$

s is the largest integer such that

$$f(z) - R_{ij}(f, z) = (z^s).$$

Conclusions:

1. $r \leq s$; and

2. $r = s \Longleftrightarrow R_{ij}(f, z) = \frac{P(z)}{Q(z)}.$

Proof: Half of conclusion 2, namely

$$R_{ij} = \frac{P}{Q} \Longrightarrow r = s, \tag{2.39}$$

follows trivially from the definition of r and s in hypothesis 4.

For the other half of conclusion 2, it suffices to show

$$r \geq s \Longrightarrow R_{ij} = \frac{P}{Q}. \tag{2.40}$$

Then (2.39) and (2.40) together imply that $r \geq s \Longrightarrow r = s$, i.e. $r \not> s$, and thus conclusion 1 is validated.

By Theorem 2.5, $R_{ij}(f, z)$ has a unique representation

$$R_{ij}(f, z) = \frac{P_{ij}(z)}{Q_{ij}(z)} \tag{2.41}$$

where the polynomials P_{ij}, Q_{ij} have the properties

(i) $P_{ij}(0) = a_0$, $Q_{ij}(0) = 1$;

(ii) $\deg P_{ij} \leq j - m$, $\deg Q_{ij} \leq i - m$;

(iii) $z^m[fQ_{ij} - P_{ij}] = (z^{i+j+1})$;

with $m \geq 0$ defined in terms of the coefficients of f.

Properties (i) and (iii), together with (2.41), give

$$f - R_{ij} = f - \frac{P_{ij}}{Q_{ij}} = (z^{i+j+1-m}).$$

From this, by hypothesis 4,

$$0 \leq i + j + 1 - m \leq s. \tag{2.42}$$

Now suppose $r \geq s$. Then, again by hypothesis 4,

$$\frac{P}{Q} - \frac{P_{ij}}{Q_{ij}} = [f - R_{ij}] - [f - \frac{P}{Q}] = (z^s) - (z^r) = (z^s)$$

$$\Longrightarrow P\, Q_{ij} - P_{ij}\, Q = Q_{ij}\, Q\, (z^s). \tag{2.43}$$

From the properties of the polynomials (P, Q) and (P_{ij}, Q_{ij}), we get

$$\deg [P\, Q_{ij}] \leq i + j - m$$

$$\deg [P_{ij}\, Q] \leq i + j - m$$

so that the left-hand side of (2.43) has no powers of z with exponent greater than $i + j - m$. But the right-hand side contains no powers of z with exponent less than s, $s \geq i + j - m + 1$ by (2.42). Therefore the two sides of (2.43) have no nontrivial terms in common, and each side must vanish identically. Thus

$$PQ_{ij} - P_{ij}Q = 0$$

and so, finally,

$$\frac{P}{Q} = \frac{P_{ij}}{Q_{ij}} = R_{ij}, \quad \text{by (2.41).}$$

This completes the proof of the Padé Theorem.

REMARKS: Padé called the theorem just proved "fundamental" to his theory. One is therefore surprised to find that the place of this important theorem in the Padé theory has been obscured in some recent work.

The distinction between the defining properties I and II of the Padé approximation, and the conclusions of the Padé theorem, is most clearly explained with the aid of an example:

Consider $f(z) = \cos z = 1 - \frac{z^2}{2} + \ldots$, and let $i = j = 1$.

The defining properties I and II give the (unique) numerator and denonimator polynomials

$$N_{11}(z) = z = D_{11}(z).$$

Check: $\deg N_{11} = 1 = \deg D_{11} = 1 = \deg D_{11}$, and

$$f(z)D_{11}(z) - N_{11}(z) = -\frac{z^3}{2} + \ldots = (z^3).$$

Now the Padé theorem asserts that the expansion of the quotient

$$R_{11}(f, z) = \frac{N_{11}(z)}{D_{11}(z)}$$

in ascending powers of z, agrees with more leading terms of the power series for f than does the expansion of any other rational function whose numerator and denominator are of degrees not exceeding j and i, respectively.

In our example, then, Theorem 2.6 claims that, of all rational functions of the form $R(z) = \frac{az+b}{cz+d}$, the unique one whose expansion agrees with the most terms of

$$\cos z = 1 - \frac{z^2}{2} + \ldots$$

is $R_{11} = 1$. The point to note is that the theorem does not yield an explicit numerical index that allows one to deduce a priori how many terms in the expansion of f are matched by the Padé approximant.

Such an index is, however, provided in the classical definition of the Padé approximant: Property II of the Padé approximant gives an explicit least upper bound, namely Z^{i+j}, on the powers of z which are matched in the expansion

$$f(z)D_{ij}(z) - N_{ij}(z).$$

This distinction between the defining properties of the Padé approximant, and the resulting properties asserted in the Padé theorem, is not always respected in the recent literature. We give three specific instances:

a. Baker, in his recent (1965) study of the convergence properties of sequences of Padé approximants, states [1, p. 3]:

"In the [N, M] Padé approximant the numerator has degree M and the denominator degree N. The coefficients are determined by equating like powers of z in the following equations:

$$f(z)Q(z) - P(z) = Az^{M+N+1} + Bz^{M+N+2} + \ldots, \quad Q(0) = 1.0$$

where $P(z)/Q(z)$ is the [N, M] Padé approximant to $f(z)$."

This characterization is clearly inconsistent, as shown by taking $f(z) = \cos z$, $M = N = 1$. The first of the two equations gives $P(z) = Q(z) = z$, $A = -\frac{1}{2}$; so $Q(0) = 0 \neq 1$.

b. Shanks, in his 1954 dissertation [14, p. 21] characterizes the Padé approximant for f by two properties:

"Property 1. R_{kn} may be written as the ratio of two polynomials:

$$R_{kn} = N_{kn}/D_{kn}$$

with the degree of $N_{kn} \leq n$ and the degree of $D_{kn} \leq k$. But D_{kn} does not vanish identically.

Property 2. The power series of R_{kn} agrees with that of $f(z)$ to a higher power of z than any other rational function with degrees of numerator and denominator no greater than n and k, respectively."

Shanks then cites Wall [17, p. 378] as a reference for the assertion:

"Property 2 is equivalent to the condition

$$f(z)D_{kn} - N_{kn} = (z^{k+n+1})."$$

Actually, Wall neither proves nor even states that the two conditions are equivalent. The weakness of Shanks' claim is evident from the same counterexample used before:

For $f(z) = \cos z$, $N_{11} = D_{11} = 1$ satisfy Shanks' properties 1 and 2, yet

$$f(z)D_{11} - N_{11} = -\frac{z^2}{2} + \ldots \neq (z^3).$$

c. In his 1962 book Matrix Iterative Analysis, Varga [15, p. 266] defines the Padé approximant for $f(z) = \sum_{v=0}^{\infty} a_v z^v$ as the quotient of polynomials $n_{pq}(z)$ and $d_{pq}(z)$ which are respectively of degree q and p. Assuming

$$d_{pq}(0) \neq 0,$$

Varga now selects for each pair of nonnegative integers p and q those polynomials $n_{pq}(z)$ and $d_{pq}(z)$ such that the Taylor's series expansion of $n_{pq}(z)/d_{pq}(z)$ about the origin agrees with as many leading terms of $f(z)$ as possible. Varga then claims "it is evident that the expression

$$d_{pq}(z)f(z) - n_{pq}(z) = O(|z|^{p+q+1}), \quad |z| \to 0,$$

gives rise to $p + q + 1$ linear equations in (the unknown coefficients), whose solution determines these unknown coefficients."

The inconsistency of Varga's assertions is readily evident from the previously used counterexample:

For $f(z) = \cos z$, $d_{11}(0) \neq 0 \Longrightarrow d_{11}(z) = 1$ and $n_{11}(z) = 1$, but

$$d_{11}(z) \cos z - n_{11}(z) = -\frac{z^2}{2} \neq O(|z|^3).$$

To place this discussion in proper perspective within the Padé approximation theory, the following should be added. First, it is of course possible to construct a consistent theory of the Padé approximation using Varga's definition. Cheney (1966) [2, p. 174] has taken this approach, carefully avoiding the pitfalls along the way.

Second, the two approaches <u>are</u> equivalent for that class of functions which have a <u>normal</u> Padé table. This assertion will be proved in Corollary 2.9, with the aid of the next theorem.

2.5 Normal Padé Approximants.

Theorem 2.7 [13, p. 425] [17, p. 394]

Hypotheses:

1. $f(z) = \sum_{k=0}^{\infty} a_k z^k$, $a_0 \neq 0$

2. (i, j) is an ordered pair of nonnegative integers, and $R_{ij}(f, z)$ is the (i, j) Padé approximant for f.

3. (P_{ij}, Q_{ij}) is the unique (i, j) Padé pair for f, with

$$\deg P_{ij} = p, \qquad \deg Q_{ij} = q. \tag{2.44}$$

Conclusions:

1. There exists a nonnegative integer r such that the power series $[f\, Q_{ij} - P_{ij}]$ starts exactly with the power $z^{p+q+r+1}$, or else $f\, Q_{ij} - P_{ij} = 0$. In the latter event, we set $r = \infty$.

2. The $(q + r_1, p + r_2)$ Padé approximant for f equals R_{ij}, with

$r_1, r_2 = 0, 1, \ldots, r$ in case r is finite,

and $r_1, r_2 = 0, 1, \ldots$ in case r is infinite.

3. No entry other than those enumerated in conclusion 2 is equal to R_{ij}.

Proof: The defining properties of the Padé approximant imply

$$0 \leq p \leq j, \qquad 0 \leq q \leq i \tag{I}$$

$$f\, Q_{ij} - P_{ij} = (z^{i+j+1}). \tag{II}$$

Assertion 1 of the theorem is immediate.

By hypothesis 3, (P_{ij}, Q_{ij}) have the following five properties (reference Theorem 2.4):

(i) $P_{ij}(0) = a_0, \quad Q_{ij}(0) = 1;$

(ii) $p \leq j - \lambda, \quad q \leq i - \lambda, \quad \lambda \geq 0;$

(iii) $\lambda + p + q + r \geq i + j;$

(iv) P_{ij} and Q_{ij} are relatively prime;

(v) $R_{ij}(f, z) = \dfrac{P_{ij}(z)}{Q_{ij}(z)}$.

To prove assertions 2 and 3 of the theorem, let (u, v) be a pair of nonnegative integers. By Theorem 2.4, necessary and sufficient conditions for

$$R_{uv}(f, z) = P_{ij}(z)/Q_{ij}(z) \tag{2.45}$$

are that there exists a nonnegative integer k such that

$$p \leq v - k, \quad q \leq u - k \tag{2.46}$$

$$k + p + q + r \geq u + v. \tag{2.47}$$

Our task is to solve these inequalities for u, v, and $k (\geq 0)$.

Condition (2.46) is equivalent to

$$u + v \geq u + k + p$$

$$u + v \geq v + k + q$$

or, combined with (2.47),

$$k + p + q + r \geq u + v \geq \begin{cases} u + k + p \\ v + k + q \end{cases} .$$

Hence

$$q + r \geq u, \qquad p + r \geq v. \tag{2,48}$$

Moreover, $k \geq 0$ and (46) implies

$$u \geq q, \qquad v \geq p$$

so that the following conditions are necessary for (2.45):

$$q + r \geq u \geq q, \qquad p + r \geq v \geq p. \tag{2.49}$$

The inequalities (49) validate assertion 3 of the theorem.

To complete the proof of the theorem, we only need to demonstrate that for each choice of (u, v) in accordance with (49), there exists an integer $k \geq 0$ satisfying (46) and (47). We choose

$$k = \min (u - q, v - p). \tag{2.50}$$

Then $k \leq u - q$, $k \leq v - p$ imply (2.46).

Suppose $u - q \leq v - p$.

Then

$$\begin{aligned} k + p + q + r &= (u - q) + p + q + r \\ &= u + p + r \\ &\geq u + v, \quad \text{by } (2.49). \end{aligned}$$

Similarly, $u - q \geq v - p \Longrightarrow k + p + q + r \geq u + v$, by (2.49).

Therefore (2.47) is satisfied by k as defined in (2.50).

It follows that (2.49) are both necessary and sufficient conditions for (2.45), and assertion 2 verified.

Assertions 2 and 3 are verified.

Remarks: Wall [17, p. 395] calls r the order of the (i, j) Padé approximant. When the approximant R_{ij} is normal, it is distinct from all other entries in the Padé table, and thus $r = 0$. For this case we have the following corollary of Theorem 2.7.

Corollary 2.8 [13, p. 425]

Hypothesis:

1. $f(z) = \sum_{k=0}^{\infty} a_k z^k, \ a_0 \neq 0.$

2. $R_{ij}(f, z) = \dfrac{N_{ij}(z)}{D_{ij}(z)}$ is the (i, j) Padé approximant for f,

with N_{ij} and D_{ij} relatively prime.

Conclusion: The following conditions are necessary and sufficient for R_{ij} to be normal:

(i) $\deg N_{ij} = j, \ \deg D_{ij} = i.$

(ii) The expansion of $f D_{ij} - N_{ij}$ in ascending powers of z starts exactly with the power z^{i+j+1} (not with a higher power).

Proof:

Sufficiency: Suppose (i) and (ii) are true. Apply Theorem 2.7. Then $p = j, \ q = i \Longrightarrow r = 0$ by conclusion 1 of Theorem 2.7, and R_{ij} is normal, by conclusion 3.

Necessity:

Let (P_{ij}, Q_{ij}) be the (i, j) Pade' pair for f. By the uniqueness of R_{ij},

$$\frac{P_{ij}}{Q_{ij}} = \frac{N_{ij}}{D_{ij}} .$$

Since both quotients are clear of common (nonconstant) factors, the numerator polynomials are equal up to a constant factor, and the denominator polynomials are equal up to the same constant factor.

Applying Theorem 2.4, there exists $\lambda \geq 0$ such that

$$p = \deg P_{ij} = \deg N_{ij} \leq j - \lambda$$

$$q = \deg Q_{ij} = \deg D_{ij} \leq i - \lambda$$

and $$[fD_{ij} - N_{ij}] = (z^{i+j+1-\lambda}).$$

By Theorem 2.7, there exists $r \geq 0$ such that

$$i + j + 1 - \lambda \leq p + q + r + 1. \tag{2.51}$$

a. Suppose $p + q < i + j$, that is, (i) is not satisfied.

If $\underline{\lambda = 0}$, then (2.51) implies

$$i + j + 1 < i + j + r + 1,$$

therefore, $\quad 0 < r.$

If $\underline{\lambda > 0}$, then $p + q \leq i + j - 2\lambda$, by definition of λ. This, together with (2.51) implies $i+j+1-\lambda \leq i+j-2\lambda+r+1$, so $0 < \lambda \leq r$ and $0 < r$.

Now $0 < r$ implies that R_{ij} is not normal, by Theorem 2.7, Conclusion 2.

By contraposition, R_{ij} is normal only if $p + q \geq i + j$. But $j \geq p$, $i \geq q$. Thus (i) is a necessary condition for R_{ij} to be normal, since

$$j \geq p \geq i + j - q \geq j \Longrightarrow p = j;$$

$$i \geq q \geq i + j - p \geq i \Longrightarrow q = i.$$

b. Suppose (ii) does not hold.

Then the expansion of $fD_{ij} - N_{ij}$ in ascending powers of z starts with $z^{i+j+r+1}$, $r > 0$. Again, by Theorem 2.7, R_{ij} occurs in at least 4 positions of the Padé table $\Longrightarrow R_{ij}$ is not normal. This completes the proof of the corollary.

Corollary 2.9.

Hypotheses:

1. $f(z) = \sum_{k=0}^{\infty} a_k z^k, \qquad a_0 \neq 0.$

2. The Padé table $[R_{uv}(f, z)]$ for f is normal.

3. (i, j) is an ordered pair of nonnegative integers.

4. N_{ij}, D_{ij} are polynomials, $\deg N_{ij} \leq j$, $\deg D_{ij} \leq i$, $D_{ij} \neq 0$.

Conclusion:

The following two conditions are equivalent:

(i) $f(z)D_{ij}(z) - N_{ij}(z) = (z^{i+j+1});$

(ii) $f(z) - \dfrac{N_{ij}(z)}{D_{ij}(z)} = (z^{i+j+1}).$

Proof:

If (i) holds, then hypothesis 4 implies

$$\frac{N_{ij}(z)}{D_{ij}(z)} = R_{ij}(f, z)$$

and

$$f - \frac{N_{ij}}{D_{ij}} = f - R_{ij}.$$

But the Padé table for f is normal. Therefore, by Corollary 2.8, the expansion of $f(z) - R_{ij}(f, z)$ in ascending powers of z starts exactly with the power z^{i+j+1}, and (ii) follows.

Now suppose (ii) holds. Let (P, Q) be the (i, j) Padé pair for f. Since f is normal, Corollary 2.8 implies that the power series $fQ - P$ starts exactly with the power z^{i+j+1}. But $Q(0) = 1$, so the expansion of $f - \frac{P}{Q} = f - R_{ij}$ starts with the same power z^{i+j+1}. By Theorem 2.6, Conclusion 1, we have for the given polynomials N_{ij}, D_{ij} (hypothesis 4)

$$f - N_{ij}/D_{ij} = (z^r), \quad r \leq i + j + 1,$$

since $\deg N_{ij} \leq j$, $\deg D_{ij} \leq i$. But $r \geq i + j + 1$, because of (ii). Theorem 2.6, Conclusion 2, shows that $r = i + j + 1$ implies

$$\frac{N_{ij}}{D_{ij}} = \frac{P}{Q} = R_{ij}. \tag{2.52}$$

The polynomials P and Q are relatively prime, and $\deg P = j$, $\deg Q = j$, because f is normal (Corollary 2.8). But this, combined with (2.52) and hypothesis 4, gives

$$j \geq \deg N_{ij} \geq \deg P = j \Longrightarrow \deg N_{ij} = j \tag{2.53}$$

$$j \geq \deg D_{ij} \geq \deg Q = i \Longrightarrow \deg Q_{ij} = i. \tag{2.54}$$

Together, (2.52 - 2.54) imply $N_{ij} = cP$, $D_{ij} = cQ$, for some constant $c \neq 0$. Therefore

$$fD_{ij} - N_{ij} = \frac{1}{c}[fQ - P] = (z^{i+j+1}), \tag{2.55}$$

by Corollary 2.8(ii). Thus we have shown (ii) $\Longrightarrow$ (i).

The proof of Corollary 2.9 is complete.

2.6 The E-Array

So far, we have considered rational approximations related to power series of the form

$$f(z) = \sum_{k=0}^{\infty} a_k z^k, \qquad a_0 \neq 0. \tag{2.56}$$

In connection with the theory of linear dynamical systems to be taken up in Chapter IV, V and VI, we will be looking for transfer functions which approximate series of the form

$$Z(s) = \sum_{k=0}^{\infty} a_k s^{-k-1}. \tag{2.57}$$

Of course, the series (2.56) and (2.57) may be transformed into each other by means of the simple relationships

$$f(z) = z^{-1}Z(z^{-1}) \tag{2.58}$$

$$Z(s) = s^{-1}f(s^{-1}). \tag{2.59}$$

However, the following important distinctions between the two series are observed.

(i) If the series (2.57) is rewritten as a power series in $z = s^{-1}$, the constant term is zero, thus violating a condition which has been assumed in the development of the Padé approximation theory.

(ii) The Padé table is essentially a symmetric structure in the sense that the power series expansion of $[f(z)]^{-1}$ has the same form as that of $f(z)$. In contrast, the power series expansion of $[Z(s)]^{-1}$ has not the same form as that of $Z(s)$.

Definition: Suppose f and Z are given, as in (2.56) and (2.57), and suppose the Padé table for f is normal. Following Wynn [30, p. 149], we define the E-array for Z to be a lower triangular matrix with rational elements $E_i^{(j)}(Z, s)$ of the form

$$E_i^{(j)}(Z, s) = \frac{P_i^{(j)}(s)}{Q_i^{(j)}(s)} \tag{2.60}$$

where $Q_i^{(j)}$ is a polynomial of ith degree,

$$Q_i^{(j)}(s) = \sum_{k=0}^{i} q_{ik}^{(j)} s^k, \tag{2.61}$$

$P_i^{(j)}$ is a function of the form

$$P_i^{(j)}(s) = s^{-j} \sum_{k=0}^{i+j-1} p_{ik}^{(j)} s^k, \qquad P_0^{(0)} = 0, \tag{2.62}$$

and the series expansion of $E_i^{(j)}(Z, s)$ in inverse powers of s agrees with that of $Z(s)$ as far as the term containing s^{-2i-j}. The elements of the E-array appear in the order shown below:

$$\begin{array}{cccccc}
E_0^{(0)} & & & & & \\
E_0^{(1)} & E_1^{(0)} & & & & \\
E_0^{(2)} & E_1^{(1)} & E_2^{(0)} & & & \\
E_0^{(3)} & E_1^{(2)} & E_2^{(1)} & E_3^{(0)} & & \\
\vdots & \vdots & \vdots & \vdots & \ddots & \\
E_0^{(j)} & E_1^{(j-1)} & E_2^{(j-2)} & E_3^{(j-3)} & \cdots & E_j^{(0)}
\end{array} \tag{2.63}$$

The element $E_i^{(j)}$ stands at the intersection of the $(i + 1)$th column and the $(j + 1)$th diagonal.

Clearly, the constituents $P_i^{(j)}$, $Q_i^{(j)}$ of the function $E_i^{(j)}$ may be displayed in the same arrangement. If we specify $q_{ii}^{(j)} = 1$ in (2.61), then, by Corollary 2.8,

$$s^{-i}Q_i^{(j)}(s) = Q_{i,i+j-1}(z), \qquad sz = 1$$

and (2.64)

$$s^{-i+1}P_i^{(j)}(s) = P_{i,i+j-1}(z), \qquad sz = 1$$

where (P_{nm}, Q_{nm}) is the (n, m) Padé pair for f.

The close relationship between the Padé table and the E-array is shown even more clearly in the following proposition.

Proposition 2.10 [30, p. 150]

Hypothesis:

Suppose $$f(z) = \sum_{k=0}^{\infty} a_k z^k, \qquad a_0 \neq 0, \tag{2.65}$$

and the Padé table for f is normal.

Conclusion: If we define $Z(s) = \sum_{k=0}^{\infty} a_k s^{-k-1}$, $sz = 1$, then the whole E-array for Z may be obtained from the Padé table for f by transposing the Padé table, deleting the terms lying above the super-diagonal (i.e., the diagonal starting with the second term of the first row in the transposed table), and placing the quantity $E_0^{(0)} = 0$ at the peak of the array.

Conversely, part of the Padé table for f may be obtained from the E-array for Z by removing the entry $E_0^{(0)}$ and transposing the E-array about the diagonal $E_i^{(1)}$, $i = 0, 1, \ldots$.

Proof: The (i, j) Padé pair (P_{ij}, Q_{ij}) may be regarded as a vector pair

$$(p_{i0}, p_{i1}, \ldots, p_{ij}), (q_{0j}, q_{ij}, \ldots, q_{ij}) \tag{2.66}$$

so that

$$\begin{aligned} P_{ij}(z) &= (p_{i0}, p_{i1}, \ldots, p_{ij})(1, z, \ldots, z^j)' \\ Q_{ij}(z) &= (q_{0j}, q_{1j}, \ldots, q_{ij})(1, z, \ldots, z^i)'. \end{aligned} \tag{2.67}$$

Similarly, the entry $E_i^{(j)}$ in the E-array may be regarded as a vector pair

$$(p_{i0}^{(j)}, p_{i1}^{(j)}, \ldots, p_{i,i+j-1}^{(j)}), (q_{i0}^{(j)}, q_{i1}^{(j)}, \ldots, q_{ii}^{(j)}) \tag{2.68}$$

so that, by (5.17) and (5.18),

$$P_i^{(j)}(s) = s^{-j}(p_{i0}^{(j)}, p_{i1}^{(j)}, \ldots, p_{i,i+j-1}^{(j)})(1, s, \ldots, s^{i+j-1})',$$

$$Q_i^{(j)}(s) = (q_{i0}^{(j)}, q_{i1}^{(j)}, \ldots, q_{ii}^{(j)})(1, s, \ldots, s^i)'. \tag{2.69}$$

Now substitute in (2.64), using the appropriate expressions shown on the right of (2.67) and (2.69), to verify the assertions made in the proposition.

2.7 Padé Approximation for Power Series of Nonnegative Order

The classical definition of the Padé table assumes that the power series to be approximated has a nonzero constant term [11], [13], [14], [17]. The assumption of zero-order series serves the purpose of simplifying the statements and proofs of theorems. The restriction can be removed without difficulty, as will now be shown.

Existence (Generalization of Theorem 2.2): Let g be a nonzero power series of order $\sigma = \sigma(g)$; that is,

$$g(z) = z^\sigma f(z) \tag{2.70}$$

where

$$f(z) = \sum_{k=0}^{\infty} a_k z^k, \qquad a_0 \neq 0.$$

Let (i,j) be an ordered pair of nonnegative integers, $j \geq \sigma$. Let $m = j - \sigma$, and let (P_{im}, Q_{im}) be the (i,m) Padé pair for f, defined as in Theorem 2.4. The (i,j) Padé approximant for g is

$$R_{ij}(g,z) = \frac{z^\sigma P_{im}(z)}{Q_{im}(z)}. \tag{2.71}$$

Proof:

$$\text{I.} \qquad \deg Q_{im} \le i \ , \qquad \text{by definition of Padé pair;}$$

$$Q_{im}(0) = 1 \ , \qquad \text{by Theorem 2.4;}$$

$$\deg[z^{\sigma}P_{im}] \le j \ , \qquad \text{by definition of Padé pair.}$$

$$\begin{aligned} \text{II.} \qquad g\,Q_{im} - z^{\sigma}P_{im} &= z^{\sigma}[f\,Q_{im} - P_{im}] \\ &= z^{\sigma}(z^{i+m+1}) \\ &= (z^{i+j+1}) \ . \end{aligned}$$

The defining conditions of the Padé approximant for g are therefore satisfied by the pair of polynomials $z^{\sigma}P_{im}(z)$ and $Q_{im}(z)$.

For $0 \le j < \sigma$, we take $R_{ij}(g,z) = 0$, corresponding to $N_{ij} = 0$, $D_{ij} = z^{i}$ in equation (2.2). Clearly,

$$\text{I.} \qquad \deg D_{ij} = i \ , \qquad D_{ij} \ne 0 \ ,$$

$$\deg N_{ij} = 0 \le j \ ;$$

$$\text{II.} \qquad g(z)D_{ij} - N_{ij} = z^{\sigma+i}f(z) = (z^{i+j+1}) \ . \tag{2.72}$$

Uniqueness (Generalization of Theorem 2.3): Having proved the existence of Padé approximants for general power series, the restriction $a_0 \ne 0$ can now be deleted from Theorem 2.3. The proof is unchanged.

Padé Representation (Generalization of Theorems 2.4 and 2.5): Suppose the power series

$$g(z) = \sum_{k=0}^{\infty} a_k z^k$$

is of order $\sigma = \sigma(g)$. Let (i,j) be an ordered pair of nonnegative integers and $R_{ij}(g,z)$ the (i,j) Padé approximant for g. The conclusions of Theorem 2.4 remain valid for g, except conclusions (i) and (iv), which must be changed to read:

(i*) $\quad P_{ij}(z) \equiv 0, \qquad Q_{ij}(z) = z^i, \quad \text{if} \quad 0 \leq j < \sigma;$

$$\lim_{z \to 0} [z^{-\sigma} P_{ij}(z)] = a_\sigma, \quad Q_{ij}(0) = 1, \quad \text{if} \quad j \geq \sigma.$$

(iv*) $\quad P$ and Q are relatively prime, provided that $j \geq \sigma$.

Proof:

1. Represent g as the product $z^\sigma f(z)$, where $f(z)$ is a power series with nonzero constant term.

2. For $j \geq \sigma$, let $m = j-\sigma$. The Padé pair (P_{im}, Q_{im}) and the integer $\lambda \geq 0$ postulated in Theorem 2.4 for f exist and are unique since f satisfies the hypothesis of Theorem 2.4.

3. It is easily verified that, for $j \geq \sigma$, the pair $(z^\sigma P_{im}, Q_{im})$ and integer λ satisfy (i*), (ii) and that

(iii) $\quad z^\lambda [f Q_{im} - P_{im}] = (z^{i+m+1})$

implies

$$z^\lambda [g Q_{im} - z^\sigma P_{im}] = (z^{i+j+1}) \; ; \tag{2.73}$$

(iv) P_{im}, Q_{im} being relatively prime and $Q_{im}(0) = 1$ imply that $z^{\sigma}P_{im}$ and Q_{im} are relatively prime;

(v)
$$R_{im}(f,z) = \frac{P_{im}(z)}{Q_{im}(z)}$$

implies

$$R_{ij}(g,z) = z^{\sigma}\frac{P_{im}(z)}{Q_{im}(z)} . \qquad (2.74)$$

4. For $j < \sigma$, the proof leading to equation (2.72) applies, and the Padé pair $(0,z^{i})$, with $\lambda = 0$, satisfies conclusions (i^*), (ii), (iii) and (v).

To extend Theorem 2.5 to the general case, it is necessary only to change conclusion (i) to (i^*), and (v) to (iv^*), as stated above, with corresponding obvious modifications in the proof.

Padé's Fundamental Proposition: Let f be a nonzero power series, and let $\sigma(f)$ denote the order of f. Theorem 2.6 was proved for $\sigma(f) = 0$. It remains true for $\sigma(f) > 0$.

Proof: Let (i,j) be an ordered pair of nonnegative integers, and let P, Q, r, s be defined as in Theorem 2.6.

For $j \geq \sigma(f)$, the proof of the generalized version of Theorem 2.6 completely parallels the proof given in section 2.4.

Suppose $0 \leq j < \sigma(f)$. Then $R_{ij}(f,z) \equiv 0$, and

$$f(z) - R_{ij}(f,z) = (z^{s}) \qquad (2.75)$$

where $s = \sigma(f) \geq j+1$.

We will show that $r \geq s \Longrightarrow \frac{P}{Q} = 0$. Then it follows that $r \not\geq s$, and the proof will be complete.

Certainly, by hypothesis 4, $r \geq s$ and $R_{ij} = 0$ imply

$$\frac{P}{Q} = [f - R_{ij}] - [f - \frac{P}{Q}] = (z^s) - (z^r) = (z^s) \Longrightarrow P = (z^s)Q \ . \tag{2.76}$$

The left-hand side of (2.76) has no powers of z with exponent greater than j, while the right-hand side contains none with exponent smaller than $s \geq j+1$. Therefore the two sides of (2.76) have no nontrivial terms in common, and each side must vanish identically. Thus

$$P = 0$$

and

$$\frac{P}{Q} = R_{ij} = 0 \ .$$

Normal Padé Approximants. Theorem 2.7 and its corollary 2.9 are generalized as follows.

Theorem 2.11

Hypotheses:

1. $f(z) = \sum_{0}^{\infty} a_k z^k$, $\sigma = \sigma(f)$ is the order of f.

2. (i,j) is an ordered pair of nonnegative integers and $R_{ij}(f,z)$ is the (i,j) Padé approximant for f.

3. (P_{ij}, Q_{ij}) is the unique (i,j) Padé pair for f, with

$$\deg P_{ij} = p \ , \qquad \deg Q_{ij} = q \ . \tag{2.77}$$

Conclusions:

1. There exists a nonnegative integer r such that the power series $[fQ_{ij} - P_{ij}]$ starts exactly with the power $z^{p+q+r+1}$, or else $fQ_{ij} - P_{ij} = 0$ and $r = \infty$.

2. The $(q+r_1, p+r_2)$ Padé approximant for f equals R_{ij}, with

$$r_1, r_2 = 0, 1, \cdots, r \qquad \text{in case } r \text{ is finite}$$

and

$$r_1, r_2 = 0, 1, \cdots \qquad \text{in case } r \text{ is infinite.}$$

3. No entry other than those enumerated in conclusion 2 is equal to R_{ij}, provided $j \geq \sigma$.

Proof: The special case $\sigma = 0$ was treated in Theorem 2.7.

Suppose $0 < \sigma \leq j$. By the generalization of Theorem 2.4,

$$\text{(i*)} \qquad \lim_{z \to 0} [z^{-\sigma} P_{ij}(z)] = a_\sigma , \qquad Q_{ij}(0) = 1 .$$

Also, there exists an integer $\lambda \geq 0$ such that

$$\text{(ii)} \qquad p \leq j - \lambda , \qquad q \leq i - \lambda ;$$

$$\text{(iii)} \qquad z^\lambda [fQ_{ij} - P_{ij}] = (z^{i+j+1}) ;$$

(iv) P_{ij} and Q_{ij} are relatively prime;

$$\text{(v)} \qquad R_{ij}(f,z) = \frac{P_{ij}(z)}{Q_{ij}(z)} .$$

From (iii),

$$fQ_{ij} - P_{ij} = (z^{i+j-\lambda+1}) .$$

But $i+j-\lambda+1 \geq p+q+\lambda+1$, so either there is an integer $r \geq \lambda$ satisfying assertion 1, or $fQ_{ij} - P_{ij} = 0$.

Let (u,v) be a pair of nonnegative integers. By the generalized Theorem 2.4, necessary and sufficient conditions for

$$R_{uv}(f,z) = \frac{P_{ij}(z)}{Q_{ij}(z)} \tag{2.78}$$

are that there exists a nonnegative integer k such that

$$p \leq v-k \;, \qquad q \leq u-k \qquad \text{(Theorem 2.4, ii)}$$

and

$$k+p+q+r \geq u+v \qquad \text{(Theorem 2.4, iii).}$$

These conditions are equivalent to

$$k+p+q+r \geq u+v \geq \begin{cases} u+k+p \\ v+k+q \end{cases} .$$

Since $k \geq 0$, we obtain

$$q+r \geq u \geq q \;, \qquad p+r \geq v \geq p \;. \tag{2.79}$$

The inequalities (2.79) validate assertion 3 of the theorem, for $j \geq \sigma$. The proof of assertion 2 is identical to that given in the context of Theorem 2.7 (see p. 34).

<u>Counterexample</u> to show Theorem 2.11 fails for $j < \sigma$.

Consider $f = z^3$ and $j = 0,1,2$. For every nonnegative integer i, the (i,j) Padé approximant is $R_{ij}(f,z)$. The Padé table for $f = z^3$ is shown below. Similarly, every function f of order $\sigma > 0$ has a Padé table whose σ leading columns are zeros. By definition, the Padé tables for all such functions are anormal.

<u>Padé Table for</u> $f = z^3$.

	$j = 0$	1	2	3	4	.
$i = 0$	0	0	0	z^3	z^3	.
1	0	0	0	z^3	z^3	.
2	0	0	0	z^3	z^3	.
3	0	0	0	z^3	z^3	.
4	0	0	0	z^3	z^3	.
.	.	.	.	.	.	

<u>Corollary 2.12</u>

Hypotheses:

1. $f(z) = \sum_0^\infty a_k z^k$, σ = order of f.

2. $R_{ij}(f,z) = \dfrac{N_{ij}(z)}{D_{ij}(z)}$, $j \geq \sigma$, is the (i,j) Padé approximant for f, with N_{ij} and D_{ij} relatively prime.

Conclusion:

The following conditions are necessary and sufficient for R_{ij} to be normal:

(i) deg $N_{ij} = j$, deg $D_{ij} = i \geq 0$.

(ii) The expansion of $[fD_{ij} - N_{ij}]$ is of order $i+j+1$.

Generalization of Corollary 2.9 and Proposition 2.10.

In view of Corollary 2.12 above, the requirement $a_o \neq 0$ is implied by the hypothesis that f has a normal Padé table. Therefore, it is not necessary to state explicitly that $a_o \neq 0$.

III. REPRESENTATION OF PADÉ APPROXIMANT BY DETERMINANTS.

In Chapter 2, we were concerned with properties of the Padé table for arbitrary power series $f(z) = \sum_{k=0}^{\infty} a_k z^k$, $a_0 \neq 0$. We continue this investigation in the present chapter, with the aim of expressing the Padé approximants for f as the ratios of determinants, explicitly in terms of the coefficients of f (Theorem 3.5).

One of the properties of the Padé table discussed in Chapter 2 was the geometrical pattern that governs the occurrence of equal approximants: If the table for a power series f contains two equal Padé approximants, then there must be a square block of $(r + 1)^2$ equal approximants (Theorem 2.7). Frank [21, pp. 92-94] gave necessary and sufficient conditions for the Padé table for f to contain a square block with corners (q, p), $(q + r, p)$, $(q + r, p + r)$, $(q, p + r)$, where p, q, and r are arbitrary nonnegative integers (Theorem 3.5). We include a proof of Frank's theorem that is, perhaps, a little easier to follow than the versions given in the original paper or by Wall [17, pp. 395-398].

Theorem 3.7 expresses the Padé approximant for the power series f explicitly in terms of the coefficients $(a_0, a_1, \ldots)$ of the series. The beauty of the method lies in the fact that it proceeds directly to the computation of relatively prime numerator and denominator polynomials. The method thus avoids any problems that might arise if the Padé approximant is to be cleared of common factors in the numerator and denominator.

3.1 Definitions.

The following notation will be used for certain frequently recurring Hankel matrices and determinants: Given the sequence $\{a_k: \ k = 0, 1, \ldots\}$ and two nonnegative integers $r, n,$ we define

$$S_r^{(n)} = \begin{bmatrix} a_n & a_{n+1} & \cdots & a_{n+r-1} \\ a_{n+1} & a_{n+2} & \cdots & a_{n+r} \\ \cdot & \cdot & \cdots & \cdot \\ a_{n+r-1} & a_{n+r} & \cdots & a_{n+2r-2} \end{bmatrix} \tag{3.1}$$

$$\begin{aligned} \Delta_r^{(n)} &= \det S_r^{(n)} \qquad (r > 0) \\ \Delta_0^{(n)} &= 1. \end{aligned} \tag{3.2}$$

Moreover, for given nonnegative integers $i, j, N,$ and for a given power series f

$$f(z) = \sum_{k=0}^{\infty} a_k z^k,$$

let A_N denote the Nth partial sum of f, that is

$$A_N(z) = \sum_{k=0}^{N} a_k z^k. \tag{3.3}$$

Taking $a_n = 0$ for $n < 0$, we define

$$T_{ij}(z) = \det \begin{bmatrix} z^i A_{j-i} & z^{i-1} A_{j-i+1} & \cdots & z^0 A_j \\ a_{j-i+1} & a_{j-i+2} & \cdots & a_{j+1} \\ \cdot & \cdot & \cdots & \cdot \\ a_j & a_{j+1} & \cdots & a_{j+i} \end{bmatrix}$$

$$U_{ij}(z) = \det \begin{bmatrix} z^i & z^{i-1} & \cdots & z^0 \\ a_{j-i+1} & a_{j-i+2} & \cdots & a_{j+1} \\ \cdot & \cdot & \cdots & \cdot \\ a_j & a_{j+1} & \cdots & a_{j+i} \end{bmatrix} \tag{3.4}$$

In particular, these definitions imply the following properties:

$$\deg T_{ij} \leq j, \quad \deg U_{ij} \leq i, \tag{3.5}$$

$$T_{0j} = A_j, \qquad U_{0j} = 1, \tag{3.6}$$

$$T_{ij}(0) = (-1)^i A_j \Delta_i^{(j-i+1)}$$

$$U_{ij}(0) = (-1)^i \Delta_i^{(j-i+1)}. \tag{3.7}$$

3.2 Some Properties of Hankel Determinants.

Certain properties of the Hankel determinants $\Delta_r^{(n)}$ will be useful later. These properties are stated in the following two classical lemmas.

Lemma 3.1. [23, p. 120][4, p. 25]

For all positive values of n and r,

$$[\Delta_r^{(n)}]^2 = \Delta_r^{(n-1)}\Delta_r^{(n+1)} - \Delta_{r+1}^{(n-1)}\Delta_{r-1}^{(n+1)} \tag{3.8}$$

the determinants being those defined in equations (3.1) and (3.2).

The proof of this lemma is straightforward, but tedious and not very enlightening. Householder [24, pp. 116-117] indicates the proof for $r = 1, 2$; Henrici [4, pp. 25-26] sketches a similar verification for all admissible r.

Lemma 3.2. [4, p. 28] (i) Bieberbach's version [25, pp. 319-321]:

Let $f(z) = \sum_{k=0}^{\infty} a_k z^k$ be rational,

$$f(z) = \frac{b_0 + b_1 z + \ldots + b_p z^p}{c_0 + c_1 z + \ldots + c_q z^q}, \quad c_0 \neq 0, \; c_q \neq 0. \tag{3.9}$$

Then

$$\Delta_{q+1}^{(n)} = 0 \quad \text{for all } n \geq \max(0, p - q + 1).$$

Conversely, let p and q be integers such that $p > q - 1$, and $\Delta_q^{(n)} \neq 0$, $\Delta_{q+1}^{(n)} = 0$ for $n \geq p - q + 1$. Then $f(z)$ is a rational function of the form (3.9), with $c_0 \neq 0$, $c_q \neq 0$.

(ii) Dienes' version [26, p. 323]:

The necessary and sufficient condition that the power series $\sum_{0}^{\infty} a_k z^k$ should represent a rational function is that there be a number q such that $\Sigma \Delta_{q+1}^{(n)} z^n = P(z)$ is a polynomial. Then the least value of q is the degree of the denominator, and the degree of $P(z)$ is not less than $p - q$, p being the degree of the numerator.

3.3 Conditions for a Block in the Padé Table.

Lemma 3.3. (See [14, p. 22] for special case where $a_0 \neq 0$.)

Hypotheses:

1. $f(z) = \sum_{k=0}^{\infty} a_k z^k$.

2. $R_{ij}(f, z)$ is the (i, j) Padé approximant for f.

Conclusions:

If

$$\Delta_i^{(j-i+1)} = 0,$$

then

$$R_{i,j-1} = R_{i-1,j} = R_{i-1,j-1} = R_{ij}.$$

Proof: To compute the Padé approximant $R_{i-1,j-1}$, we let

$$R_{i-1,j-1} = N_{i-1,j-1}/D_{i-1,j-1}$$

where

$$D_{i-1,j-1}(z) = \sum_{k=0}^{i-1} d_k z^k$$

$$N_{i-1,j-1}(z) = \sum_{k=0}^{j-1} n_k z^k$$

The denominator and numerator polynomials thus defined satisfy the condition II for the $(i - 1, j - 1)$ Padé approximant, that is

$$fD_{i-1,j-1} - N_{i-1,j-1} = (z^{i+j-1}) \tag{3.10}$$

Form the product

$$f(z)D_{i-1,\ j-1}(z) = \sum_{k=0}^{\infty} c_k z^k$$

where

$$c_k = \sum_{u+v=k} a_u d_v .$$

In order for $(N_{i-1,\ j-1}, D_{i-1,\ j-1})$ to satisfy condition II for the $(i - 1,\ j - 1)$ Padé approximant, we need

$$n_k = c_k \qquad (k = 0,\ 1,\ \ldots,\ j - 1) \tag{3.11}$$

and

$$c_k = 0 \qquad (k = j,\ j + 1,\ \ldots,\ j + i - 2).$$

The last equations, written in matrix form, are

$$\begin{bmatrix} a_{j-i+1} & a_{j-i+2} & \cdots & a_j \\ a_{j-i+2} & a_{j-i+3} & \cdots & a_{j+1} \\ \cdot & \cdot & \cdots & \cdot \\ a_{j-1} & a_j & \cdots & a_{j+i-2} \end{bmatrix} \begin{bmatrix} d_{i-1} \\ d_{i-2} \\ \cdot \\ \cdot \\ \cdot \\ d_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} . \tag{3.12}$$

Eq. (3.12) states that the i-vector $(d_{i-1}, d_{i-2}, \ldots, d_0)$ must be orthogonal to the row vectors of the $[a]$-matrix. Since there are only $i-1$ row vectors in $[a]$, they can span at most an $i-1$ dimensional subspace of R^i. Therefore (3.12) always has a nontrivial solution. Choosing any such solution, we now determine the n_k by (3.11). Then

$(N_{i-1,\ j-1}, D_{i-1,\ j-1})$ satisfy conditions I and II, and we have correctly computed $R_{i-1,\ j-1}$ since it is unique.

Suppose now that $\Delta_i^{(j-i+1)} = 0$. Then the row vector $(a_j, a_{j+1}, \ldots, a_{j+i-1})$ is in the row space of the $[a]$ matrix in (3.12) and is, therefore, orthogonal to the vector $(d_{i-1}, d_{i-2}, \ldots, d_0)$ chosen to satisfy (3.12). But this implies

$$c_{j+i-1} = 0 \tag{3.13}$$

and as a result we can improve (3.10) to read

$$f\,D_{i-1,\ j-1} - N_{i-1,\ j-1} = (z^{i+j}). \tag{3.14}$$

This increase in the order of approximation means that we may set $(N_{i-1,\ j}, D_{i-1,\ j}) = (N_{i-1,\ j-1}, D_{i-1,\ j-1})$, and

$$R_{i-1,\ j} = \frac{N_{i-1,\ j-1}}{D_{i-1,\ j-1}} = R_{i-1,\ j-1}. \tag{3.15}$$

For the polynomials of the pair $(N_{i-1,j}, D_{i-1,j})$ clearly have degrees no greater than j and $i-1$, respectively, and thus property I. Further, by (3.14), they satisfy property II of the $(i-1, j)$ Padé approximant for f. By a similar argument, we have $R_{i,j-1} = R_{i-1,j-1}$.

Again, we may set $N_{ij} = zN_{i-1,\ j-1}$ and $D_{ij} = zD_{i-1,\ j-1}$. The pair (N_{ij}, D_{ij}) satisfies both properties I and II of the (i, j) Padé approximant for f: Property I because $\deg N_{ij} \leq j$, $\deg D_{ij} \leq i$; and property II since from (3.14)

$$fD_{ij} - N_{ij} = z[fD_{i-1,\ j-1} - N_{i-1,\ j-1}] = (z^{i+j+1}).$$

This proves the lemma.

Lemma 3.4. (Frobenius) [3, pp. 1-3]

Hypotheses:

1. $f(z) = \sum_{k=0}^{\infty} a_k z^k, \quad a_0 \neq 0$

2. (i, j) is an ordered pair of nonnegative integers, and $R_{ij}(f, z)$ is the (i, j) Padé approximant for f.

Conclusions:

1. The determinants T_{ij}, U_{ij} defined in (3.4) are relatively prime (i) if, and (ii) only if

$$\Delta_i^{(j-i+1)} \neq 0.$$

2. If $\Delta_i^{(j-i+1)} \neq 0$, then $R_{ij} = T_{ij}/U_{ij}$.

Proof:

1(ii). Suppose $\Delta_i^{(j-i+1)} = 0$. Then by (3.7) and (3.8)

$$T_{ij}(0) = 0 \text{ and } U_{ij}(0) = 0.$$

Therefore, unless T_{ij} and U_{ij} are identically zero, they have a common factor z. If T_{ij}, U_{ij} are zero polynomials, they are not relatively prime, by definition [9, p. 72].

1(i) and 2. Suppose $\Delta_i^{(j-i+1)} = (-1)^i c \neq 0$. (3.16)

Then the i columns of

$$S_i^{(j-i+1)} = \begin{bmatrix} a_{j-i+1} & a_{j-i+2} & \cdots & a_j \\ a_{j-i+2} & a_{j-i+3} & \cdots & a_{j+1} \\ \cdot & \cdot & \cdots & \cdot \\ a_j & a_{j+1} & \cdots & a_{j+i-1} \end{bmatrix} \tag{3.17}$$

are linearly independent.

Define

$$A^k = \begin{bmatrix} a_{j-k+1} \\ a_{j-k+2} \\ \vdots \\ a_{j-k+i} \end{bmatrix} \qquad (k = 0, 1, \ldots, i) \tag{3.18}$$

Then $S_i^{(j-i+1)}$ can be represented as the row vector

$$S_i^{(j-i+1)} = (A^i, A^{i-1}, \ldots, A^1). \tag{3.19}$$

Each A^k is an $i \times 1$ column vector. Since the i columns $A^i, A^{i-1}, \ldots, A^1$ of $S_i^{(j-i+1)}$ are linearly independent, they must span the space R^i. Therefore A^0 is in the column space of $S_i^{(j-i-1)}$, and $m = 0$ is the largest index such that the column vectors $A^m, A^{m+1}, \ldots, A^i$ are linearly dependent.

Now consider the polynomials

$$\begin{aligned} P(z) &= \frac{1}{c} T_{ij}(z) \\ Q(z) &= \frac{1}{c} U_{ij}(z). \end{aligned} \tag{3.20}$$

From (3.7) and (3.5) respectively, we have

$$\begin{aligned} &\text{(i)} \quad P(0) = a_0, \quad Q(0) = 1 \\ &\text{(ii)} \quad \deg P \leq j, \quad \deg Q \leq i. \end{aligned} \tag{3.21}$$

To verify that

$$\text{(iii)} \quad fQ - P = (z^{i+j+1}) \tag{3.22}$$

we proceed as follows. The value of $T_{ij}(z)$ is unchanged if we add to the first row a linear combination of the other rows. In particular, we add z^{j+k} times the $(k+1)$th row, $k = 1, 2, \ldots, i$. Then the first term of the first row becomes

$$z^i A_{j-i} + z^i \sum_{k=1}^{i} z^{j+k-i} a_{j+k-i} = z^i A_j. \tag{3.23}$$

The other terms of the first row are changed similarly, increasing by i the index of each A_n $(n = j-i, j-i+1, \ldots, j)$. Now

$$T_{ij}(z) = \det \begin{bmatrix} z^i A_j & z^{i-1} A_{j+1} & \cdots & z^0 A_{j+i} \\ a_{j-i+1} & a_{j-i+2} & \cdots & a_{j+1} \\ \cdot & \cdot & \cdots & \cdot \\ a_j & a_{j+1} & \cdots & a_{j+i} \end{bmatrix}. \tag{3.24}$$

Therefore we can write

$$A_{j+i} U_{ij} - T_{ij} = \det \begin{bmatrix} z^i (A_{j+i} - A_j) & z^{i-1}(A_{j+i} - A_{j+1}) & \cdots & 0 \\ a_{j-i+1} & a_{j-i+2} & \cdots & a_{j+1} \\ \cdot & \cdot & \cdots & \cdot \\ a_j & a_{j+1} & \cdots & a_{j+i} \end{bmatrix}. \tag{3.25}$$

By inspection, z^{i+j+1} divides every polynomial appearing in the first row of the determinant in (3.25). Therefore z^{i+j+1} divides the determinant, and consequently

$$A_{j+i}U_{ij} - T_{ij} = (z^{i+j+1}). \tag{3.26}$$

But now

$$\begin{aligned} fQ - P &= [A_{j+i}Q - P] + \sum_{k=j+i+1}^{\infty} a_k z^k Q \\ &= (z^{i+j+1}) \end{aligned} \tag{3.27}$$

since the proportionality factor c in (3.20) is merely a nonzero constant.

As we have shown in the preceeding steps, the pair (P, Q) defined in (3.20) satisfies all the conditions of Theorem 2.5, with $m = 0$. By result (v) of that theorem, we have

$$R_{ij}(f, z) = P(z)/Q(z) = T_{ij}(z)/U_{ij}(z) \tag{3.28}$$

as was to be shown (Conclusion 2).

Also, by result (v) of Theorem 2.5, P and Q are relatively prime polynomials, and therefore T_{ij} and U_{ij} are relatively prime.

This completes the proof of the lemma.

Remark: The hypothesis $a_o \neq 0$ may be removed if one restricts j to values not less than the order of f and proceeds as in the generalization of Theorem 2.2. (See Section 2.7.) Smaller nonnegative values of $j \leq \sigma(f)$ give rise to the trivial case. $R_{ij}(f,z) = 0$.

Theorem 3.5. [21, p. 93][17, p. 395]

Hypotheses:

1. $f(z) = \sum_{k=0}^{\infty} a_k z^k, \qquad a_0 \neq 0.$

2. (i, j) is an ordered pair of nonnegative integers, and $R_{ij}(f, z)$ is the (i, j) Padé approximant for f.

3. (P_{ij}, Q_{ij}) is the (i, j) Padé pair of polynomials for f, postulated in Theorem 2.4, with

$$\deg P_{ij} = p, \qquad \deg Q_{ij} = q.$$

4. r is the unique nonnegative integer such that the power series $[fQ_{ij} - P_{ij}]$ starts exactly with the power $z^{p+q+r+1}$, (See Theorem 2.7.)

Conclusion:

The following five conditions are necessary and sufficient for $R_{ij}(f, z) = R_{uv}(f, z)$, with (u, v) ranging over the $(r + 1)^2$ values defined by $(q \leq u \leq q + r, \quad p \leq v \leq p + r)$:

$$\begin{array}{lll} \text{(i)} & \Delta_q^{(p-q+1)} \neq 0 & \\ \text{(ii)} & \Delta_q^{(p-q+2)} \neq 0 & \\ \text{(iii)} & \Delta_{q+1}^{(p-q)} \neq 0 & \\ \text{(iv)} & \Delta_{q+r+1}^{(p-q+1)} \neq 0 & \\ \text{(v)} & \Delta_{q+k}^{(p-q+1)} = 0 & (k = 1, 2, \ldots, r). \end{array} \tag{3.29}$$

Note: If $r = 0$, the condition (v) is not present. By Theorem 2.7, $r = 0$ implies $(q, p) = (i, j)$ and, with this substitution, conditions (i) through (iv) are necessary and sufficient for $R_{ij}(f, z)$ to be normal. [11, p. S. 34][13, p. 427].

Proof: By Theorem 2.7, the following five conditions are necessary and sufficient for $R_{ij} = R_{uv}$ $(q \leq u \leq q + r,\ p \leq v \leq p + r)$:

$$
\begin{array}{lll}
(\text{i}') & R_{qp} \neq R_{q-1,p-1} & \\
(\text{ii}') & R_{qp} \neq R_{q-1,\ p} & \\
(\text{iii}') & R_{qp} \neq R_{q,\ p-1} & \\
(\text{iv}') & R_{qp} \neq R_{q+r+1,p+r+1} & \\
(\text{v}') & R_{qp} = R_{q+k,p+k} & (k = 1, 2, \ldots, r).
\end{array}
\tag{3.30}
$$

It is convenient to include the condition (i') although, by Theorem 2.7, we have (i') $\Longleftrightarrow$ (ii') and (iii').

From Lemma 3.3, we obtain (by contraposition):

(i') $\Longrightarrow$ (i) [let $(i, j) = (q, p)$ in (3.9)].

(ii') $\Longrightarrow$ (ii) [let $(i, j) = (q, p + 1)$ in (3.9)].

(iii') $\Longrightarrow$ (iii) [let $(i, j) = (q + 1, p)$ in (3.9)].

(iv') and (v') $\Longrightarrow$ (iv) [let $(i, j) = (q + r + 1, p + r + 1)$ in (3.9)].

It remains to show (v') $\Longrightarrow$ (v). By Theorem 2.4, in order for (P_{ij}, Q_{ij}) to satisfy condition II of the $(q + k, p + k)$ Pade' approximant $(k = 1, 2, \ldots, r)$, we need

$$z^k[fQ_{ij} - P_{ij}] = (z^{p+q\ \ +2k+1}) \tag{3.31}$$

As in Theorem 2.4, let

$$P_{ij}(z) = \sum_{v=0}^{p} n_v z^v, \qquad n_0 = a_0;$$

$$Q_{ij}(z) = \sum_{v=0}^{q} d_v z^v, \qquad d_0 = 1;$$

and $$f(z)Q_{ij}(z) = \sum_{v=0}^{\infty} c_v z^v,$$

where $$c_v = \sum_{\alpha+\beta=v} a_\alpha d_\beta .$$

Then the condition (3.31) can be written

$$n_v = c_v \qquad (v = 0, 1, \ldots, p)$$

and

$$c_v = 0 \qquad (v = p+1, p+2, \ldots, p+q+k) \tag{3.32}$$

$$(k = 1, 2, \ldots, r).$$

Equations (3.32), written in matrix form, are

$$\begin{bmatrix} a_{p-q+1} & a_{p-q+2} & \cdots & a_{p+1} \\ a_{p-q+2} & a_{p-q+3} & \cdots & a_{p+2} \\ \cdot & \cdot & \cdots & \\ a_{p+k} & a_{p+k+1} & \cdots & a_{p+k+q} \end{bmatrix} \begin{bmatrix} d_q \\ d_{q-1} \\ \vdots \\ d_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad d_0 = 1 \tag{3.33}$$

$$(k = 1, 2, \ldots, r).$$

Choose any k from $\{1, 2, \ldots, r\}$. Equation (3.33) can be satisfied only if the $q + 1$ columns of the $[a]$-matrix are linearly dependent. Therefore $\Delta_{q+k}^{(p-q+1)} = 0$, $k = 1, 2, \ldots, r$, and thus (v) is necessary for (v').

We now prove that (3.29) is sufficient for (3.30). From (i) we conclude (Lemma 3.2, proof of conclusion 2) that there exists a pair of relatively prime polynomials (P, Q) such that

$$R_{qp} = P/Q, \tag{3.34}$$

$$\deg P \leq p, \quad \deg Q \leq q, \quad P(0) = a_0, \quad Q(0) = 1; \tag{3.35}$$

$$fQ - P = (z^{p+q+1}). \tag{3.36}$$

The polynomials (P, Q) are uniquely determined by the conditions (3.34) and (3.35), and the same conditions imply that the coefficients of $Q = 1 + \sum_{v=0}^{q} d_q^* z^q$ satisfy the matrix equation (3.33), with $k = 0$.

From (ii) and (iii) it follows that P and Q are exactly of degree p and q, respectively. For Q, this result is immediate from (3.33), since $k = 0$, $\Delta_q^{(p-q+2)} \neq 0$ gives $d_q^* \neq 0$. Suppose that

$$P(z) = a_0 + \sum_{v=0}^{p} n_v^* z^v.$$

Then (3.36) gives

$$n_v^* = \sum_{\alpha+\beta=v} a_\alpha d_\beta^* \qquad (v = 0, 1, \ldots, p).$$

In particular, $$n_p^* = [a_{p-q}, a_{p-q+1} \cdots a_p] \begin{bmatrix} d_q^* \\ d_{q-1}^* \\ \vdots \\ d_0^* \end{bmatrix}, \quad d_0^* = 1. \tag{3.37}$$

But (iii) implies that $S_{q+1}^{(p-q)}$ is a nonsingular matrix:

$$S_{q+1}^{(p-q)} = \begin{bmatrix} a_{p-q} & a_{p-q+1} & \cdots & a_p \\ a_{p-q+1} & a_{p-q+2} & \cdots & a_{p+1} \\ \cdot & \cdot & \cdots & \cdot \\ a_p & a_{p+1} & \cdots & a_{p+q} \end{bmatrix} \tag{3.38}$$

Therefore, the rows of $S_{q+1}^{(p-q)}$ span the space R^{q+1}. The vector $(d_q, d_{q-1}, \ldots, d_0)$ is a nonzero vector in R^{q+1}, orthogonal to all rows of $S_{q+1}^{(p-q)}$ except the first row (by equation (3.33), with $k = 0$). Therefore the scalar product in (3.37) cannot vanish, and (iii) $\Longrightarrow n_p \neq 0$.

Now $R_{qp} = P/Q$, (P, Q) relatively prime and $\deg P = p$, $\deg Q = q$, implies

(i') $R_{qp} \neq R_{q-1,p-1}$

(ii') $R_{qp} \neq R_{q-1,p}$

(iii') $R_{qp} \neq R_{q,p-1}$.

To prove (v'), let $R_{q+k,p+k} = N^{(k)}/D^{(k)}$, where

$$N^{(k)} = \sum_{v=0}^{p+k} n_v^{(k)} z^v, \qquad D^{(k)} = \sum_{v=0}^{q+k} d_v^{(k)} z^v.$$

Since $fD^{(k)} - N^{(k)} = (z^{p+q+2k+1})$, we have

$$
\begin{array}{c}
\uparrow \\ p+k+1 \\ \text{rows} \\ \downarrow
\end{array}
\begin{bmatrix}
0 & 0 & \cdots & 0 & a_0 \\
0 & 0 & \cdots & a_0 & a_1 \\
\cdot & \cdot & & \cdot & \cdot \\
a_{p-q} & a_{p-q+1} & \cdots & a_{p+k-1} & a_{p+k}
\end{bmatrix}
\begin{bmatrix}
d_{q+k}^{(k)} \\ d_{q+k-1}^{(k)} \\ \vdots \\ d_0^{(k)}
\end{bmatrix}
=
\begin{bmatrix}
n_0^{(k)} \\ n_1^{(k)} \\ \vdots \\ n_{p+k}^{(k)}
\end{bmatrix}
\qquad (3.39)
$$

where $d_0^{(k)} = 1$, $n_0^k = a_0$.

Also

$$
\begin{array}{c}
\uparrow \\ q+k \\ \text{rows} \\ \downarrow
\end{array}
\begin{bmatrix}
a_{p-q+1} & a_{p-q+2} & \cdots & a_{p+k+1} \\
a_{p-q+2} & a_{p-q+3} & \cdots & a_{p+k+2} \\
\cdot & \cdot & \cdots & \cdot \\
a_{p+k} & a_{p+k+1} & \cdots & a_{p+q+2k}
\end{bmatrix}
\begin{bmatrix}
d_{q+k}^{(k)} \\ d_{q+k-1}^{(k)} \\ \vdots \\ d_0^{(k)}
\end{bmatrix}
=
\begin{bmatrix}
0 \\ 0 \\ \vdots \\ 0
\end{bmatrix}
\qquad (3.40)
$$

Since $\Delta_{q+1}^{(p-q+1)} = 0$, comparison with (3.33) shows that a nontrivial solution of (3.40) for $k = 1$ is

$$d_0^{(1)} = 0, \quad d_v^{(1)} = d_{v-1} \qquad (v = 1, 2, \ldots, q+1). \quad (3.41)$$

Substituting in (3.39), one obtains

$$n_0^{(1)} = 0, \quad n_v^{(1)} = n_{v-1}^{(1)} \qquad (v = 1, 2, \ldots, p+1). \quad (3.42)$$

Hence $$N^{(1)}/D^{(1)} = P/Q$$

and (v') holds for $k = 1$. Then, since $\Delta_{q+2}^{(p-q-1)} = 0$, it follows that for $k = 2$, (3.40) has the nontrivial solution

$$d_0^{(2)} = 0, \quad d_v^{(2)} = d_{v-1}^{(1)} \qquad (v = 1, 2, \ldots, q+2)$$

while (3.39) gives

$$n_0^{(2)} = 0, \quad n_v^{(2)} = n_{v-1}^{(2)} \qquad (v = 1, 2, \ldots, p+2).$$

Consequently, $N^{(2)}/D^{(2)} = N^{(1)}/D^{(1)} = P/Q$. On continuing this argument, we conclude that (v') holds.

Finally, (iv') holds. For if not, we would have

$$z^{r+1}[fQ - P] = (z^{p+q+2r+3})$$

which is impossible by virtue of (iv).

This completes the proof of Theorem 3.5.

3.4 Determinantal Form of Padé Approximant.

Lemma 3.6.

Hypotheses:

1. $\{a_k\}$ is an infinite sequence of real numbers, $a_\sigma \neq 0$, $\sigma \geq 0$, $a_k = 0$ for $k < \sigma$.

2. (i, j) is an ordered pair of nonnegative integers.

3. The i-dimensional column vectors $A^k (k = 0, 1, 2, \ldots, i)$ are defined by

$$A^k = \begin{bmatrix} a_{j-k+1} \\ a_{j-k+2} \\ \vdots \\ a_{j-k+i} \end{bmatrix} . \qquad (3.43)$$

$S = S_i^{(j-i+1)}$ is the Hankel matrix defined by (3.17).

4. m is the largest index such that the column vectors $A^m, A^{m+1}, \ldots, A^i$ are linearly dependent.

Conclusions:

$$\Delta_{i-m}^{(j-i+1)} \neq 0 \quad \text{if} \quad m = 0 \quad \text{or} \quad m = i. \tag{3.44}$$

$$\Delta_{i-m+1}^{(j-i+1)} = 0 \quad \text{if} \quad 1 \leq m < i. \tag{3.45}$$

Proof:

(i) Suppose $m = 0$.

Then the columns of $S = [A^i, A^{i-1}, \ldots, A^1]$ are linearly independent and

$$\det S = \Delta_i^{(j-i+1)} \neq 0.$$

(ii) Suppose $m = i$.

Then $A^i = 0$ and, in particular, $a_{j-i+1} = 0$.

Therefore

$$\Delta_1^{(j-i+1)} = a_{j-i+1} = 0.$$

Also, trivially by definition (3.2),

$$\Delta_0^{(j-i+1)} = 1 \neq 0.$$

(iii) Suppose $1 \leq m < i$.

$\Delta_{i-m+1}^{(j-i+1)}$ is the minor consisting of the first $i - m + 1$ rows of the matrix $M = [A^i, A^{i-1}, \ldots, A^m]$. But the columns of M are linearly dependent, so

$$\Delta_{i-m+1}^{(j-i+1)} = 0.$$

This completes the proof of the lemma.

Theorem 3.7.

Hypotheses:

1. $f(z) = \sum_{k=0}^{\infty} a_k z^k, \qquad a_0 \neq 0.$

2. (i, j) is an ordered pair of nonnegative integers.

3. d is the smallest nonnegative integer such that

$$\Delta_{i-d}^{(j-i+1)} \neq 0. \tag{3.46}$$

Conclusions:

1. $0 \leq d \leq i$.

2. The polynomials $T_{i-d, j-d}$ and $U_{i-d, j-d}$, defined by equation (3.4), have at most j and i zeros, respectively, (counting multiplicities). They have no zeros in common.

3. The (i, j) Padé approximant for f is

$$R_{ij}(f, z) = \frac{T_{i-d, j-d}(z)}{U_{i-d, j-d}(z)}. \tag{3.47}$$

4. $R_{uv} = R_{ij} \qquad (u = i, i - 1, \ldots, i - d; \; v = j, j - 1, \ldots, j - d).$

Proof: Conclusion 1 is obvious from the definition $\Delta_0^{(n)} = 1$, $n \geq 0$.

Suppose $d = 0$. By Lemma 3.4, $\Delta_i^{(j-i+1)} \neq 0$ implies that T_{ij}, U_{ij} are relatively prime polynomials and

$$R_{ij}(f, z) = \frac{T_{ij}(z)}{U_{ij}(z)} . \tag{3.49}$$

From (3.5), $\deg T_{ij} \leq j$, $\deg U_{ij} \leq i$.

Suppose $d > 0$.

Then $\Delta_k^{(j-i+1)} = 0 \quad (k = i, i - 1, \ldots, i - d + 1)$.

By repeatedly applying Lemma 3.1, we obtain (3.47), since

$$R_{ij} = R_{i-d,j-d} \tag{3.50}$$

$$= \frac{T_{i-d,j-d}}{U_{i-d,j-d}}, \quad \text{by Lemma 3.2.}$$

Conclusion 2 follows directly from $\Delta_{i-d}^{(j-i+1)} \neq 0$, by Lemma 3.4.

In particular, if $d = i$, then by (3.6)

$$R_{ij} = \frac{T_{0,j-i}}{U_{0,j-i}} = A_{j-i}. \tag{3.51}$$

Conclusion 4 is a trivial consequence of (3.50), by Lemma 3.6.

Remarks: The method displayed in Theorem 3.7 depends on finding first the parameter d which, in turn, depends on the distinction between singular and nonsingular matrices. In practice, round-off errors will obscure the distinction. It is, therefore, of some interest to note the effect of an erroneously large choice of d. This will happen if a nearly singular Hankel matrix is considered to be "singular" (by the criteria used in a given computer algorithm). As a result, the computed (i, j) Padé approximant in (3.47) will be cleared of numerator and denominator factors which are not strictly cancellable.

The theorem is a refinement of Lemma 3.4. (Frobenius [3, pp. 1-3].)

Corollary 3.8 [30, p. 159]

Hypotheses:

1. $Z(s) \;=\; \sum_{k=0}^{\infty} a_k s^{-k-1}, \qquad a_0 \neq 0.$ (3.52)

2. The Padé table for $f(z) = \sum_{k=0}^{\infty} a_k z^k$ is normal.

3. (i, j) is an ordered pair of nonnegative integers.

Conclusion:

Let the partial sums Z_N of the infinite series Z be denoted by

$$\begin{aligned} Z_N &= \sum_{k=0}^{N-1} a_k s^{-k-1}, \quad N > 0 \\ Z_0 &= 0 \end{aligned} \tag{3.53}$$

and let $\Delta Z_N \;=\; Z_{N+1} - Z_N \;=\; a_N s^{-N-1}.$ (3.54)

Then the (i, j) element in the E-array for Z is

$$E_i^{(j)}(Z, s) = \frac{\det \begin{bmatrix} Z_j & Z_{j+1} & \cdots & Z_{j+i} \\ \Delta Z_j & \Delta Z_{j+1} & \cdots & \Delta Z_{j+i} \\ \Delta Z_{j+1} & \Delta Z_{j+2} & \cdots & \Delta Z_{j+i+1} \\ \cdot & \cdot & \cdots & \cdot \\ \Delta Z_{j+i-1} & \Delta Z_{j+1} & \cdots & \Delta Z_{j+2i-1} \end{bmatrix}}{\det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \Delta Z_j & \Delta Z_{j+1} & \cdots & \Delta Z_{j+i} \\ \cdot & \cdot & \cdots & \cdot \\ \Delta Z_{j+i-1} & \Delta Z_{j+i} & \cdots & \Delta Z_{j+2i-1} \end{bmatrix}} \tag{3.55}$$

Proof:

As in Chapter II, equation (2.60), we can write

$$E_i^{(j)}(Z,\ s) \quad = \quad \frac{P_i^{(j)}(s)}{Q_i^{(j)}(s)} \tag{3.56}$$

where $Q_i^{(j)}$ is a monic polynomial of degree i, and $P_i^{(j)}$ is a meromorphic function of s. [Equation (2.62)]. By Corollary 2.8, we have, as in equation (2.64)

$$\begin{aligned} s^{-i+1}P_i^{(j)}(s) \quad &= \quad P_{i,\,i+j-1}(z), \qquad sz = 1 \\ &\text{and} \\ s^i Q_i^{(j)}(s) \quad &= \quad Q_{i,\,i+j-1}(z), \qquad sz = 1 \end{aligned} \tag{3.57}$$

where $(P_{i,\,i+j-1}, Q_{i,\,i+j-1})$ is the $(i,\ i+j-1)$ Padé pair for f.

Using the result of Theorem 3.7,

$$E_i^{(j)}(Z,\ s) \quad = \quad \frac{sP_i^{(j)}(s)}{Q_i^{(j)}(s)} \quad = \quad \frac{P_{i,\,j+i-1}(z)}{zQ_{i,\,j+i-1}(z)}\ . \tag{3.58}$$

The Padé table for f is normal, so $\Delta_i^{(j)} \neq 0$ and

$$E_i^{(j)}(Z,\ s) \quad = \quad \frac{T_{i,\,j+i-1}(z)}{zU_{i,\,j+i-1}(z)}\ . \tag{3.59}$$

Now, the respective definitions of A_N in (3.3) and Z_N in (3.53), with $sz = 1$, imply

$$A_{N-1}(z) \quad = \quad \sum_{k=0}^{N-1} a_k z^k \quad = \quad s\sum_{k=0}^{N-1} a_k s^{-k-1} \quad = \quad sZ_N(s). \tag{3.60}$$

By definition (3.4),

$$T_{i,j+i-1}(z) = \det \begin{bmatrix} z^i A_{j-1} & z^{i-1} A_j & \cdots & A_{j+i-1} \\ a_j & a_{j+1} & \cdots & a_{j+i} \\ \cdot & \cdot & \cdots & \cdot \\ a_{j+i-1} & a_{j+i} & \cdots & a_{j+2i-1} \end{bmatrix}$$

$$= \det \begin{bmatrix} s^{-i-1} Z_j & s^{-i} Z_{j+1} & \cdots & s Z_{j+i} \\ s^{j+1} \Delta Z_j & s^{j+2} \Delta Z_{j+1} & \cdots & s^{j+i+1} \Delta Z_{j+i} \\ \cdot & \cdot & \cdots & \cdot \\ s^{j+i} \Delta Z_{j+i-1} & s^{j+i+1} \Delta Z_{j+i} & \cdots & s^{j+2i} \Delta Z_{j+2i-1} \end{bmatrix} \tag{3.61}$$

Similarly,

$$U_{i,j+i-1}(z) = \det \begin{bmatrix} z^i & z^{i-1} & \cdots & 1 \\ a_j & a_{j+1} & \cdots & a_{j+i} \\ \cdot & \cdot & \cdots & \cdot \\ a_{j+i-1} & a_{j+i} & \cdots & a_{j+2i-1} \end{bmatrix}$$

$$= \det \begin{bmatrix} s^{-i} & s^{-i+1} & \cdots & 1 \\ s^{j+1} \Delta Z_j & s^{j+2} \Delta Z_{j+1} & \cdots & s^{j+i+1} \Delta Z_{j+i} \\ \cdot & \cdot & \cdots & \cdot \\ s^{j+i} \Delta Z_{j+i-1} & s^{j+i+1} \Delta Z_{j+i} & \cdots & s^{j+2i} \Delta Z_{j+2i-1} \end{bmatrix} \tag{3.62}$$

where $sz = 1$.

Finally, substitution of (3.61) and (3.62) into (3.59) gives the desired result, and the corollary is proved.

Remark: Comparing Theorem 3.7 and Corollary 3.8, we note that the latter assumes normality of the Padé table as an added restriction on f. If this condition is not fulfilled, certain formal difficulties exist in the derivation of results paralleling those of Theorem 3.7. A theory of E-arrays for the general case is not yet available.

The rational expressions $R_{ij}(f, z)$ or $E_i^{(j)}(Z, s)$ which may be obtained from the set of coefficients a_k, $k = 0, 1, \ldots, 2r - 1$, are those entries in the Padé table lying upon and in the triangle whose vertices coincide with the approximants

$$\begin{array}{lcl} R_{00} & \cdots\cdots & R_{0,2r-1} \\ \vdots & & \\ \vdots & & \\ R_{2r-1,0} & & \end{array} \tag{3.63}$$

and those entries in the E-array which lie in and upon the triangle whose vertices coincide with the functions

$$\begin{array}{lcl} E_0^{(0)} & & \\ \vdots & & \\ \vdots & & E_r^{(0)} \\ \vdots & & \\ E_0^{(2r)} & & \end{array} \tag{3.64}$$

As pointed out by Wynn [30, p. 171], numerical experience supports the claim that, in general, for prescribed values of the arguments z and s, the expressions in the sets (3.63) and (3.64) for which

$$|R_{ij}(f, z) - f(z)| \quad \text{or} \quad |E_n^{(m)}(s) - Z(s)|$$

are a minimum, are given by $i + 1 = j = r$, or $i = j + j = r$, or $m = 0$, $n = r$.

IV. MINIMAL REALIZATIONS OF LINEAR DYNAMICAL SYSTEMS.

One of the problems in the theory of linear dynamical systems is to construct models from input-output data. This is variously known as the problem of "modeling", "process identification", or "constructing a realization". While special, distinctive meaning has been given to each of these terms, the object is generally to find a mathematical model which lends itself to computer simulation.

The explicit determination of such a model is the subject of B.L. Ho's dissertation [6]. In particular, Ho considers the problem of constructing state-variable models of linear, stationary, finite-dimensional, multivariate dynamical systems. His algorithm is based on the solution of the "algebraic realization problem", defined as follows:

> Given an infinite sequence of real $(p \times m)$ matrices, $\mathcal{Y} = (Y_0, Y_1, Y_2, \dots)$, to find real matrices F, G, H such that
>
> $$H F^k G = Y_k \qquad (k = 0, 1, 2, \dots) \tag{4.1}$$
>
> where
> $$\begin{aligned} F &= (n \times n) \text{ matrix} \\ G &= (n \times m) \text{ matrix} \\ H &= (p \times n) \text{ matrix.} \end{aligned}$$

Following the accepted terminology, any solution (F, G, H) of the algebraic realization problem is called a realization of $\mathcal{Y}$, n is the dimension of the realization, and a solution with the smallest possible dimension n_0 is called a minimal realization of $\mathcal{Y}$.

4.1 B.L. Ho's Existence Theorem.

Ho gave (i) necessary and sufficient conditions for a solution of the above problem to exist (Proposition 4.1), and (ii) a numerically simple algorithm for the construction of minimal realizations (Theorem 4.4). He also derived other abstract results of realization theory, some of which are reviewed and extended in this and the following chapters.

The existence of a realization is established by the following proposition which plays a fundamental role in Ho's theory.

Proposition 4.1 [6, p. 11]

The sequence $\mathcal{Y} = (Y_0, Y_1, Y_2, \ldots)$ has a realization if and only if there exists an integer r such that

$$Y_{r+j} = \sum_{i=1}^{r} \alpha_i Y_{r-i+j}, \qquad j = 0, 1, \ldots \tag{4.2}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_r$ are suitably fixed constants.

Ho gives a beautifully simple proof of this proposition. Before proceeding with it, we make some definitions and present a lemma which will be useful throughout this chapter.

Definitions:

1. Given a sequence $\mathcal{Y} = (Y_0, Y_1, \ldots)$, and two nonnegative integers r, k, we define the $(pr \times mr)$ matrix

$$S_r^{(k)} = \begin{bmatrix} Y_k & Y_{k+1} & \cdots & Y_{k+r-1} \\ Y_{k+1} & Y_{k+2} & \cdots & Y_{k+r} \\ \cdot & \cdot & \cdots & \cdot \\ Y_{k+r-1} & Y_{k+r} & \cdots & Y_{k+2r-2} \end{bmatrix} \tag{4.3}$$

Note: The definition (4.3) is consistent with that made in Chapter III when y is a scalar sequence. In this case $(p = m = 1)$, we set $\det S_r^{(k)} = \Delta_r^{(k)}$, as in (3.2).

2. If y satisfies the linear recursion relation (4.2), let M be the $(pr \times pr)$ block companion matrix

$$M = \begin{bmatrix} O_p & I_p & O_p & \cdots & O_p \\ O_p & O_p & I_p & \cdots & O_p \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ O_p & O_p & O_p & \cdots & I_p \\ \alpha_r I_p & \alpha_{r-1} I_p & \alpha_{r-2} I_p & \cdots & \alpha_1 I_p \end{bmatrix} \qquad (4.4)$$

where $O_p = (p \times p)$ zero matrix

$I_p = (p \times p)$ identity matrix.

3. If y satisfies the linear recursion relation (4.2), let N be the $(mr \times mr)$ block companion matrix

$$N = \begin{bmatrix} O_m & O_m & \cdots & O_m & \alpha_r I_m \\ I_m & O_m & \cdots & O_m & \alpha_{r-1} I_m \\ O_m & I_m & \cdots & O_m & \alpha_{r-2} I_m \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ O_m & O_m & \cdots & I_m & \alpha_1 I_m \end{bmatrix}. \qquad (4.5)$$

4. For given positive integers u, v $(u \leq v)$, let E_{uv} be the $(u \times v)$ matrix

$$E_{uv} = [I_u \; 0]. \qquad (4.6)$$

In particular, if r is a positive integer and $v = ur$, we abbreviate

$$E_u = E_{u,ur} \tag{4.7}$$

Lemma 4.2.

Hypothesis:

1. $\mathcal{Y} = (Y_0, Y_1, \ldots)$ is a sequence of matrices satisfying (4.2) for some positive integer r.

2. i, j are nonnegative integers.

Conclusion:

$$S_r^{(i+j)} = M^i S_r^{(j)} = S_r^{(j)} N^i \tag{4.8}$$

the matrices S, M, N having been defined in (4.3), (4.4), (4.5), respectively.

Proof: Let j be an arbitrary nonnegative integer. We shall prove the lemma by induction on i. (4.8) is obviously true for $i = 0$.

Suppose (4.8) holds for an arbitrary fixed integer $i \geq 0$. Then, because the coefficients $(\alpha_1, \alpha_2, \ldots, \alpha_r)$ of M and N satisfy (4.2),

$$M^{i+1} S_r^{(j)} = M S_r^{(i+j)} = S_r^{(i+j+1)} = S_r^{(i+j)} N = S_r^{(j)} N^{i+1}.$$

Therefore (4.8) holds for $i + 1$. By mathematical induction, the lemma holds for all $i \geq 0$. But j was arbitrary, so the proof is valid for all nonnegative i, j.

Proof of Proposition 4.1 [6, p. 11]:

Assume that a realization (F, G, H) of dimension n exists.

Let $\psi(z) = z^n - \beta_1 z^{n-1} - \beta_2 z^{n-2} - \dots - \beta_n$, $\beta_0 = 1$ be an annihilating polynomial of F. (The Cayley-Hamilton Theorem guarantees the existence of ψ.) Then

$$0 = HF^j\psi(F)G = HF^{n+j}G - \sum_{i=1}^{n} \beta_i HF^{n-i+j}G$$

so that (4.2) holds with $r = n$ and $\alpha_i = \beta_i$, $i = 1, 2, \dots, n$.

Conversely, suppose (4.2) is true. By Lemma 4.2, this implies

$$M^i S_r^{(0)} = S_r^{(i)} \qquad (i = 0, 1, 2, \dots).$$

The first block in $S_r^{(i)}$ is Y_i. Therefore

$$Y_i = E_p S_r^{(i)} E_m' = E_p M^i S_r^{(0)} E_m',$$

and

$$\begin{aligned} F &= M \\ G &= S_r^{(0)} E_m' \\ H &= E_p \end{aligned} \tag{4.9}$$

is a realization of $\mathcal{Y}$, by (4.1).

4.2 B.L. Ho's Realization Algorithm

The following lemma prepares the way for Ho's minimal realization algorithm.

Lemma 4.3.

Hypotheses:

1. $\mathcal{Y}$ is a sequence of matrices satisfying (4.2) for some positive integer r.

2. (F, G, H) is a realization of $\mathcal{Y}$.

Conclusion: rank $S_r^{(0)} \leq \dim F$.

Proof: By hypothesis, $Y_k = HF^kG$, all $k \geq 0$. Therefore $S_r^{(0)}$ can be factored as follows:

Let

$$\begin{aligned} V' &= [H' \quad F'H' \quad \ldots \quad (F')^{r-1}H'] \\ W &= [G \quad FG \quad \ldots \quad F^{r-1}G] \end{aligned} \tag{4.10}$$

where the prime denotes the transpose of a matrix. Then

$$S_r^{(0)} = V\,W. \tag{4.11}$$

Now rank $S_r^{(0)} \leq \min(\text{rank } V, \text{ rank } W) \leq \dim F$.

Theorem 4.4. (Ho's realization algorithm)[6, p. 13]

Hypotheses:

1. $\mathcal{Y}$ is a sequence of $(p \times m)$ matrices satisfying (4.2) for some positive integer r.

2. $S_r^{(k)}$ $(k = 0, 1, \ldots)$ are generalized Hankel matrices for $\mathcal{Y}$, as defined in (4.3), with rank $S_r^{(0)} = n$.

3. P and Q are nonsingular matrices, of dimensions $(pr \times pr)$ and $(mr \times mr)$ respectively, and such that

$$PS_r^{(0)}Q = E'_{n,pr}E_{n,mr} \tag{4.12}$$

i.e., $PS_r^{(0)}Q$ is the canonical diagonal form

$$\begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix} \text{ of } S_r^{(0)}. \text{ [22, vol. I; p. 141]}$$

Conclusion:

Let

$$F = E_{n,pr} P S_r^{(1)} Q E'_{n,mr}, \qquad (n \times n) \text{ matrix;}$$

$$G = E_{n,pr} P S_r^{(0)} E'_m, \qquad (n \times m) \text{ matrix;} \tag{4.13}$$

$$H = E_p S_r^{(0)} Q E'_{n,mr}, \qquad (p \times n) \text{ matrix.}$$

Then (F, G, H) is a minimal realization of the sequence $\mathcal{Y}$.

Proof: The existence of suitable matrices P and Q is a well-known fact from linear algebra. [22, pp. 133-141]

Let

$$S^{\#} = Q E'_{n,mr} E_{n,pr} P. \tag{4.14}$$

Then hypothesis 3 gives

$$\begin{aligned} S_r^{(0)} S^{\#} S_r^{(0)} &= S_r^{(0)}; \\ S^{\#} S_r^{(0)} S^{\#} &= S^{\#}. \end{aligned} \tag{4.15}$$

From the definitions of F, P and Q,

$$F^k = E_{n,pr} P S_r^{(k)} Q E'_{n,mr}. \qquad (k = 0,\ 1) \tag{4.16}$$

By Lemma 4.2,

$$\begin{aligned} F^2 &= E_{n,pr} P M S_r^{(0)} S^{\#} S_r^{(0)} N Q E'_{n,mr} \\ &= E_{n,pr} P M S_r^{(0)} N Q E'_{n,mr} \\ &= E_{n,pr} P S_r^{(2)} Q E'_{n,mr}. \end{aligned}$$

Using Lemma 4.2 and (4.15) repeatedly, we have the general result

$$F^k = E_{n,pr} P S_r^{(k)} Q E'_{n,mr} \qquad (k = 0, 1, 2, \ldots). \tag{4.17}$$

Therefore

$$\begin{aligned} HF^kG &= E_p S_r^{(0)} Q E'_{n,mr} E_{n,pr} P S_r^{(k)} Q E'_{n,mr} E_{n,pr} P S_r^{(0)} E'_m \\ &= E_p S_r^{(0)} S^{\#} S_r^{(0)} N^k S^{\#} S_r^{(0)} E'_m \\ &= E_p M^k S_r^{(0)} E'_m \\ &= E_p S_r^{(k)} E'_m \\ &= Y_k. \end{aligned} \tag{4.18}$$

F is an $(n \times n)$ matrix, and rank $S_r^{(0)} = n$ by hypothesis 2. By Lemma 4.3, the realization is minimal.

Remark: Theorem 4.4 solves the algebraic realization problem stated in the introductory paragraphs of this chapter, whenever such a realization exists. The following proposition, due to Ho [6, p. 48], demonstrates that the algebraic realization problem is equivalent to the problem of finding the (minimal) realization of linear, stationary, finite-dimensional dynamical systems from their input-output descriptions. A remarkable feature of this proposition is that it applies equally to discrete-time and continuous-time systems, and that the input-output data may be presented either in the time domain or in the transform domain.

For the discrete-time system

$$x(k+1) = F\,x(k) + G\,u(k)$$
$$y(k) = H\,x(k) \tag{4.19}$$

the time-domain description is given by the pulse-response function

$$Y_k = H\,F^k G, \qquad k = 0, 1, \ldots . \tag{4.20}$$

The transform-domain description is given by the z-transform transfer function

$$T(z) = H(zI - F)^{-1}G. \tag{4.21}$$

For the continuous-time system

$$\frac{dx}{dt} = Fx(t) + Gu(t)$$
$$y(t) = Hx(t) \tag{4.22}$$

the time-domain description is given by the impulse-response function

$$W(t) = H\exp(Ft)\,G, \qquad t \geq 0. \tag{4.23}$$

The transform domain description is given by the Laplace transform transfer function

$$Z(s) = H(sI - F)^{-1}G. \tag{4.24}$$

Proposition 4.5 [6, p. 48]

The following four problems are equivalent to the algebraic realization problem:

(i) Given the function $k \to Y_k$, find a triple (F, G, H) of constant matrices such that (4.20) holds.

(ii) Given the function $z \to T(z)$, find a triple (F, G, H) of constant matrices such that (4.21) holds.

(iii) Given the function $t \to W(t)$, find a triple (F, G, H) of constant matrices such that (4.23) holds.

(iv) Given the function $s \to Z(s)$, find a triple (F, G, H) of constant matrices such that (4.24) holds.

For a proof and further discussion of this proposition, the reader is referred to the paper by Ho and Kalman [36, p. 453].

4.3 The Unique Representation Theorem.

In this section, we present a specialized version of B.L. Ho's minimal realization algorithm. The specialization is accompanied by the achievement of a number of desirable new properties.

For example, the one-to-one correspondence established in Corollary 4.10 has no parallel in Ho's theory. By showing this one-to-one correspondence between minimal realizations (F, G, H) and generating matrices (V, W), we demonstrate that the new algorithm is the sharpest possible, subject to the requirement that every minimal realization may be obtained.

We note, too, that the generating pair (V, W) in Corollary 4.10 is the same pair of matrices whose rank determines the complete controllability and observability of stationary linear dynamical systems. [7; p. 201], [8; p. 170], [27; pp. 499-506], [29; p. 53]. The beauty and significance of the reciprocal relations (4.69) and (4.70) is obvious.

The close similarity, as well as the difference, between the computations for B.L. Ho's algorithm and the new algorithm are brought into sharp focus in Proposition 4.11. B.L. Ho's algorithm is phrased and proved in such a manner that the whole matrices P and Q appear. The new algorithm operates only with submatrices of the matrices P and Q, viz., with the parts which lie in the column and row spaces of $S_r^{(0)}$.

Lemma 4.6.

Hypotheses:

1. $\mathcal{Y}$ is a sequence of $(p \times m)$ matrices satisfying (4.2) for some positive integer r.

2. Rank $S_r^{(0)} = n$, where $S_r^{(0)}$ is the generalized Hankel matrix defined in (4.3).

3. (F, G, H) is a minimal realization for $\mathcal{Y}$, and

$$\begin{aligned} V' &= [H' \quad F'H' \quad \cdots \quad (F')^{r-1}H'] \\ W &= [G \quad FG \quad \cdots \quad F^{r-1}G]. \end{aligned} \tag{4.25}$$

Conclusions:

1. The columns of V are a basis for the column space of $S_r^{(0)}$. The rows of W are a basis for the row space of $S_r^{(0)}$.

2. The pseudo inverses $V^\dagger$, $W^\dagger$ are given by

$$\begin{aligned} V^\dagger &= (V'V)^{-1}V \\ W^\dagger &= W'(WW')^{-1}. \end{aligned} \tag{4.26}$$

3. $VF^kW = S_r^{(k)}, \qquad k = 0, 1, 2, \ldots$.

Note: See Appendix B for definition and properties of pseudo inverse.

Proof:

1. Since (F, G, H) is a minimal realization, we have $\dim F = n$. But

$$S_r^{(0)} = VW$$

and so we have

$$n = \text{rank } S_r^{(0)} \leq \min(\text{rank } V, \text{ rank } W) \leq \max(\text{rank } V, \text{ rank } W) \leq \dim F.$$

Therefore, $\dim F = n$ implies the well-known result [8, pp. 169, 170]

$$\text{rank } V = \text{rank } W = n.$$

Since V has exactly n columns and W has exactly n rows, Conclusion 1 follows immediately. Also, the system represented by (F, G, H) is completely controllable (since $\text{rank } W = n$) and completely observable (since $\text{rank } V = n$). [6; p. 50].

2. To show Conclusion 2, write $V = VI_n$ and $W = I_nW$. Then (4.26) is readily verified by the construction given in Appendix B, Section 2.

3. Since (F, G, H) is a realization, we have

$$Y_i = HF^iG, \qquad i = 0, 1, 2, \ldots .$$

Therefore, the product VF^kW is a block matrix whose matrix elements are precisely the elements of $S_r^{(k)}$.

This completes the proof of the lemma.

<u>Lemma 4.7.</u>

Hypotheses:

1. $\mathcal{Y}$ is a sequence of $(p \times m)$ matrices satisfying (4.2) for some positive integer r.

2. $S_r^{(k)}$ $(k = 0, 1, 2, \ldots)$ are generalized Hankel matrices for $\mathcal{Y}$, as defined in (4.3), with rank $S_r^{(0)} = n$.

3. B and C are matrices with the following properties:

(i) B is a $(pr \times n)$ matrix whose n columns are a basis for the column space of $S_r^{(0)}$;

(ii) C is a $(n \times mr)$ matrix whose n rows are a basis for the row space of $S_r^{(0)}$;

(iii) $BC = S_r^{(0)} = S$, say.

<u>Conclusion:</u>

For each pair of nonnegative integers (i, j),

$$B^\dagger M^i S N^j C^\dagger = (B^\dagger MS C^\dagger)^{i+j} = (B^\dagger S NC^\dagger)^{i+j},$$

the matrices M and N having been defined in (4.4) and (4.5), respectively.

<u>Proof:</u>

By Lemma 4.2,

$$M^i S N^j = S N^{i+j}, \text{ and } MS = SN.$$

Therefore, taking $k = i + j$, it suffices to show that

$$B^{\dagger} S\, N^{k} C^{\dagger} \;=\; (B^{\dagger} S\, NC^{\dagger})^{k}, \quad (k = 0, 1, \ldots). \tag{4.27}$$

By Hypothesis 3, we deduce (see Appendix B.4(ix))

$$S^{\dagger} \;=\; C^{\dagger} B^{\dagger}. \tag{4.28}$$

Certainly, (4.27) is true for $k = 1$.

Suppose (4.7) is true for k equal to some fixed positive integer q. Then

$$\begin{aligned}
(B^{\dagger} S\, NC^{\dagger})^{q+1} &= (B^{\dagger} S\, NC^{\dagger})(B^{\dagger} S\, N^{q} C^{\dagger}) \\
&= B^{\dagger} MSS^{\dagger} S\, N^{q} C^{\dagger}, \quad \text{by (4.28)} \\
&= B^{\dagger} MSN^{q} C^{\dagger}, \quad \text{by definition of } S^{\dagger} \text{ (see Appendix B.1)} \\
&= B^{\dagger} S\, N^{q+1} C^{\dagger}, \quad \text{by Lemma 4.2.}
\end{aligned}$$

Therefore, (4.27) is true for $k = 1, 2, 3, \ldots,$ by induction.

For $k = 0$, we have

$$B^{\dagger} SC^{\dagger} \;=\; [(B'B)^{-1} B'] BC [C'(CC')^{-1}] \;=\; I. \tag{4.29}$$

This completes the proof of the lemma.

<u>Lemma 4.8</u> [6; p. 17]

Hypotheses:

$\mathcal{Y}$ is a sequence of $(p \times m)$ matrices satisfying (4.2) for some positive integer r.

1. Any two minimal realizations (F_k, G_k, H_k), $k = 1, 2$, of the same sequence $\mathcal{Y}$ are isomorphic: There exists a nonsingular matrix T such that

$$
\begin{aligned}
F_2 &= T F_1 T^{-1} \\
G_2 &= T G_1 \\
H_2 &= H_1 T^{-1}.
\end{aligned}
\tag{4.30}
$$

2. The matrix T in (4.30) is given explicitly by

$$T = V_2^{\dagger} V_1 = W_2 W_1^{\dagger} \tag{4.31}$$

where

$$
\begin{aligned}
V_k' &= [H_k' \quad F_k' H_k' \quad \cdots \quad (F_k')^{r-1} H_k'], \\
W_k &= [G_k \quad F_k G_k \quad \cdots \quad F_k^{r-1} G_k], \quad k = 1, 2.
\end{aligned}
\tag{4.32}
$$

<u>Proof</u>:

By Lemma 4.6,

$$S_r^{(0)} = V_1 W_1 = V_2 W_2, \tag{4.33}$$

and

$$S_r^{(1)} = V_1 F_1 W_1 = V_2 F_2 W_2. \tag{4.34}$$

Define

$$T = W_2 W_1^{\dagger}, \quad U = V_1^{\dagger} V_2. \tag{4.35}$$

Assertion:

$$UT = TU = I. \tag{4.36}$$

Proof of assertion (4.36):

By Lemma 4.6, the columns of V_k are a basis for the column space of $S = V_k W_k$, and the rows of W_k are a basis for the row space of S; for $k = 1, 2$.

By the same lemma,

$$V_k^{\dagger} = (V_k' V_k)^{-1} V_k'$$

$$W_k^{\dagger} = W_k' (W_k W_k')^{-1}.$$

Therefore,

$$V_k^{\dagger} V_k = I \quad \text{and} \quad W_k W_k^{\dagger} = I. \tag{4.37}$$

Now

$$\begin{aligned} UT &= V_1^{\dagger} V_2 W_2 W_1^{\dagger} = V_1^{\dagger} V_1 W_1 W_1^{\dagger}, \quad \text{by (4.33)}, \\ &= I, \quad \text{by (4.37)}. \end{aligned}$$

Again,

$$TU = W_2 W_1^{\dagger} V_1^{\dagger} V_2 = W_2 (S_r^{(0)})^{\dagger} V_2, \quad \text{by (4.28)}.$$

Using assertion (4.33), we obtain

$$TU = W_2 W_2^{\dagger} V_2^{\dagger} V_2 = I, \quad \text{by (4.37)}.$$

The assertion (4.36) is proved.

Now (4.36) shows that T and U are nonsingular and that $U = T^{-1}$. Also, from (4.33), (4.35) and (4.37),

$$V_2T = V_2W_2W_1^{\dagger} = V_1W_1W_1^{\dagger} = V_1$$
$$UW_2 = V_1^{\dagger}V_2W_2 = V_1^{\dagger}V_1W_1 = W_1. \qquad (4.38)$$

Substitution for V_1 and W_1 from (4.38) into (4.34) gives

$$V_1F_1W_1 = V_2TF_1UW_2 = V_2F_2W_2$$

which, in turn, yields

$$F_2 = TF_1U = TF_1T^{-1} \qquad (4.39)$$

where

$$T = W_2W_1^{\dagger} = U^{-1}, \quad \text{by (4.36).}$$

But $U^{-1} = (V_1^{\dagger}V_2)^{-1} = V_2^{\dagger}V_1$, by Appendix B.4, properties (ii), (iii), and (ix). Therefore

$$T = V_2^{\dagger}V_1 = W_2W_1^{\dagger}. \qquad (4.40)$$

From (4.32), $$G_k = W_kE'_{n,mr}, \quad k = 1,\ 2; \qquad (4.41)$$

and

$$H_k = E_{n,pr}V_k, \quad k = 1,\ 2. \qquad (4.42)$$

Substitution of $W_2 = U^{-1}W_1 = TW_1$ from (4.38) into (4.41) gives

$$G_2 = TW_1E'_{n,mr} = TG_1. \qquad (4.43)$$

Similarly, substitution of $V_2 = V_1T^{-1}$ from (4.38) into (4.42) gives

$$H_2 = E_{n,pr}V_1T^{-1} = H_1T^{-1}. \qquad (4.44)$$

In view of (4.39), (4.40), (4.43) and (4.44), the lemma is true.

Theorem 4.9 (Unique Representation Theorem)

Hypotheses:

1. $\mathcal{Y}$ is a sequence of $(p \times m)$ matrices satisfying (4.2) for some positive integer r.

2. $S_r^{(k)}$ $(k = 0, 1, 2, \ldots)$ are generalized Hankel matrices for $\mathcal{Y}$, as defined in (4.3), with rank $S_r^{(0)} = n$.

3. (B, C) is an ordered pair of matrices with the following properties:

(i) B is a $pr \times n$ matrix whose n columns are a basis for the column space of $S_r^{(0)}$;

(ii) C is a $(n \times mr)$ matrix whose n rows are a basis for the row space of $S_r^{(0)}$;

(iii) $BC = S_r^{(0)} = S$, say.

Conclusion:

Let

$$\begin{aligned} F &= B^{\dagger} S_r^{(1)} C^{\dagger} && (n \times n) \text{ matrix;} \\ G &= B^{\dagger} S_r^{(0)} E_m' && (n \times m) \text{ matrix;} \\ H &= E_p S_r^{(0)} C^{\dagger} && (p \times n) \text{ matrix.} \end{aligned} \qquad (4.45)$$

Then

1. (F, G, H) is a minimal realization of the sequence $\mathcal{Y}$.

2. Given any minimal realization (F, G, H) for $\mathcal{Y}$, there exists an ordered pair of matrices (B, C) having the properties of Hypothesis 3 and generating (F, G, H) when substituted on the right-hand side of equations (4.45).

3. The pair (B, C) of Conclusion 2 is unique for each minimal realization (F, G, H), and is given explicitly by

$$(B, C) = (V, W) \tag{4.46}$$

where V and W are defined by (4.19).

<u>Proof</u>:

1. Let $S = S_r^{(0)}$ and $S^{\dagger}$ = pseudo inverse of S.

Then, as shown in Appendix B,

$$S^{\dagger} = C^{\dagger}B^{\dagger} \tag{4.47}$$

and

$$SS^{\dagger}S = S.$$

$$
\begin{aligned}
\text{Now } HF^kG &= (E_pS_r^{(0)}C^{\dagger})(B^{\dagger}S_r^{(1)}C^{\dagger})^k(B^{\dagger}S_r^{(0)}E_m') \\
&= (E_pSC^{\dagger})(B^{\dagger}SNC^{\dagger})^k(B^{\dagger}SE_m') \\
&= (E_pSC^{\dagger})(B^{\dagger}SN^kC^{\dagger})(B^{\dagger}SE_m'), \quad \text{by Lemma 4.5} \\
&= E_pSS^{\dagger}SN^kS^{\dagger}SE_m', \quad \text{by (4.47)} \qquad (4.49) \\
&= E_pM^kSE_m', \quad \text{by (4.48)} \\
&= E_pS_r^{(k)}E_m', \quad \text{by Lemma 4.2} \\
&= Y_k.
\end{aligned}
$$

This shows that (F, G, H) is a realization for $\mathcal{Y}$.

F is an $(n \times n)$ matrix and rank $S_r^{(0)} = n$, by Hypothesis 2. Therefore the realization is minimal, by Lemma 4.3. This completes the proof of Conclusion 1.

2. Suppose that (F, G, H) is an arbitrary fixed minimal realization for $\mathcal{Y}$. Then Lemma 4.6 implies that the matrices (V, W) defined by (4.25) have the properties of (B, C) postulated in Hypothesis 3 of our theorem. Therefore, substitution of

$$B^{\dagger} = V^{\dagger} = (V'V)^{-1}V$$
$$C^{\dagger} = W^{\dagger} = W'(WW')^{-1} \tag{4.50}$$

on the right-hand side of (4.45) will give a minimal realization of $\mathcal{Y}$. Let this minimal realization be (F_1, G_1, H_1):

$$F_1 = V^{\dagger}S_r^{(1)}W^{\dagger}$$
$$G_1 = V^{\dagger}S_r^{(0)}E'_m \tag{4.51}$$
$$H_1 = E_pS_r^{(0)}W^{\dagger}.$$

Let $V_1' = [H_1'F_1'H_1' \ldots (F_1')^{r-1}H_1'],\ W_1 = [G_1\ F_1G_1 \ldots F_1^{r-1}G_1].$ (4.52)

By Lemma 4.8,

$$(F, G, H) = (TF_1U, TG_1, H_1U) \tag{4.52}$$

where

$$T = V^{\dagger}V_1 = U^{-1}, \quad U = W_1W^{\dagger} = T^{-1}. \tag{4.54}$$

By Lemma 4.6,

$$S_r^{(0)} = V_1W_1 = VW, \tag{4.55}$$

and

$$S_r^{(1)} = V_1F_1W_1 = VFW. \tag{4.56}$$

Now from (4.53) and (4.54):

$$F = TF_1U = (V^\dagger V_1)F_1(W_1W^\dagger)$$

$$= V^\dagger S_r^{(1)}W^\dagger, \qquad \text{by (4.56).}$$

Using the first of the equations (4.51), we obtain

$$F = F_1. \tag{4.57}$$

Again, from (4.53) and (4.54),

$$G = TG_1 = (V^\dagger V_1)(V^\dagger S_r^{(0)}E_m'). \tag{4.58}$$

But now (using $S = S_r^{(0)}$):

$$V_1V^\dagger S = \begin{bmatrix} H_1 \\ H_1F_1 \\ \vdots \\ H_1F_1^{r-1} \end{bmatrix} V^\dagger S, \quad \text{by (4.52)}$$

$$= \begin{bmatrix} E_pSW^\dagger \\ E_pSW^\dagger V^\dagger MSW^\dagger \\ \vdots \\ E_pSW^\dagger[V^\dagger MSW^\dagger]^{r-1} \end{bmatrix} V^\dagger S, \quad \text{by (4.51)}$$

$$= \begin{bmatrix} E_pSW^\dagger V^\dagger S \\ E_pSW^\dagger(V^\dagger MSW^\dagger)V^\dagger S, \\ \vdots \\ E_pSW^\dagger(V^\dagger M^{r-1}SW^\dagger)V^\dagger S \end{bmatrix} \text{by Lemma 4.7}$$

$$= \begin{bmatrix} E_p \\ E_pM \\ \cdot \\ \cdot \\ \cdot \\ E_pM^{r-1} \end{bmatrix} S, \qquad \text{by (4.28) and Lemma 4.7}$$

$$= I_{pr}S = S. \tag{4.59}$$

Now, from (4.51), (4.58) and (4.59),

$$G = V^{\dagger}S_r^{(0)}E_m' = G_1. \tag{4.60}$$

Similarly, from (4.53) and (4.54),

$$H = H_1T^{-1} = (E_pS_r^{(0)}W^{\dagger})(W_1W^{\dagger}). \tag{4.61}$$

By proceeding as in the proof of (4.59), one can show

$$S_r^{(0)}W^{\dagger}W_1 = S_r^{(0)}. \tag{4.62}$$

Therefore

$$H = E_pS_r^{(0)}W^{\dagger} = H_1. \tag{4.63}$$

We have shown that the pair (B, C) = (V, W) generates the minimal realization (F, G, H).

3. Suppose that there exists some other pair (B_1, C_1) having the properties of Hypothesis 3 and giving

$$\begin{aligned} F &= B_1{}^{\dagger} S_r^{(1)} C_1{}^{\dagger} \\ G &= B_1{}^{\dagger} S_r^{(0)} E_m{}' \\ H &= E_p S_r^{(0)} C_1{}^{\dagger} . \end{aligned} \tag{4.64}$$

By Lemma 4.8, we must have

$$T = V^{\dagger} B_1 = I. \tag{4.65}$$

$$U = C_1 W^{\dagger} = I. \tag{4.66}$$

By Appendix B.4(iv), $VV^{\dagger}$ is the (unique) orthogonal projector for the column space of $S_r^{(0)}$. The columns of B_1 are in the column space of $S_r^{(0)}$. Therefore (4.65) implies

$$V = VV^{\dagger} B_1 = B_1. \tag{4.67}$$

Similarly, $W^{\dagger} W$ is the orthogonal projector for the row space of $S_r^{(0)}$, and (4.66) implies therefore

$$W = C_1 W^{\dagger} W = C_1. \tag{4.68}$$

The theorem is proved.

<u>Corollary 4.10</u>.

Hypothesis:

$\mathcal{Y}$ is a sequence of matrices having a realization of dimension r.

Conclusion:

There is a one-to-one correspondence between the minimal realizations (F, G, H) for $\mathcal{Y}$, and the matrix pairs $(V, W) = (B, C)$ satisfying the conditions stated in the hypotheses of Theorem 4.9.

The correspondence is established by the following transformations:

$$F = V^{\dagger} S_r^{(1)} W^{\dagger}$$

$$G = V^{\dagger} S_r^{(0)} E_m' \tag{4.69}$$

$$H = E_p S_r^{(0)} W^{\dagger}$$

$$\begin{aligned} V &= [H' \quad F'H' \quad \cdots \quad (F')^{r-1}H']' \\ W &= [G \quad FG \quad \cdots \quad F^{r-1}G]. \end{aligned} \tag{4.70}$$

(The one-to-oneness is assured by the uniqueness of the pseudo inverse of a matrix. See Appendix B, Section 3.)

Proposition 4.11.

Hypotheses:

1. $\mathcal{Y}$ is a sequence of $p \times m$ matrices and satisfies (4.2) for some positive integer r.

2. $S_r^{(k)}$ $(k = 0, 1, 2, \ldots)$ are generalized Hankel matrices for $\mathcal{Y}$, as defined in (4.3).

3. rank $S_r^{(0)} = n$.

$\mathcal{K}$ = column space of $S_r^{(0)}$; i.e. an n-dimensional subspace of R^{pr}.

$\mathcal{R}$ = row space of $S_r^{(0)}$; i.e. an n-dimensional subspace of R^{mr}.

$\mathcal{K}^{\perp}$ = orthogonal complement of $\mathcal{K}$.

$\mathcal{R}^{\perp}$ = orthogonal complement of $\mathcal{R}$.

Conclusions:

1. Suppose (P, Q) are a pair of matrices satisfying the hypotheses of Theorem 4.4, and give a minimal realization (F, G, H) for $\mathcal{Y}$, through substitution in the algorithm (4.13). Then the pair (B, C),

$$B = (E_{n,pr}P)^{\dagger} \tag{4.71}$$

$$C = (QE'_{n,mr})^{\dagger} \tag{4.72}$$

substituted in the algorithm (4.45) of Theorem 4.9, will produce the same realization (F, G, H).

2. Suppose (B, C) are a pair of matrices satisfying the hypotheses of Theorem 4.9, and give a minimal realization (F, G, H) for $\mathcal{Y}$, through substitution in the algorithm (4.45). Then the pair (P, Q),

$$P = \begin{bmatrix} B^{\dagger} \\ B^{-} \end{bmatrix}, \qquad Q = [C^{\dagger} \; C^{-}] \tag{4.73}$$

substituted in Ho's algorithm (4.13), will produce the same realization (F, G, H).

The submatrices B^{-} and C^{-} may be chosen arbitrarily, subject only to the restriction that the rows of B^{-} must be a basis for $\mathcal{K}^{\perp}$ and the columns of C^{-} must be a basis for $\mathcal{R}^{\perp}$.

Proof:

1. By Ho's algorithm, Theorem 4.4, equations (4.13):

$$\begin{aligned} F &= E_{n,pr}PS_r^{(1)}QE'_{n,mr} \\ &= B^{\dagger}S_r^{(1)}C^{\dagger} \end{aligned} \tag{4.74}$$

since $B^{\dagger} = [(E_{n,pr}P)^{\dagger}]^{\dagger} = E_{n,pr}P$, by (4.71), and

$$C^{\dagger} = QE'_{n,mr}, \text{ by (4.72).}$$

Likewise, $G = E_{n,pr}PS_r^{(0)}E'_m$. Therefore, by (4.71),

$$G = B^{\dagger}S_r^{(0)}E'_m. \tag{4.75}$$

Also, $H = E_pS_r^{(0)}QE'_{n,mr}$. Therefore, by (4.72),

$$H = E_pS_r^{(0)}C^{\dagger}. \tag{4.76}$$

Equations (4.74), (4.75), (4.76) are the same as equations (4.45) of Theorem 4.9, thus proving claim 1.

2. By Appendix B, Section 4(vii), the n rows of $B^{\dagger}$ span $\mathcal{K}$. By hypothesis, the $(pr - n)$ rows of B^- span $\mathcal{K}^{\perp}$. Therefore, $P = \begin{bmatrix} B^{\dagger} \\ B^- \end{bmatrix}$ is a nonsingular $(pr \times pr)$ matrix. Similarly, $Q = [C^{\dagger} \; C^-]$ is a nonsingular $(mr \times mr)$ matrix.

Furthermore, by hypothesis, the rows of B^- are orthogonal to the columns of $S_r^{(0)}$, and the columns of C^- are orthogonal to the rows of $S_r^{(0)}$. Therefore

$$B^-S_r^{(0)} = 0 \quad \text{and} \quad S_r^{(0)}C^- = 0. \tag{4.77}$$

Also, $B^{\dagger}S_r^{(0)}C^{\dagger} = [(B'B)^{-1}B'](BC)[C'(CC')^{-1}] = I_n$. (4.78)

Now
$$PS_r^{(0)}Q = \begin{bmatrix} B^\dagger \\ B^- \end{bmatrix} S_r^{(0)} [C^\dagger \; C^-]$$

$$= \begin{bmatrix} B^\dagger S_r^{(0)} C^\dagger & B^\dagger S_r^{(0)} C^- \\ B^- S_r^{(0)} C^\dagger & B^- S_r^{(0)} C^- \end{bmatrix} .$$

Substituting from (4.77) and (4.78), we see that

$$PS_r^{(0)}Q = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix} . \tag{4.79}$$

Thus, the pair (P, Q) satisfies the hypotheses of Theorem 4.4. In Ho's algorithm (4.13), P and Q occur always in combination with the factors $E_{n,pr}$ and $E'_{n,mr}$. By (4.73),

$$\begin{aligned} E_{n,pr}P &= E_{n,pr} \begin{bmatrix} B^\dagger \\ B^- \end{bmatrix} = B^\dagger \\ QE'_{n,mr} &= [C^\dagger \; C^-]E'_{n,mr} = C^\dagger . \end{aligned} \tag{4.80}$$

On substituting the factors (4.80) into Ho's algorithm (4.13), the identity of the resulting equations with those of (4.45) is immediately evident.

The proof is complete.

V. HANKEL DETERMINANTS OF REALIZABLE SCALAR SEQUENCES.

Using the tools of the preceding chapters, certain theoretical properties of realizable sequences and of their realizations can be easily proved. The work reported in this chapter was to provide a theoretical basis for linking the properties of realizations of a given sequence, with the properties of Padé approximants of the formal power series generated by the same sequence.

Theorem 5.4 and its corollaries show some interesting properties of symmetric matrices. Theorem 5.8 gives four different mathematical equivalents of the statement that a scalar sequence has a minimal realization of a given dimension. The theorems mentioned above are generalizations or sharper forms of theorems found in the referenced literature. In view of the great wealth of literature which is concerned with the same or related problems, it is unlikely that the results presented in this chapter are genuinely new, although the proofs are our own (except where references are given).

Proposition 5.1.

Hypotheses:

1. $\mathcal{Y} = (y_0, y_1, \ldots)$ is a sequence of real scalars having a realization; i.e. $\mathcal{Y}$ satisfies a linear recursion (4.2) for some positive integer r.

2. For each nonnegative integer k, $\rho = \rho_r^{(k)}$ is the smallest integer such that either the first $\rho + 1$ columns of $S_r^{(k)}$ are linearly dependent or $\rho = r$.

Conclusion:

$$\Delta_\rho^{(k)} \neq 0. \tag{5.1}$$

<u>Proof</u>. Let k be a fixed nonnegative integer.

By (4.2), the $(r + 1)$th column of $S_n^{(k)}$, $n > r$, is a linear combination of the first r columns.

By hypothesis 2, we have $\Delta_r^{(k)} = 0 \Longrightarrow \rho < r$, so $\rho = r \Longrightarrow \Delta_r^{(k)} \neq 0$. Therefore, the conclusion is trivial if $\rho = r$.

The result is also trivial for $\rho = 0$, since by definition, $\Delta_0^{(k)} = 1 \neq 0$.

Now suppose $\rho = r - d$, $0 < d < r$. By hypothesis 2,

$$\Delta_{\rho+1}^{(k)} = \Delta_{\rho+1}^{(k+1)} = \dots = \Delta_{\rho+1}^{(k+d-1)} = 0, \tag{5.2}$$

while at least one of the following determinants is nonzero:

$$\Delta_\rho^{(k)}, \Delta_\rho^{(k+1)}, \dots, \Delta_\rho^{(k+d)}. \tag{5.3}$$

By Lemma 3.1, for $0 \leq c \leq d$,

$$[\Delta_\rho^{(k+c)}]^2 = \Delta_\rho^{(k+c-1)}\Delta_\rho^{(k+c+1)} - \Delta_{\rho+1}^{(k+c-1)}\Delta_{\rho-1}^{(k+c+1)}. \tag{5.4}$$

Suppose $\Delta_\rho^{(k)} = 0$. Then successive substitution of $c = 1, 2, \dots, d$ in equation (5.4) and use of (5.2) gives

$$\Delta_\rho^{(k+c)} = 0 \qquad (c = 1, 2, \dots, d).$$

Therefore all the determinants (5.3) vanish, contrary to the implication of hypothesis 2. We conclude that $\Delta_\rho^{(k)} \neq 0$.

Theorem 5.2.

Hypotheses:

1. $\mathcal{Y} = (y_0, y_1, \ldots)$ is a sequence of real scalars having a realization; i.e., $\mathcal{Y}$ satisfies a linear recursion (4.2) for some positive integer r.

2. For each nonnegative integer k, $\rho = \rho_r^{(k)}$ is the smallest integer such that either the first $\rho + 1$ columns of $S_r^{(k)}$ are linearly dependent or $\rho = r$.

Conclusion:

$$\operatorname{rank} S_r^{(k)} = \rho_r^{(k)}, \quad \text{all } k \geq 0.$$

Proof: We will carry out the proof for $k = 0$. For other values of k, the subscripts of the y-element in each matrix $S_n^{(k)}$ are increased by the value of k, but the method of proof remains the same.

By hypothesis, the first $\rho = \rho_r^{(0)}$ columns of $S_r^{(0)}$ are linearly independent, and thus span a ρ-dimensional subspace, say K, of the r-dimensional Euclidean vector space R^r. Also by hypothesis, the $(\rho + 1)$th column of $S_r^{(0)}$ is in the subspace K. We will show by mathematical induction that all the columns of $S_r^{(0)}$ lie in the subspace K. Then the column rank of $S_r^{(0)}$ equals the dimension ρ of K, and the conclusion of the theorem follows.

Column No.

		1	2	3	...	ρ	ρ+1		j+1		r
Row No.	1	0	1	2	...	ρ-1	ρ	...	j	...	r-1
	2	1	2	3	...	ρ	ρ+1	...	j+1	...	r
	.	.	.	.	...	.	.	...	.	...	.
	r	r+1	r	r+1	...	r+ρ-2	r+ρ-1	...	r+j-1	...	2r-2
	r+1	r	r+1	r+2	...	r+ρ-1	r+ρ	...	r+j	...	2r-1

Table 5.1: Indices of elements in $S_r^{(0)}$ and in the row bordering $S_r^{(0)}$

Induction Hypothesis: Fix j, $0 < j$; suppose the jth column in $S_r^{(0)}$ is in K (i.e. the first ρ columns of $S_r^{(0)}$ and the jth column are linearly dependent).

Conclusion: The $(j + 1)$th column in $S_r^{(0)}$ is also in K.

Proof:

The result is trivial for $0 < j \leq \rho$.

Suppose $j > \rho$. By realizability of the given sequence $\mathcal{Y}$, the $(r + 1)$th row of y-elements bordering $S_r^{(0)}$ is a linear combination of the r rows of $S_r^{(0)}$. In other words, there exist r numbers β_i $(i = 1, \ldots, r)$ such that

$$\begin{bmatrix} y_r \\ y_{r+1} \\ \vdots \\ y_{r+j} \end{bmatrix}' = \sum_{i=1}^{r} \beta_i \begin{bmatrix} y_{i-1} \\ y_i \\ \vdots \\ y_{i+j-1} \end{bmatrix}' \qquad (5.5)$$

where the prime ($'$) denotes the transpose.

By the induction hypothesis, the jth column is in K, so that there exist (unique) numbers α_k $(k = 1, 2, \ldots, \rho)$ such that

$$\begin{bmatrix} y_{j-1} \\ y_j \\ \cdot \\ \cdot \\ \cdot \\ y_{r+j-2} \end{bmatrix} = \sum_{k=1}^{\rho} \alpha_k \begin{bmatrix} y_{k-1} \\ y_k \\ \cdot \\ \cdot \\ \cdot \\ y_{r+k-2} \end{bmatrix} \tag{5.6}$$

Now

$$\begin{aligned} y_{r+j-1} &= \sum_{i=1}^{r} \beta_i y_{k+j-2} \\ &= \sum_{i=1}^{r} \beta_i \sum_{k=1}^{\rho} \alpha_k y_{i+k-2} \\ &= \sum_{k=1}^{\rho} \alpha_k \sum_{i=1}^{r} \beta_i y_{i+k-2} \\ &= \sum_{k=1}^{\rho} \alpha_k y_{r+k-1} \end{aligned} \tag{5.7}$$

(5.7), together with (5.6), gives the desired result:

$$\begin{bmatrix} y_j \\ y_{j+1} \\ \cdot \\ \cdot \\ \cdot \\ y_{r+j-1} \end{bmatrix} = \sum_{k=1}^{\rho} \alpha_k \begin{bmatrix} y_k \\ y_{k+1} \\ \cdot \\ \cdot \\ \cdot \\ y_{r+k-1} \end{bmatrix} \tag{5.8}$$

i.e. the $(j + 1)$th column of $S_r^{(0)}$ is in the subspace K spanned by the first ρ columns.

Since the induction hypothesis is true for $j = \rho + 1$ (definition of ρ), it holds also for $j = \rho + 2, \rho + 3, \ldots$.

Lemma 5.3.

Hypotheses:

1. $\mathcal{Y} = (y_0, y_1, \ldots)$ is a sequence of real scalars having a realization; i.e. $\mathcal{Y}$ satisfies a linear recursion (4.2) for some positive integer r.

2. For each nonnegative integer k, $\rho = \rho_r^{(k)}$ is the smallest integer such that either the first $\rho + 1$ columns of $S_r^{(k)}$ are linearly dependent or $\rho = r$.

Conclusion: rank $S_n^{(k)} = \rho_r^{(k)}$, for all $n \geq \rho_r^{(k)}$, $k \geq 0$.

Proof:

For $n = \rho_r^{(k)}$, the result follows from Proposition 5.1.

For $n \geq r$, we merely replace r by n in the linear recurrence relation for $\mathcal{Y}$. The new coefficients thus arising in the recurrence relation are set equal to zero. Then Theorem 5.2 gives the desired result.

For $\rho_r^{(k)} < n < r$, the column space of $S_n^{(k)}$ has dimension at least equal to $\rho_r^{(k)}$, by Proposition 5.1. But its dimension cannot be greater than the dimensionality of the column space of $S_r^{(k)}$, which is $\rho_r^{(k)}$, by Theorem 5.2. Therefore, we again have the desired result, and the proof is complete.

Remarks: The above lemma generalizes a theorem by Ho [6, p. 28, Theorem 2.12], who proved it for the special case $k = 0$. Ho's theorem can be strengthened in another direction, as shown in the following theorem and its corollary.

<u>Theorem 5.4.</u>

Hypotheses:

1. $\mathcal{Y} = (y_0, y_1, \ldots)$ is a sequence of real scalars having a realization; i.e. $\mathcal{Y}$ satisfies a linear recursion (4.2) for some positive integer r.

2. For each nonnegative integer k, $\rho_r^{(k)}$ is the smallest integer such that either the first $1 + \rho_r^{(k)}$ columns of $S_r^{(k)}$ are linearly dependent or $\rho_r^{(k)} = r$.

3. rank $S_r^{(0)} = \rho$.

4. The $(\rho + 1)$th column of $S_n^{(0)}$, $n > \rho$, is the (unique) linear combination

$$\begin{bmatrix} y_\rho \\ y_{\rho+1} \\ \cdot \\ \cdot \\ \cdot \\ y_{\rho+n-1} \end{bmatrix} = \sum_{i=1}^{\sigma} \alpha_i \begin{bmatrix} y_{\rho-i} \\ y_{\rho-i+1} \\ \cdot \\ \cdot \\ \cdot \\ y_{\rho-i+n-1} \end{bmatrix} \tag{5.9}$$

where $$\sigma = \max \{0, i: \alpha_i \neq 0\}. \tag{5.10}$$

Conclusion:

$$\text{rank } S_n^{(k)} = \max\{\rho - k, \sigma\}, \quad \text{all } k \geq 0, \ n \geq \rho.$$

The significance of Hypothesis 4 is that it uniquely specifies σ. To show that the hypotheses are consistent, fix $n > \rho$. Independent columns in $S_\rho^{(0)}$ are also independent in $S_n^{(0)}$. By Lemma 5.3, rank $S_n^{(0)} = \rho$. Therefore the $(\rho + 1)$th column of $S_n^{(0)}$ must be a unique linear combination of the first ρ columns; that is, there exist unique numbers α_i $(i = 1, 2, \ldots, \rho)$ such that

$$\begin{bmatrix} y_\rho \\ y_{\rho+1} \\ \cdot \\ \cdot \\ \cdot \\ y_{\rho+n-1} \end{bmatrix} = \sum_{i=1}^{\rho} \alpha_i \begin{bmatrix} y_{\rho-i} \\ y_{\rho-i+1} \\ \cdot \\ \cdot \\ \cdot \\ y_{\rho-i+n-1} \end{bmatrix} \tag{5.11}$$

Some (possibly all) of the α_i may be zero. Let $\sigma = 0$ if all of the α_i are zero; otherwise let $\sigma = \max\{i\colon \alpha_i \neq 0\}$. Then the sum in (5.11) can be written as in (5.9), without loss of generality.

Now n in the preceding paragraph was an arbitrary integer greater than ρ. Therefore (5.9) implies the linear recursion relation

$$y_{\rho+j} = \sum_{i=1}^{\sigma} \alpha_i y_{\rho-i+j}, \quad \text{for } j \geq 0. \tag{5.12}$$

Let K_n denote the column space of $S_n^{(0)}$. Then (5.12) implies that the columns of $S_n^{(k)}$ all lie in K_n, for every $k \geq 0$, $n \geq \rho$. Moreover, for every $n \geq \rho$, we have from 5.12.

$$\begin{aligned} \text{rank } S_n^{(k)} &= \rho - k, & k &= 0, 1, \ldots, \rho - \sigma; \\ \text{rank } S_n^{(k)} &= \sigma, & k &> \rho - \sigma. \end{aligned} \tag{5.13}$$

The theorem is proved.

Corollary 5.5 (Generalization of B.L. Ho's Theorem [6, p. 28])

Hypotheses: Same as in Theorem 5.4.

Conclusion: Among all possible Hankel matrices $S_n^{(k)}$ constructed from the sequence $\mathcal{Y}$, the largest nonsingular ones are

(i) $S_\rho^{(0)}$, if $0 \leq \sigma < \rho$;

(ii) $S_\rho^{(k)}$, $k \geq 0$, if $\sigma = \rho$.

Corollary 5.6.

Hypotheses: Same as in Theorem 5.4.

Conclusion: The power series

$$f(z) = \sum_{i=0}^{\infty} y_i z^i \tag{5.14}$$

is a rational function with denominator of degree not more than ρ. Furthermore, if $\sigma = \rho$, then

$$f(z) = \frac{b_0 + b_1 z + \ldots + b_{\rho-1} z^{\rho-1}}{c_0 + c_1 z + \ldots + c_\rho z^\rho}, \quad c_0 \neq 0, \; c_\rho \neq 0.$$

Proof:

(i) Suppose $0 \leq \sigma < \rho$. Then, by Corollary 5.5,

$$\Delta_\rho^{(0)} \neq 0, \quad \Delta_\rho^{(k)} = 0 \quad \text{for} \quad k > 0.$$

Substitute $q = \rho - 1$ in Dienes' version of Lemma 3.2. Then $f(z)$ is seen to be a rational function with denominator of degree at most $\rho - 1$.

(ii) Suppose $\sigma = \rho$. Then, by Corollary 5.5, $\Delta_\rho^{(k)} \neq 0$, $\Delta_{\rho+1}^{(k)} = 0$, for $k \geq 0$. Substitute $p = \rho - 1$, $q = \rho$ in the Bieberbach version of Lemma 3.2. Then the desired result follows immediately from the lemma.

Lemma 5.7. [27, pp. 302-305][28, pp. 1010, 1011]

Hypothesis:

F is a constant $n \times n$ matrix.

Conclusions:

1.
$$(sI - F)^{-1} = \frac{B(s)}{d(s)} \tag{5.15}$$

where
$$d(s) = \det (sI - F) = s^n + d_1 s^{n-1} + \ldots + d_n,$$
$$\tag{5.16}$$
and
$$B(s) = B_0 s^{n-1} + B_1 s^{n-2} + \ldots + B_{n-1}.$$

2. The coefficients d_i of the polynomial $d(s)$ and the matrix coefficients B_j of the matrix polynomial $B(s)$ are given by the recursion formulas

$$\begin{aligned} d_k &= -\frac{1}{k}\, t_r(B_{k-1}F), \quad \text{for } k = 1, 2, \ldots, n; \\ B_k &= B_{k-1}F + d_k I, \quad \text{for } k = 1, 2, \ldots, n-1; \\ B_0 &= I. \end{aligned} \tag{5.17}$$

3.
$$B_{n-1}F + d_n I = 0. \tag{5.18}$$

For a proof of the lemma, see Desoer [27], pp. 302-305.

Comment:

The matrix rational function $\Phi(s) = (sI - F)^{-1}$ is called the resolvent of F. [29, p. 52] $\Phi(s)$ is also the Laplace transform of the state transition matrix, $\Phi(t) = \exp (Ft)$.

It may happen that all the n^2 elements of the matrix polynomial $B(s)$ have one or more factors in common with $d(s)$. Cancellation of all such common factors leads to a simplified expression, namely,

$$\Phi(s) = \frac{P(s)}{m(s)}$$

where the polynomials $m(s)$ and $P(s)$ are the result of these cancellations. Then $m(s)$ is the minimal polynomial of the matrix F, i.e. the (monic) polynomial of least degree such that $m(F) = 0$; also, every eigenvalue of F is a pole of $(sI - F)^{-1}$, i.e. a zero of $m(s)$.

For the purpose of computing $\Phi(s)$, (5.17) is a more efficient procedure than Cramer's rule, the latter requiring nearly $(n - 1)!$ times as many multiplications as (5.17). [27, pp. 302, 306]

Theorem 5.8.

Hypothesis:

$\mathcal{Y} = (y_0, y_i, \ldots)$ is a sequence of real scalars.

Conclusion:

The following five statements are equivalent:

(i) $\mathcal{Y}$ has a minimal realization of dimension ρ.

(ii) $\mathcal{Y}$ satisfies a linear recursion of the form

$$y_{\rho+j} = \sum_{i=1}^{\rho} \alpha_i y_{\rho-i+j}, \quad j = 0, 1, \ldots \tag{5.19}$$

where ρ is the smallest positive integer for which (5.19) is true.

(iii) The function $f(z) = \sum_{k=0}^{\infty} y_k z^k$ is rational of the form

$$f(z) = \frac{b_0 + b_1 z + \ldots + b_{\rho-1} z^{\rho-1}}{c_0 + c_1 z + \ldots + c_\rho z^\rho}, \quad c_0 \neq 0, \quad c_\rho \neq 0.$$

(iv) $\Delta_{\rho+1}^{(k)} = 0, \quad k = 0, 1, \ldots$ (5.20)

For each $N \geq 0$, there is $n = n(N) \geq N$ such that

$$\Delta_\rho^{(n)} \neq 0. \tag{5.21}$$

(v) There exists a positive integer ρ such that

$$\rho = \min \{r: \text{ rank } S_n^{(0)} = r, \text{ all } n \geq r\}. \tag{5.22}$$

Proof:

The proof will be accomplished in two cycles:

First cycle: (ii) $\Longrightarrow$ (i) $\Longrightarrow$ (iii) $\Longrightarrow$ (iv); $\Longrightarrow$ (ii);

Second cycle: (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (v) $\Longrightarrow$ (i).

(ii) $\Longrightarrow$ (i).

By Lemma 4.2, $S_\rho^{(i)} = M^k S_\rho^{(0)}$ (i = 0, 1, 2, ...).

The first element in $S_\rho^{(i)}$ is y_i. Therefore

$$y_i = E_1 S_\rho^{(i)} E_1' = E_1 M^i S_\rho^{(0)} E_1' \qquad (5.23)$$

and

$$\begin{aligned} F &= M \\ G &= S_\rho^{(0)} E_1' \\ H &= E_1 \end{aligned} \qquad (5.24)$$

is a realization of $\mathcal{Y}$, of dimension ρ.

By Lemma 4.3, the dimension of any minimal realization of $\mathcal{Y}$ is equal to the rank of $S_\rho^{(0)}$.

It remains to show that rank $S_\rho^{(0)} = \rho$, i.e. $\Delta_\rho^{(0)} \neq 0$. Since ρ is the smallest integer for which (5.19) is true, then for each integer σ, $1 \leq \sigma \leq \rho$, there is an integer $r(\sigma)$ such that the σth column of $S_{r(\sigma)}^{(0)}$ is linearly independent of the preceding $\sigma - 1$ columns. Let $r = \max_{1 \leq \sigma \leq \rho} \{r(\sigma), \rho\}$. Then ρ is the smallest integer such that the first $\rho + 1$ columns of $S_r^{(0)}$ are linearly dependent or $\rho = r$. By Proposition 5.1, we have $\Delta_\rho^{(0)} \neq 0$. This completes the proof of (ii) $\Longrightarrow$ (i).

(i) $\Longrightarrow$ (iii).

Let (F, G, H) be a minimal realization of $\mathcal{Y}$. The transfer function of the system (F, G, H) is

$$Z(s) = H(sI - F)^{-1} G \qquad (5.25)$$

where I is the ρ-dimensional unit matrix. We can write

$$Z(s) = \frac{1}{s} H\left(I - \frac{1}{s} F\right)^{-1} G$$

and, for sufficiently large s, express $(I - \frac{1}{s} F)^{-1}$ as a geometric series. The result is

$$Z(s) = \frac{1}{s} H \sum_{k=0}^{\infty} (\frac{1}{s} F)^k G = \sum_{k=0}^{\infty} y_k s^{-k-1}. \tag{5.26}$$

It is convenient to consider functions that are regular at 0 rather than at infinity. We therefore put

$$z = \frac{1}{s}, \qquad sZ(s) = f(z) = \sum_{k=0}^{\infty} y_k z^k. \tag{5.27}$$

By Lemma 5.7, Z is a rational function of s, of denominator degree $q \leq \rho$ and numerator degree $p = q - 1$. It follows that f is a rational function of z, with numerator and denominator degrees the same as Z.

To show that $q = \rho$, we use the following two facts:

(a) (F, G, H) is the minimal realization of a scalar sequence; therefore F is a nonderogatory matrix, i.e. the minimal and the characteristic polynomials of F coincide. [6, p. 47, Corollary 3.7]

(b) Suppose all factors common to the numerator and denominator of $Z(s)$ have been cancelled. Then the denominator is the minimal polynomial of the matrix F. [27, p. 306]

The characteristic polynomial of F, $\Delta(s) = \det [F - sI]$, is of degree ρ. By (a), F is nonderogatory; so, the minimal polynomial is of the same degree ρ. Therefore, (b) implies $q = \rho$.

By Lemma 5.7, equation (5.18), the denominator of f has a nonzero constant term. (Otherwise s would be a cancellable factor.) Thus, f is a rational function,

$$f(z) = \frac{b_0 + b_1 z + \dots + b_{\rho-1} z^{\rho-1}}{c_0 + c_1 z + \dots + c_\rho z^\rho}, \quad c_0 \neq 0, \quad c_\rho \neq 0. \tag{5.28}$$

(iii) $\Longrightarrow$ (iv)

Invoking Lemma 3.2(i), we find that (5.28) implies

$$\Delta_{\rho+1}^{(k)} = 0 \quad \text{for all} \quad k \geq 0. \tag{5.20}$$

Again, by Lemma 3.2(ii), ρ is the least value of q such that $\sum_{k=0}^{\infty} \Delta_{q+1}^{(k)} z^k$ is a polynomial. Therefore, given any nonnegative integer N, there is $n = n(N) \geq N$ such that

$$\Delta_\rho^{(n)} \neq 0. \tag{5.21}$$

(iv) $\Longrightarrow$ (ii).

Condition (5.20) implies that, for all $n \geq \rho + 1$, the first $\rho + 1$ columns of $S_n^{(0)}$ are linearly dependent. Condition (5.21), however, implies that for sufficiently large n, the first ρ columns of $S_n^{(0)}$ are linearly independent. Therefore, $q = \rho$ is the smallest integer such that the first $q + 1$ columns of $S_n^{(0)}$ are linearly dependent for all $n \geq 0$. This, in turn, implies that $q = \rho$ is the smallest integer such that for all $n \geq 0$, the $(q + 1)$th column of $S_n^{(0)}$ is a linear combination of the preceding q columns. The statement (ii) is an immediate consequence.

(i) $\Longrightarrow$ (ii) $\Longrightarrow$ (v) $\Longrightarrow$ (i). [6; p. 25]

If a minimal realization for $\mathcal{Y}$ exists and has dimension ρ, then (5.19) is true. This implies rank $S_n^{(0)}$ = rank $S_\rho^{(0)} = \rho$, all $n \geq \rho$. Since rank $S_{\rho-1}^{(0)} \leq \rho - 1$, we have shown that (ii) $\Longrightarrow$ (v).

Conversely, suppose (5.22) is true for some positive integer ρ. We appeal to the following lemma proved in Ho's dissertation [6; p. 20, Lemma 2.7]:

> "Suppose that rank $S_r^{(0)}$ = rank $S_{r+1}^{(0)}$ for some integer r. Applying the algorithm (4.13) to the matrix $S_r^{(0)}$ produces (F, G, H) such that the fundamental relation $HF^kG = Y_k$ is satisfied for every element of $S_{r+1}^{(0)}$, that is, for $k = 0, 1, 2, \ldots, 2r$."

Now (5.22) implies rank $S_n^{(0)} = \rho$, all $n \geq \rho$. Therefore applying the algorithm (4.13) to the matrix $S_\rho^{(0)}$ produces (F, G, H) such that $HF^kG = y_k$, $k = 0, 1, 2, \ldots$. Therefore $\mathcal{Y}$ has a realization (F, G, H) and the realization is minimal of dimension ρ, by Theorem 4.4.

The proof of the theorem is complete.

VI. PARTIAL REALIZATION OF SCALAR SEQUENCES.

The last theorem of Chapter 5 gave four distinct sets of conditions, each set being necessary and sufficient for a given scalar sequence to be realizable in the strict sense used by B. L. Ho. We now turn our attention to the following two problem areas: One is the approximate realization of sequences which do not meet the realizability criteria of Theorem 5.8. The other problem concerns the partial, approximate realization of sequences which are known to be realizable in the strict sense.

The two problems can be treated as one, mathematically. This becomes evident from the following definitions.

6.1 Definitions.

Suppose $\mathcal{Y} = (y_0, y_1, \ldots)$ is a sequence of real numbers.

The sequence $\mathcal{Y}$ is called <u>ρ-realizable</u> if it is realizable and if ρ is the smallest positive integer for which the recursion formula

$$y_{\rho+j} = \sum_{i=1}^{\rho} \alpha_i y_{\rho-i+j}, \quad j = 0, 1, \ldots \tag{6.1}$$

is true. If $\mathcal{Y}$ is not realizable, we set $\rho = \infty$.

Suppose r is a positive integer, and (F_r, G_r, H_r) is an ordered triple of matrices computed from Ho's minimal realization algorithm using the matrices $S_r^{(0)}$ and $S_r^{(1)}$ associated with $\mathcal{Y}$. Then (F_r, G_r, H_r) is called a <u>linear model of order r</u> for $\mathcal{Y}$.

For the sake of clarity, the phrase <u>realizable in the strict sense</u> will be used if a sequence $\mathcal{Y}$ is ρ-realizable, $\rho < \infty$.

A linear model of order r is called a partial realization (of order r) for $\mathcal{Y}$ if $\mathcal{Y}$ is ρ-realizable, $\rho > r$, or if $\mathcal{Y}$ is not realizable in the strict sense.

Remarks.

If $\mathcal{Y}$ is ρ-realizable, $\rho < \infty$, then any linear model of order ρ (or greater than ρ) is a minimal realization, while any linear model of order less than ρ is a partial realization. This follows immediately from the definitions and the theorems of Chapter 4, especially Theorem 4.9 and its corollary.

Suppose (F_r, G_r, H_r) is a partial realization for $\mathcal{Y}$, and that the elements x_k of the sequence $\mathcal{X} = (x_0, x_1, \ldots)$ are given by

$$x_k = H_r F_r^{\,k} G_r, \quad k = 0, 1, \ldots . \tag{6.2}$$

The sequence $\mathcal{X}$ is uniquely determined by $\mathcal{Y}$, r, and by the choice of the submatrices $B^{\dagger}$ and $C^{\dagger}$ of P and Q, respectively. (See conclusion 2 of Proposition 4.11.) The map $\mathcal{Y} \to \mathcal{X}$ and its approximating properties thus may possibly depend on the particular choice of the partial realization. The related questions can be usefully studied as projection problems. [37]

6.2 Elementary Properties of Partial Realizations.

Proposition 6.1.

Hypotheses:

1. $\mathcal{Y}$ is a ρ-realizable sequence, $\rho < \infty$.

2. (F_r, G_r, H_r) is a linear model of order r for $\mathcal{Y}$.

Conclusion:

1. (F_r, G_r, H_r) is a minimal realization for $\mathcal{Y}$ if, and only if, $r \geq \rho$.

2. $\dim F_r = \rho$ if $r \geq \rho$; $\dim F_r \leq r$ if $r < \rho$.

Proof:

Since $\mathcal{Y}$ is ρ-realizable, $\rho < \infty$, a linear recursion of the form (6.1) holds for $\mathcal{Y}$, with upper limit ρ. By Proposition 4.1, $\mathcal{Y}$ has a realization.

By definition, ρ is the smallest integer for which the recursion (6.1) is true. Therefore, Theorem 5.8 implies that every minimal realization for $\mathcal{Y}$ has dimension ρ. By Theorem 4.4, $\dim F_r = \operatorname{rank} S_r^{(0)}$.

(i) Suppose $r \geq \rho$.

Then, $\operatorname{rank} S_r^{(0)} = \rho$ [6; p. 25, Corollary 2.9], and, by Theorem 4.4, (F_r, G_r, H_r) is a minimal realization for $\mathcal{Y}$, $\dim F_r = \rho$.

(ii) Suppose $r < \rho$.

Clearly, $\operatorname{rank} S_r^{(0)} \leq r$, since the elements of $\mathcal{Y}$ are scalars. But $\dim F_r = \operatorname{rank} S_r^{(0)}$, by Theorem 4.4. Therefore, $\dim F_r \leq r < \rho$. Every minimal realization for $\mathcal{Y}$ has dimension ρ. Therefore (F_r, G_r, H_r) is not a minimal realization.

The proof is complete.

Corollary 6.2.

Partial realization for a given scalar sequence $\mathcal{Y}$ are not realizations in the strict sense.

Theorem 6.3.

Hypotheses:

1. $\mathcal{Y} = (y_0, y_1, \ldots)$ is a sequence of real numbers.

2. r is a positive integer such that $\Delta_r^{(0)} \neq 0$.

3. (F, G, H) is a linear model of order r for $\mathcal{Y}$.

Conclusion:

$HF^kG = y_k, \quad k = 0, 1, \ldots, 2r - 1.$

Proof:

By hypothesis 2, $S = S_r^{(0)}$ has an inverse S^{-1}. Define the $(r \times 1)$ vector α by

$$\alpha = \begin{bmatrix} \alpha_r \\ \alpha_{r-1} \\ \vdots \\ \alpha_1 \end{bmatrix} = S^{-1} \begin{bmatrix} y_r \\ y_{r+1} \\ \vdots \\ y_{2r-1} \end{bmatrix} \tag{6.3}$$

Then $S_r^{(0)}\alpha = \begin{bmatrix} y_r \\ y_{r+1} \\ \vdots \\ y_{2r-1} \end{bmatrix}$, which means that

$$\sum_{i=1}^{r} \alpha_i y_{k-i} = y_k, \quad k = r, r+1, \ldots, 2r-1. \tag{6.4}$$

Define a sequence $\mathscr{X} = (x_0, x_1, \ldots)$ by

$$x_k = y_k, \quad k = 0, 1, \ldots, r - 1 \tag{6.5}$$

$$x_k = \sum_{i=1}^{r} \alpha_i x_{k-i}, \quad k = r, r + 1, \ldots \tag{6.6}$$

We observe:

(i) The sequence $\mathscr{X}$ satisfies a linear recurrence relation and therefore has a realization, by Proposition 4.1.

(ii) The first $2r$ terms of the $\mathscr{X}$-sequence are the same as the first $2r$ terms of the $\mathscr{Y}$-sequence.

(iii) The matrices $S_r^{(0)}$ and $S_r^{(1)}$ are completely determined by the first $2r$ terms of the $\mathscr{Y}$-sequence.

(iv) The coefficients $\{\alpha_i\}$ of the vector α are unique because S is nonsingular.

By Theorem 4.4, (F, G, H) determined from the first $2r$ terms of the $\mathscr{X}$-sequence is a minimal realization for the $\mathscr{X}$-sequence. Because of hypothesis 2, $\dim F = r$. Now

$$HF^kG = x_k, \quad \text{for all} \quad k \geq 0. \tag{6.7}$$

But, as observed in (ii), $x_k = y_k$, $k = 0, 1, \ldots, 2r - 1$. Therefore

$$HF^kG = y_k, \quad \text{for} \quad k = 0, 1, \ldots, 2r - 1, \tag{6.8}$$

where (F, G, H) are determined by $S_r^{(0)}$ and $S_r^{(1)}$.

The proof of the theorem is complete.

Corollary 6.4. (Note: This is a special case of B.L. Ho's Lemma 2.7 [6, p. 20] and [38, Theorem 2].)

Hypotheses:

1. $\mathcal{Y} = (y_0, y_1, \ldots)$ is a sequence of real numbers.

2. r is a positive integer such that $\Delta_r^{(0)} \neq 0$, $\Delta_{r+1}^{(0)} = 0$.

3. (F, G, H) is a linear model of order r for $\mathcal{Y}$.

Conclusion:

$$HF^kG = y_k, \quad k = 0, 1, \ldots, 2r.$$

Proof:

By Theorem 6.3, $HF^kG = y_k$, $k = 0, 1, \ldots, 2r - 1$. (6.8)

$\Delta_r^{(0)} \neq 0$ implies the first r columns of $S_{r+1}^{(0)}$ are linearly independent. Let their span be $\mathcal{K}$. Then $\Delta_{r+1}^{(0)} = 0$ implies that the last column of $S_{r+1}^{(0)}$ is in $\mathcal{K}$. Therefore there exist unique coefficients α_i, $i = 1, 2, \ldots, r$, such that

$$y_{r+j} = \sum_{i=1}^{r} \alpha_i y_{r+j-i}, \quad j = 0, 1, \ldots, r. \tag{6.9}$$

The first r equations of the set (6.9) are identical with (6.4). The coefficients $\{\alpha_i\}$ are uniquely determined by (6.12), as was observed in comment (iv) of Theorem 6.3. Therefore the $\{\alpha_i\}$ in (6.4) and (6.9) must be the same.

Proceeding as in Theorem 6.3, we get a realizable sequence $\mathcal{X} = (x_0, x_1, \ldots)$ and a triple of matrices (F, G, H) such that

$$HF^kG = x_k, \quad \text{all } k \geq 0 \tag{6.10}$$

$$x_k = \sum_{i=1}^{r} \alpha_i x_{k-i}, \quad k = r, r+1, \ldots \tag{6.11}$$

$$x_k = y_k, \quad k = 0, 1, \ldots, r - 1. \tag{6.12}$$

Now (6.8) to (6.12) imply

$$HF^kG = x_k = y_k, \quad \text{for} \quad k = 0, 1, \ldots, 2r. \tag{6.13}$$

The corollary is proved.

6.3 Approximating Properties of Partial Realizations.

Theorem 6.5.

Hypotheses:

1. $\mathcal{Y} = (y_0, y_1, \ldots)$ is a sequence of real numbers.

2. r is a positive integer such that $\Delta_r^{(0)} \neq 0$.

3. (F, G, H) is a linear model of order r for $\mathcal{Y}$.

Conclusion:

The rational function

$$R(z) = H(I - zF)^{-1}G \tag{6.14}$$

is the $(r, r - 1)$ Padé approximant for the power series

$$f(z) = \sum_{k=0}^{\infty} y_k z^k. \tag{6.15}$$

Proof:

The hypotheses are the same as those of Theorem 6.3. As shown in that theorem, (F, G, H) is a minimal realization for the sequence $\mathcal{X} = (x_0, x_1, \ldots)$ defined by (6.4) to (6.6). The dimension of the realization is r, by hypothesis 2.

The transfer function of the system (F, G, H) is

$$\begin{aligned} Z(s) &= H(sI - F)^{-1}G \\ &= \frac{1}{s}H(I - \frac{1}{s}F)^{-1}G. \end{aligned} \tag{6.26}$$

Let $z = \frac{1}{s}$, then

$$\frac{1}{z} Z(\frac{1}{z}) = H(I - zF)^{-1}G = R(z). \tag{6.27}$$

Now $R(z)$ is a rational function of the form

$$R(z) = \frac{b_0 + b_1 z + \ldots + b_{r-1} z^{r-1}}{c_0 + c_1 z + \ldots + c_r z^r}, \quad c_0 \neq 0, \ c_r \neq 0. \tag{6.28}$$

The proof is contained in the proof of Theorem 5.8, equation (5.28). Because of (6.28), $R(z)$ satisfies the defining condition I of the $(r, r - 1)$ Padé approximant for f:

$$\deg B \leq r - 1, \tag{6.29}$$

$$\deg C \leq r; \tag{6.30}$$

where B = numerator polynomial of R,

C = denominator polynomial of R.

It remains to show that

$$f(z)C(z) - B(z) = (z^{2r}). \tag{6.31}$$

For sufficiently small values of z, we have

$$R(z) = H \sum_{k=0}^{\infty} (zF)^k G = \sum_{k=0}^{\infty} x_k z^k \tag{6.32}$$

where x_k, $k = 0, 1, 2, \ldots,$ are the elements of the sequence $\mathfrak{X}$.

By Theorem 6.3,

$$x_k = y_k, \quad k = 0, 1, \ldots, 2r - 1. \tag{6.33}$$

Therefore,

$$f(z) - R(z) = \sum_{k=2r}^{\infty} (y_k - x_k) z^k = (z^{2r}). \qquad (6.34)$$

Multiply both sides of (6.34) by $C(z) = c_0 + c_1 z + \ldots + c_r z^r$. Then (6.31) is obtained, and condition II of the $(r, r-1)$ Padé approximant for f is satisfied by $R(z)$.

The proof of the theorem is complete.

Corollary 6.6.

Hypotheses:

1. $$f(z) = \sum_{k=0}^{\infty} y_k z^k, \quad y_0 \neq 0 \qquad (6.35)$$

is a power series whose Padé table is normal.

2. (F, G, H) is a linear model of order r for the sequence $y = (y_0, y_1, \ldots)$.

Conclusion:

The transfer function of the system (F, G, H) is the element $E_r^{(0)}$ in the E-array for the function

$$T(s) = \sum_{k=0}^{\infty} y_k s^{-k-1}. \qquad (6.36)$$

Proof:

Since f has a normal Padé table, Theorem 3.5 implies that $\Delta_r^{(0)} \neq 0$. Therefore, $H(I - zF)^{-1}G$ is the $(r, r-1)$ Padé approximant for f, as shown in Theorem 6.5.

The transfer function of the system (F, G, H) is

$$Z(s) = H(sI - F)^{-1}G = \sum_{k=0}^{\infty} x_k s^{-k-1} \qquad (6.37)$$

where $x_k = HF^kG$, $k = 0, 1, \ldots$. (6.38)

Let $(P_{r,\, r-1}, Q_{r,r-1})$ be the $(r, r-1)$ Padé pair for f.

By Theorem 6.5,

$$\begin{aligned} P_{r,r-1}(z) &= b_0 + b_1 z + \ldots + b_{r-1} z^{r-1}; \\ Q_{r,r-1}(z) &= c_0 + c_1 z + \ldots + c_r z^r, \quad c_0 \neq 0, \quad c_r \neq 0. \end{aligned} \tag{6.39}$$

By Proposition 2.10 and the definition of the E-array,

$$E_r^{(0)}(T, s) = \frac{P_r^{(0)}(s)}{Q_r^{(0)}(s)} \tag{6.40}$$

where [equations (2.64)]

$$P_r^{(0)}(s) = z^{-r+1} P_{r,r-1}(z), \qquad sz = 1$$

and

$$Q_r^{(0)}(s) = z^{-r} Q_{r,r-1}(z), \qquad sz = 1. \tag{6.41}$$

Now (6.39) and (6.40) give

$$\begin{aligned} E_r^{(0)}(T, s) &= \frac{zP_{r,r-1}(z)}{Q_{r,r-1}(z)}, \qquad sz = 1, \\ &= zR(z), \quad \text{by (6.28)} \\ &= \sum_{k=0}^{\infty} x_k s^{-k-1}, \quad \text{by (6.32)} \\ &= Z(s). \end{aligned} \tag{6.42}$$

The proof of the corollary is complete.

VII. CONCLUSIONS AND FUTURE RESEARCH AREAS.

Starting with B. L. Ho's algorithm for computing linear models from input-output data, we have studied the relation between the realized system matrices (F, G, H) and certain rational approximations related to the formal power series whose coefficients are the Markov parameters. For scalar sequences of Markov parameters, $\mathcal{Y} = (y_0, y_1, \cdots)$, the system matrices (F, G, H) computed by B. L. Ho's algorithm are shown to have the following property: Suppose $\mathcal{Y}$ is a normal sequence (i.e., $\sum_0^\infty y_k z^k$ has a normal Padé table), then the transfer function $H(sI - F)^{-1}G$ lies on the diagonal of the E-array for $\mathcal{Y}$.

Deeper results require research into the properties of E-arrays for nonnormal sequences. With this aim, we have developed an explicit determinantal expression for Padé approximants which is valid for both the normal and the nonnormal case (Theorem 3.7). The extension of this work to the E-array should be straight-forward, even if tedious.

By concentrating attention on the leading terms of a given sequence $\mathcal{Y}$, the Padé approximation emphasizes the high-frequency response of a linear model for $\mathcal{Y}$. This approach, while desirable for many applications and interesting from a theoretical standpoint, has certain limitations. For instance, in modeling a linear system from noisy input-output data, one would want to emphasize a pass-band rather than the high end of frequency-response spectrum. This important problem is therefore likely to require some modification of the methods of the Padé approach. A possible alternative to be considered is the uniform (Chebychev) approximation.

Besides confining attention to the Padé approach, our research on the approximating properties of linear models has been restricted to single input-single output systems. The restriction allowed the important issues and steps to stand out in the investigation. Extension of the results to multivariate systems may be possible, at the cost of increased complexity in the derivations.

As a byproduct of our study, we derived reciprocal relations between a minimal realization (F, G, H) and the corresponding pair of matrix factors (V, W). The result, found in the Unique Representation Theorem and its corollary, is a refinement of Ho's algorithm. The intrinsic elegance of the relations presented in the corollary is accompanied by computational advantages compared with the earlier formulation of the algorithm by B. L. Ho. Future research may be able to exploit the one-to-one correspondence established here for the first time, and uncover its deeper theoretical significance.

Appendix A. Some Definitions in Algebra

1. Formal Power Series [9, p. 146]

Let X be a letter and let N be the set of nonnegative integers (i.e. the natural numbers). Let G be the monoid of functions from the set $\{X\}$ to N.

If $k \in N$, let X^k denote the function in G whose value at X is k. Then $G = (X^0, X^1, X^2, \ldots, X^k, \ldots)$, and X^k is a monomial whose index v is called its degree. As a matter of notation, "degree" is often abbreviated "deg".

Let R be a commutative ring, and let $R[[X]]$ be the set of functions from G into R, without any restriction. Then an element of $R[[X]]$ may be viewed as assigning to each monomial X^k a coefficient $a_k \in R$. We denote this element by

$$\sum_{k=0}^{\infty} a_k X^k.$$

The summation symbol here is not a sum, but the expression is also written in the form

$$a_0 X^0 + a_1 X^1 + \ldots$$

and is called a formal power series in one variable, with coefficients $a_0, a_1, \ldots$ in R.

Addition and multiplication of two elements in $R[[X]]$, say

$$f = \sum_{k=0}^{\infty} a_k X^k \quad \text{and} \quad g = \sum_{k=0}^{\infty} b_k X^k,$$

are defined as follows:

$$f + g = \sum_{k=0}^{\infty} (a_k + b_k)X^k$$

$$fg = \sum_{k=0}^{\infty} c_k X^k$$

where $c_k = \sum_{u+v=k} a_u b_v$. (Note: With these definitions of addition and multiplication, the set $R[[X]]$ becomes a commutative ring.)

Let $f = \sum_{k=0}^{\infty} a_k X^k$ be a nonzero power series. The smallest index k for which $a_k \neq 0$ is called the <u>order</u> of f, denoted by $\sigma(f)$. The zero element of $R[[X]]$ is said to be of order $+\infty$. [19, p. 129]

2. <u>Polynomials</u> [9, p. 118]

<u>Polynomials</u> in one variable with coefficients in R can be identified with formal power series as follows:

If $f \in R[X]$ and $f = a_0X^0 + a_1X^1 + \ldots + a_mX^m$, then we identify f with the power series $\sum_{k=0}^{\infty} a_k X^k$, where $a_k = 0$, $\forall\, k > m$. Thus, the polynomials in one variable in $R[X]$ are identified with the subset of functions $G \to R$ in $R[[X]]$ which are zero for almost all elements of G. [9, p. 110]

The <u>degree</u> of f, denoted by deg f, is the largest index k for which $a_k \neq 0$. The zero polynomial is said to be of degree $-\infty$. If $\deg f = m$, then $a_m \neq 0$ by definition, and a_m is the <u>leading coefficient</u> of f. A <u>monic</u> polynomial has leading coefficient equal to unity.

If $f, g \in R[X]$, then we have:

$$\deg (f, g) \leq \max(\deg f, \deg g).$$

Also $\deg(fg) = \deg f + \deg g$ if R is an integral domain, provided at least one of the leading coefficients of f, g is not a divisor of zero.

If $f, g \in R[[X]]$, then [Zariski and Samuel, II p. 129]

$$\sigma(f, g) \geq \min(\sigma(f), \sigma(g)).$$

Also

$$\sigma(fg) = \sigma(f) + \sigma(g)$$

if R is an integral domain.

3. Rational Functions [9, p. 116]

If K is the quotient field of an integral domain R, the quotient field of $R[X]$ is denoted by $K(X)$. An element of $K(X)$ is called a rational function. A rational function can be written as a quotient $f(X)/g(X)$ where f, g are polynomials.

Two nonzero polynomials f, g are called relatively prime if f and g have no common factors other than constants. If f and g are relatively prime, the rational function $f(X)/g(X)$ is sometimes called "irreducible". [2, p. 153], [10, p. 106], [17, p. 398].

Appendix B. The Pseudo Inverse of a Matrix

1. Definition [7, p. 197]

Let A be an arbitrary (finite) matrix. A matrix $A^\dagger$ is called the pseudo inverse of A if the following hold:

$$AA^\dagger A = A, \qquad A^\dagger AA^\dagger = A^\dagger \tag{B.1}$$

$$(A^\dagger A)' = A^\dagger A, \qquad (AA^\dagger)' = AA^\dagger \tag{B.2}$$

For an alternative, equivalent definition, see Moore's theorem in Section 3.

2. Construction [31, p.9]

Let A be an arbitrary $p \times m$ matrix, rank $A = n$. Suppose B and C are matrices with the following properties:

(i) B is a $p \times n$ matrix whose n columns are a basis for the column space of A.

(ii) C is a $n \times m$ matrix whose n rows are a basis for the row space of A.

(iii) $A = BC$. (B.3)

Then, by the "Theorem of Corresponding Minors" [31, pp. 14, 15], $(B'B)$ and (CC') are nonsingular $n \times n$ matrices. The pseudo inverse of A is given by

$$A^\dagger = C'(CC')^{-1}(B'B)^{-1}B'. \tag{B.4}$$

3. Existence and Uniqueness Theorems [34, pp. 600, 601]

Moore's Theorem [35, p. 197-203]:

Given a finite matrix A, there is a unique matrix $A^\dagger$ (called "general reciprocal" by Moore) such that, for suitable matrices L and R,

$$AA^\dagger A = A, \qquad A^\dagger = LA^* = A^*R. \tag{B.5}$$

This $A^\dagger$ satisfies

$$A^\dagger AA^\dagger = A^\dagger \tag{B.6}$$

and $AA^\dagger$, $A^\dagger A$ are Hermitian matrices. (Note: A^* is the conjugate transpose of A.)

Proof: [34, p. 600]

(i) Let $n = \text{rank } A$. There are nonsingular matrices P and Q such that

$$PAQ = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix} = J \tag{B.7}$$

is the canonical diagonal form of A. [22, vol. I; p. 141]

Then $A^\dagger = QJ^*P$ satisfies $AA^\dagger A = A$. (B.8)

Express the columns of $A^\dagger$ as sums of vectors in the column subspace of A^* and vectors orthogonal to the columns of A^*:

$$A^\dagger = A^*X_2 + X_3, \qquad AX_3 = 0. \tag{B.9}$$

Similarly, write

$$X_2 = X_4A^* + X_5, \qquad X_5A = 0. \tag{B.10}$$

Then

$$A \;=\; AA^{\dagger}A \;=\; A[A^{*}(X_4A^{*} + X_5) + X_3]A$$

$$= \; AA^{*}X_4A^{*}A.$$

Hence (B.5) holds if we take

$$\begin{aligned} A^{\dagger} &= A^{*}X_4A^{*} \\ L &= A^{*}X_4 \\ R &= X_4A^{*}. \end{aligned} \tag{B.11}$$

This proves the existence part of Moore's theorem.

(ii) If $AXA = 0$, $X = YA^{*} = A^{*}Z$, then

$$(AX)(AX)^{*} \;=\; AXX^{*}A^{*} \;=\; AX(YA^{*})^{*}A^{*} \;=\; (AXA)Y^{*}A^{*} \;=\; 0$$

implies $AX = 0$. But then we have

$$X^{*}X \;=\; (A^{*}Z)^{*}X \;=\; Z^{*}(AX) \;=\; 0 \qquad X \;=\; 0.$$

Hence all solutions $A^{\dagger}$ of (B.5) are the same. This proves the uniqueness part of Moore's theorem.

(iii) Since $(AA^{\dagger}A)A^{\dagger}A \;=\; A$, and, by (B.11),

$$A^{\dagger}AA^{\dagger} \;=\; (A^{*}X_4A^{*})A(A^{*}X_4A^{*}) \;=\; L_1A^{*} \;=\; A^{*}R_1,$$

where $L_1 \;=\; A^{*}X_4A^{*}AA^{*}X_4$, $R_1 \;=\; X_4A^{*}AA^{*}X_4A^{*}$,

it follows that $(A^{\dagger}AA^{\dagger})$ satisfies Moore's conditions (B.5) for the pseudo inverse. By uniqueness of $A^{\dagger}$, we have

$$A^{\dagger}AA^{\dagger} \;=\; A^{\dagger}.$$

$$\text{(iv)}\quad AA^{\dagger} = (AA^{*}X_4)A^{*} = A(A^{*}X_4A^{*})A^{\dagger *}A^{*} = AA^{\dagger}A^{\dagger *}A^{*}$$

$$= (AA^{\dagger}A^{\dagger *}A^{*})^{*} = (AA^{\dagger})^{*}.$$

Similarly, $A^{\dagger}A = (A^{\dagger}A)^{*}$. This completes the proof of Moore's theorem.

Penrose's Theorem [33, pp. 17-19]

Given a finite matrix A, there is a unique matrix $A^{\dagger}$ (called "generalized inverse" by Penrose) such that

$$\begin{aligned} AA^{\dagger}A &= A & \qquad (AA^{\dagger})^{*} &= AA^{\dagger} \\ A^{\dagger}AA^{\dagger} &= A^{\dagger} & \qquad (A^{\dagger}A)^{*} &= A^{\dagger}A. \end{aligned} \tag{B.12}$$

Proof: [34, p. 601]

Let A be fixed. If $A^{\dagger}$, L, R satisfy (B.5) and (B.6), then (B.12) holds, so $A^{\dagger}$ exists.

Conversely, if (B.12) holds, then

$$A^{\dagger} = A^{\dagger}(AA^{\dagger}) = A^{\dagger}(AA^{\dagger})^{*} = (A^{\dagger}A^{\dagger *})A^{*}$$

$$A^{\dagger} = (A^{\dagger}A)A^{\dagger} = (A^{\dagger}A)^{*}A^{\dagger} = A^{*}(A^{\dagger *}A^{\dagger}).$$

Hence (B.5) follows, with $L = A^{\dagger}A^{\dagger *}$, $R = A^{\dagger *}A^{\dagger}$. Therefore, by Moore's theorem, (B.12) has exactly one solution $A^{\dagger}$.

This completes the proof of Penrose's theorem, and furthermore proves the

Corollary:

Moore's and Penrose's definitions, of the pseudo inverse $A^{\dagger}$ of a given finite matrix A, are equivalent.

4. Properties [31, pp. 8, 9][32, p. 15]

Let A be an arbitrary $p \times m$ matrix, rank $A = n$, and let $A^\dagger$ be the pseudo inverse of A, as defined in Section 1. Let B and C be constructed as in Section 2.

(i) $A^\dagger$ exists and is unique. (Penrose's theorem, see Section 3.)

(ii) If A is nonsingular, then $A^\dagger = A^{-1}$.

(iii) $(A^\dagger)^\dagger = A$.

(iv) $AA^\dagger = B(B'B)^{-1}B'$ is the unique orthogonal projector for the column space of A, i.e. given any $p \times 1$ vector x, $AA^\dagger x$ is the orthogonal projection of x upon the column space of A.

Similarly, $A^\dagger A = C'(CC')^{-1}C$ is the unique orthogonal projector for the row space of A.

(v) $AA^\dagger$ and $A^\dagger A$ are symmetric, idempotent matrices.

(vi) The row space of $A^\dagger$ is the column space of A; the column space of $A^\dagger$ is the row space of A.

(vii) $B^\dagger$ is a $n \times p$ matrix whose rows span the column space of A; $C^\dagger$ is a $m \times n$ matrix whose columns span the row space of A.

(viii) $B^\dagger B = I_n$; $CC^\dagger = I_n$.

(ix) $A^\dagger = C^\dagger B^\dagger$.

REFERENCES

1. George A. Baker, "The Theory and Application of the Padé Approximant Method", Advances in Theoretical Physics 1 (1965), 1-58.

2. E. W. Cheney, INTRODUCTION TO APPROXIMATION THEORY, McGraw-Hill (1966).

3. G. Frobenius, "Über Relationen zwischen Näherungsbrüchen von Potenzreihen", J. für reine u. angew. Math. 90 (1881), 1-17.

4. Peter Henrici, "The Quotient-Difference Algorithm", N.B.S. Applied Mathematics Series 49 (1958), 23-46.

5. Peter Henrici, ELEMENTS OF NUMERICAL ANALYSIS, John Wiley, New York (1964).

6. B.-L. Ho, "On Effective Construction of Realizations from Input-Output Descriptions", Ph.D. Dissertation, Stanford University (1966).

7. R. E. Kalman, Y. C. Ho and K. S. Narendra, "Controllability of Linear Dynamical Systems", Contrib. to Diff. Eq., Vol. 1, No. 2; John Wiley, New York (1962), pp 189-213.

8. R. E. Kalman, "Mathematical Description of Linear Dynamical Systems", J. SIAM Control, Ser. A, Vol. 1, No. 2 (1963), pp 152-192.

9. Serge Lang, ALGEBRA, Addison-Wesley (1965).

10. D. F. Mayers, "Interpolation by Rational Functions", Ch. 12, Methods of Numerical Approximation, D. C. Hanscomb (ed.), Oxford, Pergamon Press (1966).

11. M. H. Padé, "Sur la représentation approchée d'une fonction par des fractions rationelles", Ann. Scientifiques de l'École Normale Supérieure 3e Série, 9 (1892), Supplement pp. S.1-S.96.

12. M. H. Padé, "Sur la distribution des réduites anormales d'une fonction", Comptes Rendus 130 (1900), pp. 102-104

13. Oskar Perron, DIE LEHRE VON DEN KETTENBRÜCHEN, Teubner (1913).

14. Daniel Shanks, "Nonlinear Transformations of Divergent and Slowly Convergent Sequences", J. Math. and Phys. 34 (1955), pp 1-42.

15. Richard S. Varga, MATRIX ITERATIVE ANALYSIS, Prentice-Hall (1962).

16. B. L. van der Waerden, MODERN ALGEBRA I, 2nd Edition, Ungar, New York (1953).

17. H. S. Wall, ANALYTIC THEORY OF CONTINUED FRACTIONS, Van Nostrand (1948).

18. P. Wynn, "Upon Systems of Recursions which Obtain Among the Quotients of the Padé Table", Numerische Mathematik 8 (1966), pp 264-269.

19. O. Zarisky and P. Samuel, COMMUTATIVE ALGEBRA, Vol. II, Van Nostrand, Princeton (1960).

20. Rudolf Zurmühl, MATRIZEN UND IHRE TECHNISCHEN ANWENDUNGEN, 4th ed., Springer, Berlin (1964).

21. Evelyn Frank, "Corresponding Type Continued Fractions", Am. J. of Math., 68 (1946), pp 89-108.

22. F. R. Gantmacher, THE THEORY OF MATRICES, 2 vols., Chelsea Publishing Company, New York (1959).

23. J. Hadamard, "Essai sur l'étude des functions données par leur dévelopment de Taylor", J. Math. Pure et Appl. (Ser. 4) 8 (1892), pp. 101-186.

24. A. S. Householder, PRINCIPLES OF NUMERICAL ANALYSIS, McGraw-Hill, New York (1953).

25. Ludwig Bieberbach, LEHRBUCH DER FUNKTIONENTHEORIE, Vol. II: MODERNE FUNKTIONENTHEORIE, 2nd ed., B. G. Teubner, Leipzig (1931); Chelsea Publishing Co., New York (1945).

26. P. Dienes, THE TAYLOR SERIES, Clarendon Press, Oxford (1931).

27. Lotfi A. Zadeh and Charles A. Desoer, LINEAR SYSTEM THEORY: The State Space Approach, McGraw-Hill, New York (1963).

28. J. M. Souriau, "Une méthode pour la décomposition spectrale et l'inversion des matrices", Comptes Rendus 227 (1948), pp. 1010,1011.

29. R. E. Kalman, "When is a Linear Control System Optimal?", Trans. ASME 86D (1964), pp. 51-60.

30. P. Wynn, "The Rational Approximation of Functions which are Formally Defined by a Power Series Expansion", Math. of Computation 14 (1960), pp. 147-186.

31. A. S. Householder, THE THEORY OF MATRICES IN NUMERICAL ANALYSIS, Blaisdell (1964).

32. T. N. E. Greville, "Some Applications of the Pseudoinverse of a Matrix", SIAM Review 2 (1960), pp. 15-22.

33. R. Penrose, "On Best Approximate Solutions of Linear Matrix Equations", Proc. Cambridge Phil. Soc., 52 (1956), pp. 17-19; "A Generalized Inverse for Matrices", ibid., 52 (1956), p. 406.

34. R. Rado, "Note on Generalized Inverses of Matrices", Proc. Cambridge Phil. Soc., 52 (1956), pp. 600-601.

35. E. H. Moore, GENERAL ANALYSIS - Part I, Memoirs Am. Phil. Soc., 1 (1935), vi + 231.

36. B. L. Ho and R. E. Kalman, "Effective Construction of Linear State-Variable Models from Input/Output Data", Proc. Third Annual Allerton Conf. on Circuit and System Theory, Monticello, Illinois (1965), pp. 449-459.

37. John R. Rice, "Nonlinear Approximation", Proc. Symp. on Approximation of Functions, G. M. Res. Labs., Warren, Mich., 1964; H. L. Garabedian, editor; Elvesier (1965), pp. 111-133.

38. B. L. Ho and R. E. Kalman, "Effective Construction of Linear State-Variable Models from Input/Output Functions", Regelungstechnik 14, No. 12 (1966), pp. 545-548.