Vladimír Kučera

Algebraic theory of discrete optimal control for multivariable systems [I.]

Kybernetika, Vol. 10 (1974), No. Suppl, (1),3--56

Persistent URL: http://dml.cz/dmlcz/124938

Kybernetika

Algebraic Theory of Discrete Optimal Control for Multivariable Systems

Vladimír Kučera

ACADEMIA

PRAHA

PREFACE

This work provides a new algebraic theory of optimal control for discrete linear constant systems. The author has developed the essentials of this approach in a series of recent papers dealing with single-variable systems. Here is a natural but nontrivial generalization to multivariable systems.

The subject is divided into six chapters. In Chapter 1 (Preliminaries) the author builds a new mathematical machinery for analysis and synthesis of discrete linear constant systems. It is essentially the polynomial algebra and algebra of polynomial matrices. Further, a matrix linear Diophantine equation is introduced, the solvability criterion established, and an algorithm to find its general solution is presented. It will turn out that all problems of optimal control and stabilization will reduce to solving such an equation.

In Chapter 2 (Systems) we shall define, in an axiomatic way, a finite dimensional, linear, discrete, constant system over an arbitrary field and give the relations to its input-output description by a transfer function matrix. Further some important factorizations of polynomials and matrices, needed in the sequel, are defined and the associated computational aspects are briefly discussed.

Chapter 3 (Open-Loop Control) is devoted to the simplest and basic problem of control. The open-loop configuration is defined and optimal control problems in the sense of minimum transfer time and least squares of the error are posed and solved. The solution is complete, contains the existence condition, and is not restricted to stable systems.

Chapter 4 (Closed-Loop Stability) contains the basic preliminary material for synthesizing closed-loop systems. The explicit formulas for the characteristic and invariant polynomials of the closed-loop system are established. They make it possible to formulate and solve the problem of assigning desired invariant polynomials (and hence a characteristic polynomial) to the closed-loop system by dynamical output feedback. The main result is the fundamental necessary and sufficient condition of stability for the closed-loop system. This condition takes on the form of two coupled linear Diophantine equations and it is indispensable in the closed-loop system theory.

In Chapter 5 (Closed-Loop Control) we discuss the most common control problem. The usual closed-loop configuration is considered and optimal control problems are posed and solved in the sense of minimum transfer time and least squares of the error. A particular emphasis is placed on stability of the closed-loop system. The solution is complete, contains the existence condition, and it is not restricted to stable systems.

In Chapter 6 (Decoupling a Multivariable System) we focus our attention on some problems intrinsically relevant to multivariable systems. We first discuss the invertibility of a system, find a minimum-delay inverse of minimal dimension, and then pose and solve the stable decoupling problem. The ultimate purpose of decoupling is to control the system and hence all usual control problems are solved for the decoupled system.

For the sake of better orientation we shall summarize here some notation conventions. Throughout the paper, unless otherwise stated, the following notation is used:

rings, fields	$\mathfrak{G}, \mathfrak{F}, \mathfrak{R}, \ldots$
numbers (in a field)	$\lambda, \varphi, \ldots$
matrices over numbers	$\Lambda, \Phi, \ldots$
polynomials, integers	$a, b, \ldots, n, \ldots$
polynomial matrices	$A, B, \ldots$
rational functions	$\boldsymbol{s}, \boldsymbol{r}, \ldots$
rational matrices	$\boldsymbol{S}, \boldsymbol{R}, \ldots$
state vectors	$\mathbf{x}, \mathbf{z}, \ldots$
state matrices	$\mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots$
systems, sets	$\mathscr{S}, \mathscr{R}, \ldots, \mathscr{X}, \ldots$

REFERENCES

[1] Ackermann J.: Abtastregelung. Springer, Berlin 1972.

[2] Åström K. J.: Introduction to Stochastic Control Theory. Academic Press, New York 1970.

[3] Åström K. J., Jury E. J., Agniel R. G.: A numerical method for the evaluation of complex integrals. IEEE Trans. Automatic Control *AC-15* (1970), 4, 468—471.

[4] Barnet S.: Matrices, polynomials, and linear time-invariant systems. IEEE Trans. Automatic Control *AC-18* (1973), 1, 1—10.

[5] Bourbaki N.: Algèbre, Chapter 5. Hermann & Cie, Paris 1956.

[6] Bourbaki N.: Algèbre, Chapter 8. Hermann & Cie, Paris 1966.

[7] Bourbaki N.: Algèbre Commutative. Hermann & Cie, Paris 1961.

[8] Brockett R. W.: Finite Dimensional Linear Systems. Wiley, New York 1970.

[9] Chang S. S. L.: Synthesis of Optimum Control Systems. McGraw-Hill, New York 1961.

[10] Chen C. T.: Representation of linear time-invariant composite systems. IEEE Trans. Automatic Control *AC-13* (1968), 3, 277—283.

[11] Davison E. J.: On pole assignment in linear systems with incomplete state feedback. IEEE Trans. Automatic Control *AC-16* (1970), 3, 348—350.

[12] Гантмахер Ф. Р.: Теория матриц. Гостехиздат, Москва 1953.

[13] Hardy G. H., Wright E. M.: An Introduction to the Theory of Numbers. Clarendon Press, Oxford 1960.

[14] Heymann M.: Comment on pole assignment in multi-input linear systems. IEEE Trans. Automatic Control *AC-13* (1968), 6, 748—749.

[15] Янушевский Р. Т.: Теория линейных многосвязных систем управления. Наука, Москва 1973.

[16] Johnson C. D.: Stabilization of linear dynamical systems with output feedback. Proc. 5th IFAC Congress, Paris 1972.

[17] Kalman R. E.: Contributions to the theory of optimal control. Bol. Soc. Mat. Mexicana *5* (1960), 102—119.

[18] Kalman R. E.: On the general theory of control systems. Proc. 1st IFAC Congress, Moscow 1960.

[19] Kalman R. E.: Mathematical description of linear dynamical systems. SIAM J. Control *1* (1963), 152—192.

[20] Kalman R. E.: Irreducible realizations and the degree of a rational matrix. SIAM J. Control *13* (1965), 2, 520—544.

[21] Kalman R. E.: Algebraic theory of linear systems. Proc. 3rd Allerton Conference on Circuit and System Theory, Illinois 1965, 563—577.

[22] Kalman R. E.: Lectures on Controllability and Observability. Centro Internazionale Matematico Estivo 1968.

[23] Kalman R. E., Falb P. L., Arbib M. A.: Topics in Mathematical System Theory. McGraw-Hill, New York 1969.

[24] Kalman R. E., Bertram J. E.: A unified approach to the theory of sampling systems. Journal of the Franklin Institute *267* (1959), 5, 405—436.

[25] Kučera V.: The discrete Riccati equation of optimal control. Kybernetika *8* (1972), 5, 430—447.

[26] Kučera V.: A contribution to matrix quadratic equations. IEEE Trans. Automatic Control *AC-17* (1972), 3, 344—347.

[27] Kučera V.: On nonnegative definite solutions of matrix quadratic equations. Proc. 5th IFAC Congress, Paris 1972.

[28] Kučera V.: On nonegative definite solutions of matrix quadratic equations. Automatica *8* (1972), 4, 413—423.

[29] Kučera V.: A review of the matrix Riccati equation. Kybernetika *9* (1973), 1, 42—61.
[30] Kučera V.: Algebraic theory of discrete optimal control for single-variable systems I — Preliminaries. Kybernatika *9* (1973), 2, 94—107.
[31] Kučera V.: Algebraic theory of discrete optimal control for single-variable systems II — Open-loop control. Kybernatika *9* (1973), 3, 206—221.
[32] Kučera V.: Algebraic theory of discrete optimal control for single-variable systems III — Closed-loop control. Kybernetika *9* (1973), 4, 291—312.
[33] Kučera V.: Closed-loop stability of discrete linear single-variable systems. Kybernetika *10* (1974), 2, 162—187.
[34] Kučera V.: Constrained optimal control — The algebraic approach. Kybernetika *10* (1974), 4, 317—349.
[35] Kučera V.: Algebraic approach to discrete optimal control. To appear in IEEE Trans. Automatic Control.
[36] Kučera V.: On the minimal order inverses for linear time-invariant discrete systems. Proc. Joint Automatic Control Conference, St. Louis 1971, 312—318.
[37] Kučera V.: The pole assignment in linear systems. Proc. International Conference on Systems and Control, Coimbatore, India, 1973.
[38] Kuroš A. G.: Kapitoly z obecné algebry. Academia, Prague 1968.
[39] Lambek J.: Lectures on Rings and Modules. Blaisdell, Waltham, Mass. 1966.
[40] Lang S.: Algebra. Addison-Wesley, Reading, Mass. 1965.
[41] MacLane S., Birghoff G.: Algebra. Macmillan, New York 1967.
[42] Михелович Ш. Х.: Теория чисел. Высшая школа, Москва 1967.
[43] Mitter S. K., Foulkes R.: Controllability and pole assignment for discrete-time linear systems defined over arbitrary fields. SIAM J. Control *9* (1971), 1, 1—8.
[44] Mordell L. J.: Diophantine Equations. Academic Press, London 1969.
[45] Morse A. S., Wonham W. M.: Decoupling and pole assignment by dynamic compensation. SIAM J. Control *8* (1970), 3, 317—337.
[46] Nekolný J.: Numerická interpolace a náhrada analytických funkcí. Report No. 357, ÚTIA ČSAV, Prague 1971.
[47] Peterka V., Halousková A.: Tally estimate of Åström model for stochastic systems. Identification and Process Parameter Estimation, 2nd Prague IFAC Symposium, 1970.
[48] Peterka V.: On steady-state minimum variance control strategy. Kybernetika *8* (1972), 3, 219—232.
[49] Rosenbrock H. H.: State Space and Multivariable Theory. Wiley, New York 1970.
[50] Roth W. E.: The equations $AX - YB = C$ and $AX - XB = C$ in matrices. Proc. Amer. Math. Soc. *3* (1952), 390—396.
[51] Sage A. P.: Optimum Systems Control. Prentice-Hall, Englewood Cliffs, N. J. 1968.
[52] Sain M. K., Massey J. L.: Invertibility of linear time-invariant systems. IEEE Trans. Automatic Control *AC-14* (1969), 2, 141—149.
[53] Silverman L. M.: Inversion of multivariable linear systems. IEEE Trans. Automatic Control *AC-14* (1969), 3, 270—276.
[54] Silverman L. M., Payne H. J.: Input-output structure of linear systems with application to the decoupling problem. SIAM J. Control *9* (1971), 2, 199—233.
[55] Strejc V. et al.: Syntéza regulačních obvodů s číslicovým počítačem. NČSAV, Prague 1965.
[56] Tou J. T.: Digital and Sampled-data Control Systems. McGraw-Hill, New York 1959.
[57] van der Waerden B. L.: Moderne Algebra, vol. I. Springer, Berlin 1930.
[58] Vostrý Z.: Нумерический метод спектральной факторизации полиномов. Kybernetika *8* (1972), 4, 323—332.
[59] Vostrý Z.: A numerical method of matrix spectral factorization. Kybernetika *8* (1972), 5, 448—470.

[60] Vostrý Z.: Syntéza lineárního diskrétního řízení podle kvadratického kritéria pro soustavy s nestejným počtem vstupů a výstupů. C.Sc. Thesis, ÚTIA ČSAV, Prague 1973.
[61] Wang S. H., Davison E. J.: A minimization algorithm for the design of linear multivariable systems. IEEE Trans. Automatic Control *AC-18* (1973), 3, 220–225.
[62] Wonham W. M.: On pole assignment in multi-input controllable linear systems. IEEE Trans. Automatic Control *AC-12* (1967), 6, 660–665.
[63] Wonham W. M., Morse A. S.: Decoupling and pole assignment in linear multivariable systems — A geometric approach. SIAM J. Control *8* (1970), 1, 1–18.
[64] Youla D. C.: On the factorization of rational matrices. IRE Trans. Information Theory *IT-17* (1961), 3, 172–189.
[65] Zadeh L. A., Desoer C. A.: Linear Systems Theory — The State Space Approach. McGraw-Hill, New York 1963.
[66] Zariski O., Samuel P.: Commutative Algebra, vols. I and II. Van Nostrand, Princeton, N. J. 1958.

1. PRELIMINARIES

1.1. Introduction

There are two major approaches to the analysis and synthesis of control systems. It is the complex-domain approach and the time-domain approach. Either methods has its own advantages and objections and its own fields of applications. An attempt to compare the two methods has been made in [29, 33].

In the former approach [9, 15, 48, 55, 56, 60] we transform the problem into the language of functions of a complex variable. This simplifies and visualizes the manipulations but requires a rather advanced mathematical tool (the theory of analytic functions, contour integration, the residue theorem, the principle of argument, the Z-transform theory, etc.) to lend mathematical respectability to those methods. Moreover, we are not able to give a rigorous definition of a system within this framework since we are confined to input-output properties. This usually leads to considering stable systems only. Further, this theory does not apply to systems defined over arbitrary fields and is limited to linear constant systems. From the engineering point of view, a great advantage of this method consists in the fact that it requires just the available output information to implement the optimal control. However, the computations associated with obtaining the optimal control are burdensome and not suited for machine processing.

On the other hand, the latter approach [1, 8, 17, 18, 19, 23, 24, 49, 51, 65] introduces the idea of state thus making an exact definition of a system possible. It works with the notion of state space and profits from the theory of recursive equations in matrix form. Finite automata and related systems become a special part of the theory. This approach is particularly useful for nonconstant or nonlinear systems. However, a control engineer may be disappointed. The state of a system is in general an abstract entity and frequently not accessible in a real system. In applications the state must be, therefore, recovered by an observer [23] and the overall system becomes unnecessarily complex. Another objection involves computational aspects. Though the computations can easily be algorithmized, the use of matrices leads to superfluous operations [25, 26, 27, 28, 29].

The new algebraic approach reflects the most recent trend in linear system theory [21, 22, 23]. To find the optimal control for single-variable systems [30, 31, 32, 33, 34, 35] we work with the input-output responses viewed as ratios of two abstract polynomials. The whole theory is based on polynomial algebra and the synthesis procedure reduces to solving linear Diophantine equations in polynomials. This is conceptually simple, requires no advanced mathematical background, applies to systems defined over an arbitrary field no matter they are stable or unstable, and yields effective and unified computational algorithms. The algebraic approach also retains the advantage of utilizing only the available output information for control. However, the method does not seem to generalize for nonconstant or nonlinear systems.

The input-output response of a multivariable system is a matrix, not a single function, and hence it should be viewed as a product of a polynomial matrix and the inverse of another polynomial matrix in order that the algebraic approach may be applicable. Once this is appropriately done, the multivariable theory becomes a natural generalization of the single-variable theory. Even though the generalization is natural, it is by no means trivial. Matrix multiplication is not commutative unlike the multiplication of polynomials and hence the order in which a signal enters two systems in cascade is essential. Another complication stems from the fact that matrices often contain some degrees of freedom in their structure and hence the solution is, as a rule, not unique. This is a double-edged property. It makes the synthesis procedure depend upon somewhat arbitrary choices and, therefore, more complicated and less suited for machine processing. On the other hand, it leaves more room for the engineer to realize the optimal system according to additional requirements.

1.2. Rings and fields

In [30] we have defined a commutative ring as a basic algebraic structure relevant to the description of single-variable systems. To develop an adequate mathematical machinery for multivariable systems, we shall need the following more general concepts [5, 6, 7, 12, 38, 39, 40, 41, 66].

A set $\mathfrak{C}$ in which two laws of composition are given, the first written additively and the second multiplicatively, is called a (noncommutative) *ring* if the following axioms hold.

A_0 (closedness): $a, b \in \mathfrak{C}$ implies $a + b \in \mathfrak{C}$,
A_1 (associativity): $a, b, c \in \mathfrak{C}$ implies $a + (b + c) = (a + b) + c$,
A_2 (commutativity): $a, b \in \mathfrak{C}$ implies $a + b = b + a$,

A_3	(zero element):	$a \in \mathfrak{C}$, there exists $0 \in \mathfrak{C}$ such that $0 + a = a$,
A_4	(additive inverse):	$a \in \mathfrak{C}$, there exists $-a \in \mathfrak{C}$ such that $-a + a = 0$,
M_0	(closedness):	$a, b \in \mathfrak{C}$ implies $ab \in \mathfrak{C}$,
M_1	(associativity):	$a, b, c \in \mathfrak{C}$ implies $a(bc) = (ab)c$,
M_2	(identity element):	$a \in \mathfrak{C}$, there exists $1 \in \mathfrak{C}$ such that $1a = a1 = a$,
D	(distributivity):	$a, b, c \in \mathfrak{C}$ implies $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$.

We do not exclude the possibility that $0 = 1$. Then for any $a \in \mathfrak{C}$ we have $a = 1a = 0a = 0$ so that the ring contains just the element 0.

If an element $e \in \mathfrak{C}$ has a multiplicative inverse, i.e. an element $e^{-1} \in \mathfrak{C}$ exists such that $ee^{-1} = e^{-1}e = 1$, we call e a *unit* of $\mathfrak{C}$.

If every nonzero element of $\mathfrak{C}$ has a multiplicative inverse, if multiplication is commutative, and if $0 \neq 1$ we call $\mathfrak{C}$ a *field.*

Consider elements $a, b \in \mathfrak{C}$. If $a = be_2$, where e_2 is a unit of $\mathfrak{C}$, the elements a and b are called *right associates* in $\mathfrak{C}$ and if $a = e_1b$ where e_1 is a unit of $\mathfrak{C}$, the a and b are called *left associates* in $\mathfrak{C}$. If $a = e_3be_4$, where e_3, e_4 are units of $\mathfrak{C}$, the a and b are called simply *associates* in $\mathfrak{C}$.

It is seen that a ring is not closed under division. Consider elements $a, b \in \mathfrak{C}$, $b \neq 0$. We say that b *divides a on the left* if there exists an element $c_2 \in \mathfrak{C}$ such that $a = bc_2$ and we say that b *divides a on the right* if there exists an element $c_1 \in \mathfrak{C}$ such that $a = c_1b$. Finally, we simply say that b *divides* a and write $b \mid a$, if there exist elements $c_3, c_4 \in \mathfrak{C}$ such that $a = c_3bc_4$. The b is called respectively a *left divisor*, a *right divisor*, and a *divisor* of a.

An element $a \in \mathfrak{C}$ is a *divisor of zero* if there exists a $b \in \mathfrak{C}$, $b \neq 0$ such that either $ab = 0$ or $ba = 0$.

Consider elements $a, b \in \mathfrak{C}$. A *greatest common left divisor* of a and b is element $d_1 \in \mathfrak{C}$ such that

(a) d_1 divides both a and b on the left,

(b) $c_1 \in \mathfrak{C}$, c_1 divides both a and b on the left implies that c_1 divides d_1 on the left.

A *greatest common right divisor* of a and b is an element $d_2 \in \mathfrak{C}$ such that

(a) d_2 divides both a and b on the right,

(b) $c_2 \in \mathfrak{C}$, c_2 divides both a and b on the right implies that c_2 divides d_2 on the right.

Finally, a *greatest common divisor* of a and b is an element $d \in \mathfrak{C}$, denoted by (a, b), such that

(a) $d \mid a$, $d \mid b$,

(b) $c \in \mathfrak{C}$, $c \mid a$, $c \mid b$ implies $c \mid d$.

It is to be noted that if d is a greatest common (left, right) divisor of $a, b \in \mathfrak{C}$ then all (left, right) divisors of a and b are (right, left) associates of d in $\mathfrak{C}$.

Consider elements $a, b \in \mathfrak{C}$. If all greatest common left divisors of a and b are units of $\mathfrak{C}$, the a and b are said to be *left coprime* in $\mathfrak{C}$. If all greatest common right divisors of a and b are units of $\mathfrak{E}$, the a and b are said to be *right coprime* in $\mathfrak{E}$. If all greatest common divisors of a and b are units of $\mathfrak{E}$, i.e. when $(a, b) = 1$ up to a unit of $\mathfrak{E}$, the a and b are said to be *coprime* in $\mathfrak{E}$.

A nonzero, nonunit element $p \in \mathfrak{E}$ is *prime* in $\mathfrak{E}$ if $a \mid p$ implies that a is either a unit of $\mathfrak{E}$ or an associate of p.

For example, the set $\mathfrak{Z}$ of integers constitute a commutative ring. The units of $\mathfrak{Z}$ are ± 1, the only divisor of zero is 0, and the primes of $\mathfrak{Z}$ are prime numbers. The rationals $\mathfrak{Q}$, algebraic numbers $\mathfrak{A}$, reals $\mathfrak{R}$, and complex numbers $\mathfrak{C}$ all form fields. The set $\mathfrak{Z}_n$ of residue classes of integers modulo an integer n is an example of a finite ring. To recall, $u, v \in \mathfrak{Z}$ belong to the same residue class modulo n if $n \mid u - v$, which is written as $u = v \bmod n$. Each residue class contains exactly one element less than n. Hence the $\mathfrak{Z}_n$ is (as a set) isomophic with the set of integers $\{0, 1, \ldots, n - 1\}$. The units of $\mathfrak{Z}_n$ are integers coprime with n while the divisors of zero are integers not coprime with n. If $n = p$, a prime integer, the $\mathfrak{Z}_p$ becomes a (finite) field.

We shall give further examples of commutative rings. Given a field $\mathfrak{F}$, the *polynomials*

$$a = \alpha_0 + \alpha_1 z + \ldots + \alpha_n z^n \,, \quad \alpha_k \in \mathfrak{F} \,, \quad n < \infty$$

over $\mathfrak{F}$ in the indeterminate z with the usual definition of addition and multiplication constitute a ring $\mathfrak{F}[z]$, see [30]. If $\alpha_n \neq 0$ the n is the *degree* of a, denoted as ∂a. By convention, $\partial 0 = -\infty$. If $\alpha_n = 1$ then the a is a *monic* polynomial. The units of $\mathfrak{F}[z]$ are polynomials of zero degree (which are viewed as isomorphic with $\mathfrak{F}$), the only divisor of zero is 0, and the primes of $\mathfrak{F}[z]$ are polynomials irreducible in $\mathfrak{F}[z]$.

The set $\mathfrak{F}[z]_a$ of *residue classes of polynomials* of $\mathfrak{F}[z]$ modulo a polynomial $a \in \mathfrak{F}[z]$ is also a ring. To recall, two polynomials $u, v \in \mathfrak{F}[z]$ belong to the same residue class modulo a if $a \mid u - v$, which is written as $u = v \bmod a$. In each residue class there is exactly one polynomial with degree less than ∂a. Hence the $\mathfrak{F}[z]_a$ is (as a set) isomorphic with the set of all polynomials of $\mathfrak{F}[z]$ with degree less than ∂a. The units of $\mathfrak{F}[z]_a$ are polynomials coprime with a, while the divisors of zero are polynomials not coprime with a. If $a = p$, a polynomial prime in $\mathfrak{F}[z]$, the $\mathfrak{F}[z]_p$ is a field.

It is well-known that the field $\mathfrak{Q}$ of rationals is the *quotient field* of the ring $\mathfrak{Z}$ of integers, i.e. the set of all ratios q/p, where $q \in \mathfrak{Z}$, $p \in \mathfrak{Z} - \{0\}$. Similarly, denote $\mathfrak{F}(z^{-1})$ the quotient field of $\mathfrak{F}[z]$. It is called the field of *rational functions* over $\mathfrak{F}$ and its elements have the form

$$\boldsymbol{a} = \frac{q}{p} \,, \quad q \in \mathfrak{F}[z] \,, \quad p \in \mathfrak{F}[z] - \{0\}$$

or

$$\boldsymbol{a} = \alpha_n z^{-n} + \alpha_{n+1} z^{-(n+1)} + \ldots, \quad \alpha_k \in \mathfrak{F} \,, \quad n \in \mathfrak{Z} \,,$$

which is obtained by formal long division q/p into ascending powers of z^{-1}. If $\alpha_n \neq 0$ the n is the *order* of $\boldsymbol{a}$, denoted as $\mathcal{O}\boldsymbol{a}$. In fact, $\mathcal{O}\boldsymbol{a} = \partial p - \partial q$.

Now consider the set of rational functions over $\mathfrak{F}$ with nonnegative order. They form a ring denoted by $\mathfrak{F}\{z^{-1}\}$ with elements of the form

$$a = \frac{q}{p}, \quad \partial q \leqq \partial p \tag{1.1}$$

or

$$a = \alpha_0 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \ldots, \quad \alpha_k \in \mathfrak{F}. \tag{1.2}$$

The $\mathfrak{F}\{z^{-1}\}$ is sometimes called the ring of *realizable rational functions*. The units of $\mathfrak{F}\{z^{-1}\}$ are elements of order 0, the only divisor of zero is 0, and the only prime of $\mathfrak{F}\{z^{-1}\}$ is z^{-1}. All elements of $\mathfrak{F}\{z^{-1}\}$ having the same order are associates.

Further consider a subset of $\mathfrak{F}\{z^{-1}\}$ consisting of elements (1.2) such that the sequence $\{\alpha_0, \alpha_1, \alpha_2, \ldots\}$ converges to zero. They constitute a ring of *stable realizable rational functions*, which will be denoted as $\mathfrak{F}^+\{z^{-1}\}$.

This motivates the following fundamental definition. A polynomial $a \in \mathfrak{F}[z]$ is said to be *stable* if $1/a \in \mathfrak{F}^+\{z^{-1}\}$. Then we can characterize the units of $\mathfrak{F}^+\{z^{-1}\}$ as elements (1.1) for which $\mathcal{O}a = 0$ and q is stable.

Of course, $\mathfrak{F}[z]$ is a subring of $\mathfrak{F}^+\{z^{-1}\}$ and $\mathfrak{F}^+\{z^{-1}\}$ is a subring of $\mathfrak{F}\{z^{-1}\}$ which, in turn, is contained in the field $\mathfrak{F}(z^{-1})$.

1.3. Matrix polynomials and rational functions

Given a field $\mathfrak{F}$, the set $\mathfrak{F}_{m,m}$ of $m \times m$ matrices over $\mathfrak{F}$ is an example of a non-commutative ring if $m > 1$. For $m = 1$ the $\mathfrak{F}_{1,1}$ is viewed as isomorphic with $\mathfrak{F}$. The identity and zero elements of the ring $\mathfrak{F}_{m,m}$ are respectively the identity matrix I_m and the zero matrix 0_m. A matrix A is a unit of $\mathfrak{F}_{m,m}$ if $\det A \neq 0$ (such an A is said to be *nonsingular*) and it is a divisor of zero in $\mathfrak{F}_{m,m}$ if $\det A = 0$ (such an A is called *singular*). On the other hand, the set $\mathfrak{F}_{l,m}$ of $l \times m$ matrices over $\mathfrak{F}$ is not a ring since the product of two $l \times m$ matrices is not defined whenever $l \neq m$ and hence axiom M_0 is not satisfied.

Given a field $\mathfrak{F}$, consider the set $\mathfrak{F}_{l,m}[z]$ of $l \times m$ matrices over the ring $\mathfrak{F}[z]$. These *polynomial matrices* are not a ring, either, unless $l = m$. The identity and zero elements of the ring $\mathfrak{F}_{m,m}[z]$ are the I_m and 0_m, respectively. A matrix $A \in \mathfrak{F}_{m,m}[z]$ is a unit of $\mathfrak{F}_{m,m}[z]$ if and only if $\det A$ is a unit of $\mathfrak{F}[z]$, i.e. if $\det A \in \mathfrak{F}$, $\det A \neq 0$. On the other hand, the A is a divisor of zero in $\mathfrak{F}_{m,m}[z]$ if and only if $\det A$ is a divisor of zero in $\mathfrak{F}[z]$, i.e. if $\det A = 0$.

A polynomial matrix $A \in \mathfrak{F}_{l,m}[z]$ may also be viewed as a polynomial

$$A = A_0 + A_1 z + \ldots + A_n z^n, \quad A_i \in \mathfrak{F}_{l,m}, \quad n < \infty$$

over $\mathfrak{F}_{l,m}$ in the indeterminate z, called a *matrix polynomial.* For example,

$$A = \begin{bmatrix} 1-z & 0 \\ z & 1-z^2 \\ 0 & z^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -1 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} z + \begin{bmatrix} 0 & 0 \\ 0 & -1 \\ 0 & 1 \end{bmatrix} z^2 .$$

If $A_n \neq 0$ the n is the *degree* of A, denoted as ∂A. We define $\partial 0 = -\infty$. If $l = m$ and $\det A_n \neq 0$, then the A is said to be a *regular* matrix polynomial. Observe the following property: if A, $B \in \mathfrak{F}_{m,m}[z]$, then $\partial AB \leqq \partial A + \partial B$ the equality holding if and only if A and/or B is a regular matrix polynomial.

We can say that a matrix polynomial is the same thing as a polynomial matrix. This interpretation of polynomial matrices makes it obvious that essentially the same mathematical machinery is being used for both multivariable and single-variable systems.

Even though the $\mathfrak{F}_{l,m}[z]$ is not a ring, many concepts can be defined similarly. Consider matrices A, $B \in \mathfrak{F}_{l,m}[z]$. If $A = BE_2$ where E_2 is a unit of $\mathfrak{F}_{m,m}[z]$, the A and B are called *right associates* and if $A = E_1 B$, where E_1 is a unit of $\mathfrak{F}_{l,l}[z]$, the A and B are called *left associates.* If $A = E_3 B E_4$, where E_3 is a unit of $\mathfrak{F}_{l,l}[z]$ and E_4 is a unit of $\mathfrak{F}_{m,m}[z]$, the A and B are called simply *associates.*

If $A \in \mathfrak{F}_{l,m}[z]$, $B \in \mathfrak{F}_{l,q}[z]$, $B \neq 0$ we say that B *divides A on the left* if there exists a matrix $C_2 \in \mathfrak{F}_{q,m}[z]$ such that $A = BC_2$. If $A \in \mathfrak{F}_{l,m}[z]$, $B \in \mathfrak{F}_{p,m}[z]$, $B \neq 0$ we say that B *divides A on the right* if there exists a matrix $C_1 \in \mathfrak{F}_{l,p}[z]$ such that $A = C_1 B$. If $A \in \mathfrak{F}_{l,m}[z]$, $B \in \mathfrak{F}_{p,q}[z]$, $B \neq 0$ we say that B divides A, and write $B \mid A$, if there exist matrices $C_3 \in \mathfrak{F}_{l,p}[z]$ and $C_4 \in \mathfrak{F}_{q,m}[z]$ such that $A = C_3 B C_4$.

Given matrices $A \in \mathfrak{F}_{l,m}[z]$, $B \in \mathfrak{F}_{l,q}[z]$, *a greatest common left divisor* of A and B is a matrix $D_1 \in \mathfrak{F}_{l,l}[z]$ such that

(a) D_1 divides both A and B on the left,

(b) $C_1 \in \mathfrak{F}_{l,l}[z]$, C_1 divides both A and B on the left implies that C_1 divides D_1 on the left.

Given matrices $A \in \mathfrak{F}_{l,m}[z]$, $B \in \mathfrak{F}_{p,m}[z]$, *a greatest common right divisor* of A and B is a matrix $D_2 \in \mathfrak{F}_{m,m}[z]$ such that

(a) D_2 divides both A and B on the right,

(b) $C_2 \in \mathfrak{F}_{m,m}[z]$, C_2 divides both A and B on the right implies that C_2 divides D_2 on the right.

Given matrices A, $B \in \mathfrak{F}_{m,m}[z]$, a *greatest common divisor* of A and B is a matrix $D \in \mathfrak{F}_{m,m}[z]$, denoted by (A, B), such that

(a) $D \mid A$, $D \mid B$,

(b) $C \in \mathfrak{F}_{m,m}[z]$, $C \mid A$, $C \mid B$ implies $C \mid D$.

If all greatest common left divisors of $A \in \mathfrak{F}_{l,m}[z]$ and $B \in \mathfrak{F}_{l,q}[z]$ are units of $\mathfrak{F}_{l,l}[z]$, the A and B are said to be *left coprime.* If all greatest common right divisors of $A \in \mathfrak{F}_{l,m}[z]$ and $B \in \mathfrak{F}_{p,m}[z]$ are units of $\mathfrak{F}_{m,m}[z]$, the A and B are said to be *right coprime.* If, finally, all greatest common divisors of A, $B \in \mathfrak{F}_{m,m}[z]$ are units of $\mathfrak{F}_{m,m}[z]$, the A and B are said to be *coprime.*

For convenience, given a polynomial $b \in \mathfrak{F}[z]$ and a matrix $A \in \mathfrak{F}_{l,m}[z]$ with elements $a_{ij} \in \mathfrak{F}[z]$, we define $b \mid A$ if and only if $b \mid a_{ij}$, $i = 1, 2, \ldots, l$ and $j = 1, 2, \ldots, m$. Also $(b, A) = (b, (a_{11}, a_{12}, \ldots, a_{lm}))$.

In [30] we have applied the well-known *division algorithm* in $\mathfrak{F}[z]$, viz. given $a, b \in \mathfrak{F}[z]$, polynomials q, r exist in $\mathfrak{F}[z]$ such that

$$a = bq + r$$

and they are uniquely determined by $\partial r < \partial b$. In a like manner, given matrices $A \in \mathfrak{F}_{l,m}[z]$ and $B \in \mathfrak{F}_{l,l}[z]$, B regular, we can perform the *left division algorithm*

$$A = BQ_1 + R_1 ,$$

where matrices $Q_1, R_1 \in \mathfrak{F}_{l,m}[z]$ are uniquely determined by $\partial R_1 < \partial B$. Similarly, given matrices $A \in \mathfrak{F}_{l,m}[z]$ and $B \in \mathfrak{F}_{m,m}[z]$, B regular, we can perform the *right division algorithm*

$$A = Q_2 B + R_2 ,$$

where matrices $Q_2, R_2 \in \mathfrak{F}_{l,m}[z]$ are uniquely determined by $\partial R_2 < \partial B$. See [12] for proofs and further details.

Unfortunately, the euclidean algorithm is not defined over $\mathfrak{F}_{l,m}[z]$ since repeated left or right division may not be executable (the divisor may not be regular) [4].

The structure of polynomial matrices is given by the classical invariant-factor theorem [12]. An arbitrary matrix $A \in \mathfrak{F}_{l,m}[z]$ is associated with a matrix

$$\operatorname{diag}\{a_1, a_2, \ldots, a_r, 0, \ldots, 0\} \in \mathfrak{F}_{l,m}[z]$$

all of whose elements are zero except those on the main diagonal, which are $a_1, a_2, \ldots, a_r$, possibly followed by zeros.

The $a_1, a_2, \ldots, a_r$ belong to $\mathfrak{F}[z]$ and $a_k \mid a_{k+1}$, $k = 1, 2, \ldots, r - 1$. They are uniquely determined by A to within units of $\mathfrak{F}[z]$ and are called the *invariant polynomials* of A. The integer r is the *rank* of A.

The a_k may be computed directly as

$$a_k = \frac{d_k}{d_{k-1}}, \quad k = 1, 2, \ldots, r$$

where

$d_0 = 1$ by convention,

$d_k =$ greatest common divisor of all $k \times k$ minors of A.

The representation

$$A = E_1 \operatorname{diag}\{a_1, a_2, \ldots, a_r, 0, \ldots, 0\} E_2 \tag{1.3}$$

is called the *canonical representation* of A. Not only are the units $E_1 \in \mathfrak{F}_{l,l}[z]$ and $E_2 \in \mathfrak{F}_{m,m}[z]$ not unique, but they may be of arbitrarily high degree.

Example:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & z \end{bmatrix} = \begin{bmatrix} 1 & z^{n-1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & z \end{bmatrix} \begin{bmatrix} 1 & 1 - z^{n} \\ 0 & 1 \end{bmatrix}.$$

The largest monic invariant polynomial a_r of A is called the *minimal polynomial* of A and is denoted by mp$A$.

It is instructive to describe an algorithm for obtaining representation (3). The algorithm is based on the elementary matrix transformations [12] of the following three types.

(a) Multiplying the ith row (jth column) by a unit e of $\mathfrak{F}[z]$. It is effected by premultiplying A by the (diagonal) matrix

$$\begin{bmatrix} 1 & & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & 1 & & & \\ & & e & & \\ & & & 1 & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & & \vdots & & 1 \end{bmatrix}$$
$$i\text{th column}$$

(postmultiplying A by the matrix

$$\begin{bmatrix} 1 & & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & 1 & & & \\ & & e & \cdots & \cdots \\ & & & 1 & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & & 1 \end{bmatrix} \ldots j\text{th row}).$$

(b) Interchanging the ith and jth rows (columns). This is effected by premultiplying (postmultiplying) A by the matrix

$$\begin{bmatrix} 1 & & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & 0 & \cdots & 1 & \cdots \\ & \vdots & & \vdots & \\ & 1 & \cdots & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & & 1 \end{bmatrix} \begin{matrix} \\ \\ \ldots i\text{th row} \\ \\ \ldots j\text{th row.} \\ \\ \\ \end{matrix}$$

(c) Adding the ith row (column) multiplied by a polynomial $q \in \mathfrak{F}[z]$ to the jth row (column). This is effected by premultiplying A by the matrix

$$\begin{bmatrix} 1 & & & \\ \cdots & \cdots & \cdots & \cdots \\ & 1 & & \\ & \vdots & & \\ & q & \cdots \; 1 & \\ \cdots & \cdots & \cdots & \cdots \\ & \vdots & \vdots & 1 \end{bmatrix}$$

$$\begin{matrix} i\text{th} & j\text{th} \\ \text{columns} & \end{matrix}$$

(postmultiplying A by the matrix

$$\begin{bmatrix} 1 & & & \\ \cdots & \cdots & \cdots & \cdots \\ & 1 & \cdots \; q \cdots & \\ & & \vdots & \\ & & 1 \cdots & \\ \cdots & \cdots & \cdots & \cdots \\ & & & 1 \end{bmatrix} \begin{matrix} \\ \\ \ldots i\text{th row} \\ \\ \ldots j\text{th row}). \\ \\ \\ \end{matrix}$$

Denoting the ith row jth column element of A as a_{ij}, the canonical representation of A is obtained as follows. Among all nonzero elements of A take the one having least degree and make it a_{11} by reordering the rows and columns of A. Carry out the divisions

$$a_{i1} = q_{i1}a_{11} + r_{i1}\,, \quad \partial r_{i1} < \partial a_{11}\,, \quad i = 2, 3, \ldots, l\,,$$

and

$$a_{1j} = a_{11}q_{1j} + r_{1j}\,, \quad \partial r_{1j} < \partial a_{11}\,, \quad j = 2, 3, \ldots, m\,.$$

If at least one of the r_{i1}'s or r_{1j}'s, say the r_{1k}, is nonzero add the first column multiplied by $-q_{1k}$ to the kth column in order to replace the a_{1k} by the element r_{1k} with smaller degree. Now we can again reduce the upper left-hand corner element by locating there an element of A with least degree.

If all the r_{i1}'s and r_{1j}'s are zero we add the first row multiplied by $-q_{i1}$ to the ith row, $i = 1, 2, \ldots, l$, and the first column multiplied by $-q_{1j}$ to the jth column, $j = 1, 2, \ldots, m$, to bring A on the form

$$\begin{bmatrix} a_{11} & 0 & \ldots & 0 \\ 0 & a_{22}^{(1)} & \ldots & a_{2m}^{(1)} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & a_{l2}^{(1)} & \ldots & a_{lm}^{(1)} \end{bmatrix}. \tag{1.4}$$

Should not some $a_{ij}^{(1)}$ be divisible by a_{11} we add the respective column to the first column and again we can start reducing the degree of a_{11}.

After a finite number of steps we shall obtain the matrix

$$\begin{bmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_{22}^{(2)} & \dots & a_{2m}^{(2)} \\ \dots & \dots & \dots & \dots \\ 0 & a_{l2}^{(2)} & \dots & a_{lm}^{(2)} \end{bmatrix}$$

in which a_1 divides all the remaining elements. If at least one $a_{ij}^{(2)}$ is nonzero, we can repeat the whole process for the rows $i = 2, 3, \dots, l$ and the columns $j = 2, 3, \dots, m$ to bring A on the form

$$\begin{bmatrix} a_1 & 0 & 0 & \dots & 0 \\ 0 & a_2 & 0 & \dots & 0 \\ 0 & 0 & a_{33}^{(3)} & \dots & a_{3m}^{(3)} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & a_{l3}^{(3)} & \dots & a_{lm}^{(3)} \end{bmatrix}$$

where a_1 divides a_2 and a_2 divides all the remaining elements. Continuing the process we shall finally arrive at the matrix

$$\operatorname{diag}\{a_1, a_2, \dots, a_r, 0, \dots, 0\} .$$

Using (a) we can scale the coefficients of the invariant polynomials and make them monic.

Representing each step of the algorithm by the corresponding elementary matrix we get

$$E_l A E_r = \operatorname{diag}\{a_1, \dots, a_r, 0, \dots, 0\} ,$$

where E_l is the product of all left elementary transformations and E_r is the product of all right elementary transformations. As determinants of the elementary matrices are units of $\mathfrak{F}[z]$, the same is true of E_l and E_r. Consequently, $E_1 = E_l^{-1}$ and $E_2 = E_r^{-1}$.

Example 1.1. Compute a canonical representation for the matrix

$$A = \begin{bmatrix} -z + z^3 & 0 \\ 1 - z^3 & 1 - z^2 \\ -z^2 + z^3 + z^5 & z + z^4 \end{bmatrix}$$

over the field $\mathfrak{R}$.

First we put $1 - z^2$ to the upper left-hand corner by means of the elementary matrices

$$E_l^{(1)} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_r^{(2)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

The result is

$$\begin{bmatrix} 1 - z^2 & 1 - z^3 \\ 0 & -z + z^3 \\ z + z^4 & -z^2 + z^3 + z^5 \end{bmatrix}.$$

Since

$$z + z^4 = (-1 - z^2)(1 - z^2) + 1 + z\,,$$

we obtain

$$\begin{bmatrix} 1 - z^2 & 1 - z^3 \\ 0 & -z + z^3 \\ 1 + z & 1 \end{bmatrix}$$

by applying

$$E_l^{(3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 + z^2 & 0 & 1 \end{bmatrix}.$$

Now we make 1 into the upper left-hand corner element by using

$$E_l^{(4)} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \qquad E_r^{(5)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Thus we get

$$\begin{bmatrix} 1 & 1 + z \\ -z + z^3 & 0 \\ 1 - z^3 & 1 - z^2 \end{bmatrix}.$$

Since

$$\begin{aligned} -z + z^3 &= (-z + z^3)\,1 + 0\,, \\ 1 - z^3 &= (1 - z^3)\,1 \quad + 0\,, \\ 1 + z \ &= (1 + z)\,1 \quad\ + 0\,, \end{aligned}$$

the transformations

$$E_l^{(6)} = \begin{bmatrix} 1 & 0 & 0 \\ z - z^3 & 1 & 0 \\ -1 + z^3 & 0 & 1 \end{bmatrix}, \qquad E_r^{(7)} = \begin{bmatrix} 1 & -1 - z \\ 0 & 1 \end{bmatrix}$$

will bring A on the form

$$\begin{bmatrix} 1 & 0 \\ 0 & z + z^2 - z^3 - z^4 \\ 0 & -z - z^2 + z^3 + z^4 \end{bmatrix}.$$

As 1 divides the other elements, we can start operating exclusively on them. We have

$$-z - z^2 + z^3 + z^4 = (-1)(z + z^2 - z^3 - z^4) + 0$$

and hence

$$E_l^{(8)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

yields the final form

$$\begin{bmatrix} 1 & 0 \\ 0 & z + z^2 - z^3 - z^4 \\ 0 & 0 \end{bmatrix}$$

where

$$a_1 = 1 ,$$

$$a_2 = z + z^2 - z^3 - z^4 .$$

Therefore,

$$E_l = E_l^{(8)} E_l^{(6)} E_l^{(4)} E_l^{(3)} E_l^{(1)} = \begin{bmatrix} 0 & 1 + z^2 & 1 \\ 1 & z - z^5 & z - z^3 \\ 1 & z - z^2 + z^3 & -1 + z \end{bmatrix},$$

$$E_r = E_r^{(2)} E_r^{(5)} E_r^{(7)} = \begin{bmatrix} 0 & 1 \\ 1 & -1 - z \end{bmatrix}$$

and, in turn,

$$E_1 = E_l^{-1} = \begin{bmatrix} -z + z^3 & 1 & 0 \\ 1 - z^3 & -1 & 1 \\ -z^2 + z^3 + z^5 & 1 + z^2 & -1 - z^2 \end{bmatrix},$$

$$E_2 = E_r^{-1} = \begin{bmatrix} 1 + z & 1 \\ 1 & 0 \end{bmatrix}.$$

Example 1.2. Given

$$A = \begin{bmatrix} 1 - z & 0 \\ 0 & 1 + z \end{bmatrix}$$

over the field $\mathfrak{Q}$, find a canonical representation.

We are already given A in the form (1.4) but $1 - z$ does not divide $1 + z$. Therefore, we add the second column to the first one using

$$E_r^{(1)} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

to obtain

$$\begin{bmatrix} 1 - z & 0 \\ 1 + z & 1 + z \end{bmatrix}.$$

Since

$$1 + z = (-1)(1 - z) + 2 ,$$

we get

$$\begin{bmatrix} 1 - z & 0 \\ 2 & 1 + z \end{bmatrix}$$

on applying

$$E_l^{(2)} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Now

$$E_l^{(3)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

yields

$$\begin{bmatrix} 2 & 1 + z \\ 1 - z & 0 \end{bmatrix}$$

and, since

$$1 - z = (\tfrac{1}{2} - \tfrac{1}{2}z)\, 2 + 0\,,$$

$$1 + z = 2(\tfrac{1}{2} + \tfrac{1}{2}z) + 0\,,$$

the matrices

$$E_l^{(4)} = \begin{bmatrix} 1 & 0 \\ -\tfrac{1}{2} + \tfrac{1}{2}z & 1 \end{bmatrix}, \qquad E_r^{(5)} = \begin{bmatrix} 1 & -\tfrac{1}{2} - \tfrac{1}{2}z \\ 0 & 1 \end{bmatrix}$$

will successively bring A on the form

$$\begin{bmatrix} 2 & 0 \\ 0 & -\tfrac{1}{2} + \tfrac{1}{2}z^2 \end{bmatrix}.$$

Using

$$E_r^{(6)} = \begin{bmatrix} \tfrac{1}{2} & 0 \\ 0 & 2 \end{bmatrix}$$

we can scale the leading coefficients of the invariant polynomials to unity. Therefore,

$$a_1 = 1\,,$$

$$a_2 = z^2 - 1$$

and

$$E_l = E_l^{(4)} E_l^{(3)} E_l^{(2)} = \begin{bmatrix} 1 & 1 \\ \tfrac{1}{2} + \tfrac{1}{2}z & -\tfrac{1}{2} + \tfrac{1}{2}z \end{bmatrix},$$

$$E_r = E_r^{(1)} E_r^{(5)} E_r^{(6)} = \begin{bmatrix} \tfrac{1}{2} & -1 - z \\ \tfrac{1}{2} & 1 - z \end{bmatrix}.$$

It follows that

$$E_l = \begin{bmatrix} \tfrac{1}{2} - \tfrac{1}{2}z & 1 \\ \tfrac{1}{2} + \tfrac{1}{2}z & -1 \end{bmatrix},$$

$$E_2 = \begin{bmatrix} 1 - z & 1 + z \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

In view of the invariant factor theorem a matrix $A \in \mathfrak{F}_{m,m}[z]$ is a unit of $\mathfrak{F}_{m,m}[z]$ if and only if all invariant polynomials of A are units of $\mathfrak{F}[z]$, and the A is a divisor of zero in $\mathfrak{F}_{m,m}[z]$ if and only if rank $A < m$.

Further matrices $A, B \in \mathfrak{F}_{l,m}[z]$ are associates if and only if the invariant polynomials a_k of A and b_k of B are associates in $\mathfrak{F}[z]$, $k = 1, 2, \ldots$. Otherwise speaking, $a_k = b_k$, $k = 1, 2, \ldots$ modulo units of $\mathfrak{F}[z]$.

Given matrices $A \in \mathfrak{F}_{l,m}[z]$ and $B \in \mathfrak{F}_{p,q}[z]$, we have $B \mid A$ if and only if $b_k \mid a_k$, $k = 1, 2, \ldots$ where a_k and b_k are invariant polynomials of A and B respectively [22].

A matrix $D_1 \in \mathfrak{F}_{l,l}[z]$ is a greatest common left divisor of $A \in \mathfrak{F}_{l,m}[z]$ and $B \in \in \mathfrak{F}_{l,q}[z]$ if and only if the matrices

$$[A \; B] \quad \text{and} \quad [D_1 \; 0]$$

are right associates; a matrix $D_2 \in \mathfrak{F}_{m,m}[z]$ is a greatest common right divisor of $A \in \mathfrak{F}_{l,m}[z]$ and $B \in \mathfrak{F}_{p,m}[z]$ if and only if the matrices

$$\begin{bmatrix} A \\ B \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} D_2 \\ 0 \end{bmatrix}$$

are left associates. A matrix $D \in \mathfrak{F}_{m,m}[z]$ is a greatest common divisor of $A, B \in \in \mathfrak{F}_{m,m}[z]$ if and only if the matrices

$$\operatorname{diag}\{(a_1, b_1), (a_2, b_2), \ldots, (a_m, b_m)\} \quad \text{and} \quad D$$

are associates.

If follows that matrices $A \in \mathfrak{F}_{l,m}[z]$ and $B \in \mathfrak{F}_{l,q}[z]$ are left coprime if and only if the matrices

$$[A \; B] \quad \text{and} \quad [I_l \; 0]$$

are associates, matrices $A \in \mathfrak{F}_{l,m}[z]$ and $B \in \mathfrak{F}_{p,m}[z]$ are right coprime if and only if the matrices

$$\begin{bmatrix} A \\ B \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} I_m \\ 0 \end{bmatrix}$$

are associates, and matrices $A, B \in \mathfrak{F}_{m,m}[z]$ are coprime if and only if a_k and b_k are coprime in $\mathfrak{F}[z]$.

At last, given a polynomial $b \in \mathfrak{F}[z]$ and a matrix $A \in \mathfrak{F}_{l,m}[z]$, we have $b \mid A$ if and only if $b \mid a_1$. It follows that $(b, A) = (b, a_1)$.

Given a field $\mathfrak{F}$, we can also consider the set $\mathfrak{F}_{l,m}(z^{-1})$ of $l \times m$ matrices over $\mathfrak{F}(z^{-1})$. They are called *rational matrices* and they can be written in the form

$$A = \frac{Q}{p}, \quad Q \in \mathfrak{F}_{l,m}[z], \quad p \in \mathfrak{F}[z] - \{0\}$$

or in the form

$$A = A_n z^{-n} + A_{n+1} z^{-(n+1)} + \dots, \quad A_k \in \mathfrak{F}_{l,m}, \quad n \in \mathfrak{Z}.$$

If $A_n \neq 0$ then n is called the *order* of A, denoted as $\mathcal{O}A$. In fact, $\mathcal{O}A = \partial p - \partial Q$. The $\mathfrak{F}_{l,m}(z^{-1})$ is not a field and it is not even a ring unless $l = m$. A matrix A is a unit of the ring $\mathfrak{F}_{m,m}(z^{-1})$ if and only if $\det A \neq 0$ and it is a divisor of zero in $\mathfrak{F}_{m,m}(z^{-1})$ if and only if $\det A = 0$.

The set of $l \times m$ matrices over $\mathfrak{F}\{z^{-1}\}$ will be denoted by $\mathfrak{F}_{l,m}\{z^{-1}\}$ and its elements can be written in the form

$$A = \frac{Q}{p}, \quad \partial Q \leqq \partial p$$

or in the form

$$A = A_0 + A_1 z^{-1} + A_2 z^{-2} + \dots, \quad A_k \in \mathfrak{F}_{l,m}.$$

The $\mathfrak{F}_{l,m}\{z^{-1}\}$ is sometimes called the set of *realizable rational matrices* and it is not a ring unless $l = m$. A matrix

$$A = A_0 + A_1 z^{-1} + \dots \in \mathfrak{F}_{m,m}\{z^{-1}\}$$

is a unit in $\mathfrak{F}_{m,m}\{z^{-1}\}$ if and only if A_0 is a unit in $\mathfrak{F}_{m,m}$.

Similarly $\mathfrak{F}^+_{l,m}\{z^{-1}\}$ denotes the set of $l \times m$ matrices over $\mathfrak{F}^+\{z^{-1}\}$, called *stable realizable rational matrices*. They can again be written in the form

$$A = \frac{Q}{p}, \quad \partial Q \leqq \partial p, \quad p \text{ stable}.$$

The $\mathfrak{F}^+_{l,m}\{z^{-1}\}$ is not a ring unless $l = m$. An element

$$A = \frac{Q}{p} = A_0 + A_1 z^{-1} + \dots \in \mathfrak{F}^+_{m,m}\{z^{-1}\}$$

is a unit in $\mathfrak{F}^+_{m,m}\{z^{-1}\}$ if and only if A_0 is a unit in $\mathfrak{F}_{m,m}$ and $\det Q$ is a stable polynomial.

Of course, $\mathfrak{F}_{m,m}[z]$ is a subring of $\mathfrak{F}^+_{m,m}\{z^{-1}\}$ and $\mathfrak{F}^+_{m,m}\{z^{-1}\}$ is a subring of $\mathfrak{F}_{m,m}\{z^{-1}\}$ which, in turn, is a subring of $\mathfrak{F}_{m,m}(z^{-1})$.

It is to be emphasized that we regard a polynomial or rational matrix as an algebraic object, not as a function of a complex variable z or z^{-1}. They are simply an alternate way of viewing finite or infinite sequences in $\mathfrak{F}_{l,m}$, the indeterminate z or z^{-1} being just a position-marker.

1.4. Matrix linear Diophantine Equations

Consider the equation

$$AX + YB = C\,, \tag{1.5}$$

where $A \in \mathfrak{F}_{l,p}[z]$, $B \in \mathfrak{F}_{q,m}[z]$, $C \in \mathfrak{F}_{l,m}[z]$ are given matrices and $X \in \mathfrak{F}_{p,m}[z]$, $Y \in \mathfrak{F}_{l,q}[z]$ are unknown matrices. By analogy with a similar equation in integers [13, 42, 44] or polynomials [30, 35] we shall call (1.5) a *matrix linear Diophantine equation.*

Any pair X, Y satisfying (1.5) will be called a solution. It is not a simple task to obtain a solution of (1.5), in general. The approach presented here is based on elementary transformations [50], thereby converting equation (1.5) to a set of linear Diophantine equations over $\mathfrak{F}[z]$.

Theorem 1.1. *Equation* (1.5) *has a solution if and only if the matrices*

$$\begin{bmatrix} A & C \\ 0 & B \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \tag{1.6}$$

are associates in $\mathfrak{F}_{l+q,p+m}[z]$.

Proof. Necessity: If X, Y is a solution of (1.5), we have

$$\begin{bmatrix} A & C \\ 0 & B \end{bmatrix} = \begin{bmatrix} A & AX + YB \\ 0 & B \end{bmatrix} = \begin{bmatrix} I_l & Y \\ 0 & I_q \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} I_p & X \\ 0 & I_m \end{bmatrix}.$$

Since

$$\det \begin{bmatrix} I_l & Y \\ 0 & I_q \end{bmatrix} = 1\,, \quad \det \begin{bmatrix} I_p & X \\ 0 & I_m \end{bmatrix} = 1\,,$$

matrices (1.6) are associates in $\mathfrak{F}_{l+q,p+m}[z]$.

Sufficiency: Assuming that matrices (1.6) are associates we are to prove the existence of a solution for (1.5).

Let

$$\begin{aligned} A &= E_{1A} \operatorname{diag}\{a_1, a_2, \ldots, a_r, 0, \ldots, 0\}\, E_{2A}\,, \\ B &= E_{1B} \operatorname{diag}\{b_1, b_2, \ldots, b_s, 0, \ldots, 0\}\, E_{2B} \end{aligned} \tag{1.7}$$

be the canonical representations. Then equation (1.5) is equivalent to the equation

$$\operatorname{diag}\{a_1, \ldots, a_r, 0, \ldots, 0\}\, \bar{X} + \bar{Y} \operatorname{diag}\{b_1, \ldots, b_s, 0, \ldots, 0\} = \bar{C}\,, \tag{1.8}$$

where

$$\begin{aligned} X &= E_{2A}^{-1} \bar{X} E_{2B}\,, \quad Y = E_{1A} \bar{Y} E_{1B}^{-1}\,, \\ C &= E_{1A} \bar{C} E_{2B}\,, \end{aligned} \tag{1.9}$$

and also denote

$$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} = \begin{bmatrix} E_{1A} & 0 \\ 0 & E_{1B} \end{bmatrix} M \begin{bmatrix} E_{2A} & 0 \\ 0 & E_{2B} \end{bmatrix},$$

$$\begin{bmatrix} A & C \\ 0 & B \end{bmatrix} = \begin{bmatrix} E_{1A} & 0 \\ 0 & E_{1B} \end{bmatrix} N \begin{bmatrix} E_{2A} & 0 \\ 0 & E_{2B} \end{bmatrix},$$

where

$$M = \begin{bmatrix} \operatorname{diag}\{a_1, \ldots, a_r, 0, \ldots, 0\} & 0 \\ 0 & \operatorname{diag}\{b_1, \ldots, b_s, 0, \ldots, 0\} \end{bmatrix},$$

$$N = \begin{bmatrix} \operatorname{diag}\{a_1, \ldots, a_r, 0, \ldots, 0\} & \bar{C} \\ 0 & \operatorname{diag}\{b_1, \ldots, b_s, 0, \ldots, 0\} \end{bmatrix}.$$

The M and N are associates by hypothesis.

Denote $\bar{\bar{c}}_{ij}$ the elements of $\bar{\bar{C}}$. In view of equation (1.8) we are to prove that elements $\bar{\bar{x}}_{ij}$ of $\bar{\bar{X}}$ and $\bar{\bar{y}}_{ij}$ of $\bar{\bar{Y}}$ exist in $\mathfrak{F}[z]$ and satisfy the equations

$$a_i\bar{\bar{x}}_{ij} + \bar{\bar{y}}_{ij}b_j = \bar{\bar{c}}_{ij} \tag{1.10}$$

for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, s$; the equations

$$a_i\bar{\bar{x}}_{ij} = \bar{\bar{c}}_{ij} \tag{1.11}$$

for $i = 1, 2, \ldots, r$ and $j = s + 1, \ldots, m$; the equations

$$\bar{\bar{y}}_{ij}b_j = \bar{\bar{c}}_{ij} \tag{1.12}$$

for $i = r + 1, \ldots, l$ and $j = 1, 2, \ldots, s$; and that the remaining elements $\bar{\bar{c}}_{ij}$ for $i = r + 1, \ldots, l$ and $j = s + 1, \ldots, m$ are identically zero.

To prove the existence of $\bar{\bar{x}}_{ij}$ and $\bar{\bar{y}}_{ij}$ in (1.10) we shall show that $(a_i, b_j) \mid \bar{\bar{c}}_{ij}$. Indeed, let w be any polynomial prime in $\mathfrak{F}[z]$ which is common to a_r and b_s. Then the invariant polynomials of A and B are

$$a_i = w^{f_i}a_i', \quad 0 \leqq f_1 \leqq f_2 \leqq \ldots \leqq f_r,$$

$$b_j = w^{g_j}b_j', \quad 0 \leqq g_1 \leqq g_2 \leqq \ldots \leqq g_s,$$

respectively, where $a_i', b_j' \in \mathfrak{F}[z]$ are coprime with w. Consequently the invariant polynomials of M (of N) are

$$m_k = w^{h_k}m_k', \quad 0 \leqq h_1 \leqq h_2 \leqq \ldots \leqq h_{r+s},$$

where h_k's is a permutation of the exponents f_i and g_j in nondescending order, and where $m_k' \in \mathfrak{F}[z]$ is coprime with w.

Further let d_k stand for a greatest common divisor of all $k \times k$ minors of M (and of N). Then $d_{i+j-1} = \prod_{k=1}^{i+j-1} w^{h_k} m'_k$. It contains the factor w^{f_i} of a_i or the factor w^{g_j} of b_j for in forming the sequence $\{h_1, h_2, \ldots, h_{i+j-1}\}$ either all $f_1, f_2, \ldots, f_i$ or all $g_1, g_2, \ldots, g_j$ must be taken in order to get its $i + j - 1$ terms. On the other hand, both w^{f_i} and w^{g_j} cannot be factors of d_{i+j-1} for both f_i and g_j cannot occur in the sequence $\{h_1, h_2, \ldots, h_{i+j-1}\}$ which has only $i + j - 1$ terms. Hence h_{ij}, the lesser of f_i and g_j, is in the sequence and d_{i+j-1} must contain $w^{h_{ij}}$.

Now let M^0 and N^0 be the matrices obtained from M and N respectively by deleting their ith and $(l + j)$th rows and ith and $(p + j)$th columns, and let d_k^0 be their greatest common divisor of all minors of order k. Note that d_{i+j-2}^0 will not contain $w^{h_{ij}}$ because the rows and columns containing a_i and b_j were deleted in forming M^0 and N^0 but it will contain as factors all the remaining powers of w that occur in d_{i+j-1}. Now $\bar{c}_{ij} d_{i+j-2}^0$ is a minor of order $i + j - 1$ of N and as a consequence it is a multiple of $\prod_{k=1}^{i+j-1} w^{h_k}$, a factor of d_{i+j-1}. Therefore $\bar{c}_{ij}$ must be a multiple of $w^{h_{ij}}$, the highest power of w which is common to a_i and b_j. Since w is *any* polynomial prime in $\mathfrak{F}[z]$ and common to a_r and b_s, it follows that $(a_i, b_j) \mid \bar{c}_{ij}$.

To prove the existence of $\bar{x}_{ij}$ in (1.11) we shall show that $a_i \mid \bar{c}_{ij}$. Indeed, $d_{r+s} = \prod_{k=1}^{r} a_k \prod_{t=1}^{s} b_t$ must divide $\bar{c}_{ij} \prod_{k=1}^{i-1} a_k \prod_{k=i+1}^{r} a_k \prod_{t=1}^{s} b_t$, a minor of order $r + s$ of N. It follows that $a_i \mid \bar{c}_{ij}$.

To prove the existence of $\bar{y}_{ij}$ in (1.12) we shall show that $b_j \mid \bar{c}_{ij}$. Indeed, $d_{r+s} = \prod_{k=1}^{r} a_k \prod_{t=1}^{s} b_t$ must divide $\bar{c}_{ij} \prod_{k=1}^{r} a_k \prod_{t=1}^{j-1} b_t \prod_{t=j+1}^{s} b_t$, another minor of order $r + s$ of N. It follows that $b_j \mid \bar{c}_{ij}$.

Finally, if any $\bar{c}_{ij}$, $i = r + 1, \ldots, l$ and $j = s + 1, \ldots, m$ were not identically zero, N would have the nonzero minor $\bar{c}_{ij} \prod_{k=1}^{r} a_k \prod_{t=1}^{s} b_t$ of order $r + s + 1$, which is impossible. Hence these $\bar{c}_{ij}$ are zero and the remaining elements $\bar{x}_{ij}$ and $\bar{y}_{ij}$, which are not coupled by equation (1.8), may be chosen arbitrarily within $\mathfrak{F}[z]$.

The existence of a solution X, Y to (1.5) has been proved. □

As a consequence of Theorem 1.1, a particular solution of equation (1.5) reads

$$(1.13) \qquad X_0 = E_{2A}^{-1} \begin{bmatrix} \bar{X}_{0,11} & \bar{X}_{0,12} \\ 0 & 0 \end{bmatrix} E_{2B} \in \mathfrak{F}_{p,m}[z] ,$$

$$Y_0 = E_{1A} \begin{bmatrix} \bar{Y}_{0,11} & 0 \\ \bar{Y}_{0,21} & 0 \end{bmatrix} E_{1B}^{-1} \in \mathfrak{F}_{l,q}[z] ,$$

where the elements $\bar{x}_{0,ij}$ of $\bar{X}_{0,11} \in \mathfrak{F}_{r,s}[z]$ and the elements $\bar{y}_{0,ij}$ of $\bar{Y}_{0,11} \in \mathfrak{F}_{r,s}[z]$ are particular solutions of (1.10), the elements $\bar{x}_{0,ij}$ of $\bar{X}_{0,12} \in \mathfrak{F}_{r,m-s}[z]$ are particular solutions of (1.11), and the elements $\bar{\bar{y}}_{0,ij}$ of $\bar{\bar{Y}}_{0,21} \in \mathfrak{F}_{l-r,s}[z]$ are particular solutions of (1.12).

An effective method of finding $\bar{x}_{0,ij}$ and $\bar{y}_{0,ij}$ is presented in [30].

Theorem 1.2. *If X_0, Y_0 is a particular solution of equation* (1.5) *then all solutions are of the form*

$$X = X_0 + E_{2A}^{-1} T E_{2B}, \qquad (1.14)$$
$$Y = Y_0 - E_{1A} S E_{1B}^{-1},$$

where

$$T = \begin{bmatrix} T_{11} & 0 \\ T_{21} & T_{22} \end{bmatrix}, \qquad S = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix}.$$

The elements of $T_{11} \in \mathfrak{F}_{r,s}[z]$ are $t_{ij}[b_j/(a_i, b_j)]$ and the elements of $S_{11} \in \mathfrak{F}_{r,s}[z]$ are $[a_i/(a_i, b_j)]\, t_{ij}$, where t_{ij} is an arbitrary polynomial of $\mathfrak{F}[z]$. The matrices $T_{21} \in \mathfrak{F}_{p-r,s}[z]$, $T_{22} \in \mathfrak{F}_{p-r,m-s}[z]$ and $S_{12} \in \mathfrak{F}_{r,q-s}[z]$, $S_{22} \in \mathfrak{F}_{l-r,q-s}[z]$ are arbitrary polynomial matrices.

Proof. Equation (1.5) can be converted into the set of equations (1.10), (1.11) and (1.12). The general solutions of (1.10) read

$$\bar{x}_{ij} = \bar{x}_{0,ij} + t_{ij} \frac{b_j}{(a_i, b_j)},$$

$$\bar{y}_{ij} = \bar{y}_{0,ij} - \frac{a_i}{(a_i, b_j)} t_{ij},$$

where t_{ij} are arbitrary elements of $\mathfrak{F}[z]$, the general solutions of (1.11) read

$$\bar{\bar{x}}_{ij} = \bar{\bar{x}}_{0,ij} = \frac{\bar{c}_{ij}}{a_i},$$

and the general solutions of (1.12) become

$$\bar{\bar{y}}_{ij} = \bar{\bar{y}}_{0,ij} = \frac{\bar{c}_{ij}}{b_j}.$$

The remaining elements $\bar{x}_{ij}$ and $\bar{y}_{ij}$, if any, can be chosen arbitrarily within $\mathfrak{F}[z]$.

Hence our claim follows by virtue of (1.9). □

To illustrate how the above theorems work we compute several examples.

Example 1.3. Consider equation (1.5) over the field $\mathfrak{R}$, where

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & z-1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \qquad B = [0 \quad z-1], \qquad C = \begin{bmatrix} 1 & 0 \\ 0 & z-1 \\ 1 & z-1 \end{bmatrix}.$$

The canonical representations of A and B are

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & z-1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad r = 2,$$

$$B = \begin{bmatrix} z-1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad s = 1.$$

Hence we set

$$X = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \bar{X} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad Y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \bar{Y}$$

to obtain the set of equations

(1.15)
$$\bar{x}_{11} + \bar{y}_{11}(z-1) = 0, \qquad \bar{x}_{12} = 1,$$
$$(z-1)\,\bar{x}_{21} + \bar{y}_{21}(z-1) = z - 1, \qquad (z-1)\,\bar{x}_{22} = 0.$$
$$\bar{y}_{31}(z-1) = z - 1,$$

It is seen that equations (1.15) are solvable, which is equivalent to saying that matrices (1.6) are associates, and hence our equation has a solution.

By solving (1.15) we obtain

$$\bar{\bar{X}} = \begin{bmatrix} 0 + (z-1)\,t_{11} & 1 \\ 1 + t_{21} & 0 \\ t_{31} & t_{32} \end{bmatrix}, \qquad \bar{\bar{Y}} = \begin{bmatrix} 0 - t_{11} \\ 0 - t_{21} \\ 1 \end{bmatrix},$$

where $t_{11}, t_{21} \in \Re[z]$ are arbitrary polynomials generating the general solutions of (1.15) and $t_{31}, t_{32} \in \Re[z]$ are arbitrary polynomials which do not appear in (1.15).

Hence

$$X_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad Y_0 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

by (1.13) and since

$$T = \begin{bmatrix} (z-1)\,t_{11} & 0 \\ t_{21} & 0 \\ t_{31} & t_{32} \end{bmatrix}, \qquad S = \begin{bmatrix} t_{11} \\ t_{21} \\ 0 \end{bmatrix},$$

we end up with the general solution

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -t_{32} & (z-1)\,t_{11} - t_{31} \\ 0 & t_{21} \\ t_{32} & t_{31} \end{bmatrix},$$

$$Y = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} t_{11} \\ t_{21} \\ t_{11} \end{bmatrix}.$$

In synthesizing the optimal systems we shall frequently encounter a special case of equation (1.5), viz.

$$AX + Yb = C, \tag{1.16}$$

where $A \in \mathfrak{F}_{l,p}[z]$, $b \in \mathfrak{F}[z]$, $C \in \mathfrak{F}_{l,1}[z]$ and $X \in \mathfrak{F}_{p,1}[z]$, $Y \in \mathfrak{F}_{l,1}[z]$. That is $m = q = 1$. We write

$$A = E_1 \operatorname{diag}\{a_1, \ldots, a_r, 0, \ldots, 0\} E_2 .$$

Then equation (1.16) is equivalent to the equation

$$\operatorname{diag}\{a_1, \ldots, a_r, 0, \ldots, 0\} \overline{X} + \overline{Y}b = \overline{C},$$

where

$$X = E_2^{-1}\overline{X}, \quad Y = E_1\overline{Y}, \quad C = E_1\overline{C},$$

and which, in turn, yields the set of polynomial equations

$$a_i\bar{x}_i + \bar{y}_i b = \bar{c}_i, \quad i = 1, 2, \ldots, r, \tag{1.17}$$

and

$$\bar{y}_i b = \bar{c}_i, \quad i = r + 1, \ldots, l. \tag{1.18}$$

A particular solution X_0, Y_0 of (1.16) is then obtained as

$$X_0 = E_2^{-1} \begin{bmatrix} \overline{X}_{0,1} \\ 0 \end{bmatrix},$$

$$Y_0 = E_1 \begin{bmatrix} \overline{Y}_{0,1} \\ \overline{Y}_{0,2} \end{bmatrix},$$

where the elements $\bar{x}_{0,i}$ of $\overline{X}_{0,1} \in \mathfrak{F}_{r,1}[z]$ and the elements $\bar{y}_{0,i}$ of $\overline{Y}_{0,1} \in \mathfrak{F}_{r,1}[z]$ are particular solutions of (1.17), and the elements $\bar{y}_{0,i}$ of $\overline{Y}_{0,2} \in \mathfrak{F}_{l-r,1}[z]$ are particular solutions of (1.18).

The general solution of (1.16) then reads

$$X = X_0 + D^{-1}Tb, \tag{1.19}$$

$$Y = Y_0 - AD^{-1}T,$$

where

(1.20) $$D = \operatorname{diag}\{(a_1, b), \ldots, (a_r, b), (0, b), \ldots, (0, b)\}\, E_2 \in \mathfrak{F}_{p,p}[z]$$

and T is an arbitrary matrix of $\mathfrak{F}_{p,1}[z]$.

In applications we often seek for a particular solution X^0, Y^0 of (1.16) such that the degree of one matrix polynomial, say Y^0, is *minimal*.

Unfortunately, no general algorithm to solve for X^0, Y^0 is available at present. If, however, the AD^{-1} is a regular matrix polynomial, we can use the left division algorith. Write

$$Y = Y_0 - AD^{-1}T\,,$$

$$Y_0 = AD^{-1}Q_0 + R_0\,, \quad \partial R_0 < \partial AD^{-1}\,.$$

Then

$$Y = R_0 + AD^{-1}(Q_0 - T)$$

and evidently,

$$Y^0 = R_0\,, \quad X^0 = X_0 + D^{-1}Q_0 b$$

is uniquely determined on setting $T = Q_0$. The solution is not unique is general, however.

Sometimes the solution X^0, Y^0 satisfying $\partial Y^0 < \partial A$ is useful. The two solutions are identical whenever D is a unit of $\mathfrak{F}_{p,p}[z]$.

Example 1.4. We are to solve equation (1.16), where

$$A = \begin{bmatrix} z^2 & z & z^2 \\ z & z & z \end{bmatrix}, \quad b = [1]\,, \quad C = \begin{bmatrix} z^2 + z - 1 \\ z + 1 \end{bmatrix}$$

are matrices over $\mathfrak{R}[z]$, for a solution X^0, Y^0 such that $\partial Y^0 = \min$.

The canonical representation of A is

$$A = \begin{bmatrix} z & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} z & 0 & 0 \\ 0 & z(z-1) & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and since matrices (1.6) are associates our equation has a solution. Setting

$$X = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \bar{X}\,, \quad Y = \begin{bmatrix} z & -1 \\ 1 & 0 \end{bmatrix} \bar{Y}\,,$$

we are to solve equations

(1.21) $$z\,\bar{x}_1 + \bar{y}_1 = z + 1\,,$$

$$z(z-1)\,\bar{x}_2 + \bar{y}_2 = 1$$

obtaining

$$\bar{X} = \begin{bmatrix} 1 + t_1 \\ 0 + t_2 \\ t_3 \end{bmatrix}, \quad \bar{Y} = \begin{bmatrix} 1 + zt_1 \\ 1 + z(z - 1)\, t_2 \end{bmatrix}.$$

Here $t_1, t_2 \in \Re[z]$ are arbitrary polynomials generating the general solutions of (1.21) and $t_3 \in \Re[z]$ is arbitrary due to its absence in (1.21).

Further

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and, therefore, the general solution reads

$$X = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}$$

$$Y = \begin{bmatrix} z - 1 \\ 0 \end{bmatrix} - \begin{bmatrix} z^2 & z & z^2 \\ z & z & z \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}$$

on using (1.19).

The solution X^0, Y^0 satisfying $\partial Y^0 = \min$ (and also $\partial Y^0 < \partial A$) is obtained by an appropriate choice of t_1 and t_2. It is seen that any $t_1 = t_2 = \tau_0 \in \Re$ will yield $\partial Y^0 = 1$ regardless of t_3 and no other choice will give Y^0 of less degree. Therefore

$$X^0 = \begin{bmatrix} 1 - t_3 \\ \tau_0 \\ t_3 \end{bmatrix}, \quad Y^0 = \begin{bmatrix} -1 + (1 - \tau_0)\, z \\ 1 - \tau_0 z \end{bmatrix}.$$

Note that this solution is not unique and that it cannot be obtain via the left division algorithm.

As far as the linear Diophantine equation

$$(1.22) \qquad ax + by = c$$

over $\mathfrak{F}[z]$ is concerned, it may be thought a special case of (1.5) for $l = m = p = q = 1$. Then the solvability condition of Theorem 1.1 reduces to the condition $(a, b) \mid c$, derived in [30]. Indeed, in this case

$$M = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}, \quad N = \begin{bmatrix} a & c \\ 0 & b \end{bmatrix}$$

and the greatest common divisors d_k^M and d_k^N of all $k \times k$ minors of M and N respectively are

$$d_0^{\mathrm{M}} = 1\,, \qquad d_0^{\mathrm{N}} = 1\,,$$
$$d_1^{\mathrm{M}} = (a, b)\,, \quad d_1^{\mathrm{N}} = ((a, b), c)\,,$$
$$d_2^{\mathrm{M}} = ab\,, \qquad d_2^{\mathrm{N}} = ab\,.$$

The d_k^{M} and d_k^{N}, $k = 1, 2$ are associates if and only if $(a, b) = ((a, b), c)$ i.e. when $(a, b) \mid c$.

Also the general solution of (1.22) can be obtained, for instance, via (1.19). In this case $D = (a, b)$ and (1.19) becomes

$$x = x_0 + \frac{b}{(a, b)}\,t\,,$$
$$y = y_0 - \frac{a}{(a, b)}\,t\,.$$

If the matrices appearing in (1.5) are not matrices over $\mathfrak{F}[z]$ but, more generally, over $\mathfrak{F}\{z^{-1}\}$ or $\mathfrak{F}^+\{z^{-1}\}$, all results remain valid with $\mathfrak{F}[z]$ replaced by $\mathfrak{F}\{z^{-1}\}$ or $\mathfrak{F}^+\{z^{-1}\}$. Consider, for instance, the following example.

Example 1.5. Find the general solution of equation (1.22) where

$$a = 1 - 0{\cdot}5z^{-1}\,, \quad b = z^{-1} - 0{\cdot}5z^{-2}\,, \quad c = 1$$

are elements of $\mathfrak{F}^+\{z^{-1}\}$.

First of all $(a, b) = 1 - 0{\cdot}5z^{-1}$ which is a unit of $\mathfrak{F}^+\{z^{-1}\}$ and as such it divides c. As a result, our equation has a solution.

A particular solution is evidently

$$x_0 = \frac{1}{1 - 0{\cdot}5z^{-1}}\,,$$
$$y_0 = 0$$

and, therefore, the general solution becomes

$$x = \frac{1}{1 - 0{\cdot}5z^{-1}} + z^{-1}t\,,$$
$$y = -t$$

for arbitrary $t \in \mathfrak{F}^+\{z^{-1}\}$.

2. SYSTEMS

2.1. System Description

For a rigorous treatment of optimal control problems we cannot do with the intuitive engineering notion of a system. We shall give an axiomatic definition of the class of systems to be dealt with in the paper. It is understood that in this way we define a mathematical model of a physical system and, therefore, the optimal control theory deals with such models rather than with the actual physical systems [19; 23].

Following [22; 23; 30], let

$\mathscr{T}$ = time set = $\mathfrak{Z}$ = (ordered) set of integers,
$\mathscr{U}$ = input values = $\mathfrak{F}^m$ = vector space of m-tuples over a field $\mathfrak{F}$,
$\mathscr{Y}$ = output values = $\mathfrak{F}^l$,
$\mathscr{X}$ = state space = $\mathfrak{F}^n$.

Then a *finite dimensional, discrete, constant, linear, m-input, l-output system* $\mathscr{S}$ *over a field* $\mathfrak{F}$ is a quadruple $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ of linear maps

$$\mathbf{A} : \mathscr{X} \to \mathscr{X} ,$$

$$\mathbf{B} : \mathscr{U} \to \mathscr{X} ,$$

$$\mathbf{C} : \mathscr{X} \to \mathscr{Y} ,$$

$$\mathbf{D} : \mathscr{U} \to \mathscr{Y} .$$

The n is dimension of $\mathscr{S}$.

We think of the quadruple $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ as defining the dynamical equations

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k ,$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k ,$$

where $k \in \mathfrak{Z}$, $\mathbf{x} \in \mathscr{X}$, $\mathbf{u} \in \mathscr{U}$, and $\mathbf{y} \in \mathscr{Y}$.

We shall usually not make a distinction between $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ as linear maps or as matrices representing these maps with respect to a given basis.

The definition covers a fairly large class of systems. In particular, if $\mathfrak{F} = \mathfrak{R}$ we have a real sampled-data or intrinsically discrete linear system. If $\mathfrak{F} = \mathfrak{Z}_p$ or an algebraic extension of $\mathfrak{Z}_p$, we have a linear finite automaton. If $l = m = 1$ we have a single-input single-output system.

The above definition gives the *internal* description of $\mathscr{S}$, it is not confined to the external behavior of the system.

The matrix

$$\boldsymbol{S} = \mathbf{C}(z\mathbf{I}_n - \mathbf{A})^{-1}\,\mathbf{B} + \mathbf{D} \in \mathfrak{F}_{l,m}\{z^{-1}\} \tag{2.1}$$

is called the *impulse response matrix of* $\mathscr{S}$. It reflects just the input-output properties of the system.

Conversely, any quadruple $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ satisfying (2.1) is called a *realization* of $\boldsymbol{S}$. Apparently, there are many realizations of $\boldsymbol{S}$. A realization $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ is said to be completely reachable and completely observable, or equivalently, to be a *minimal realization* of $\boldsymbol{S}$, if

$$\text{(2.2)} \qquad \operatorname{rank}[\mathbf{B}\ \mathbf{AB} \ldots \mathbf{A}^{n-1}\mathbf{B}] = n\,,$$

$$\operatorname{rank}\begin{bmatrix}\mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{n-1}\end{bmatrix} = n\,.$$

The minimal realization has least dimension among all realizations of $\boldsymbol{S}$ and it is unique up to a coordinatization of its state space, see [20; 23]. We recall that there is a one-to-one correspondence between $\boldsymbol{S}$ and its realization if and only if the realization is minimal. Otherwise speaking, nonminimal realizations contain certain parts which have no relation to $\boldsymbol{S}$. That is why the notion of minimal realization is of fundamental importance in the mathematical system theory.

The monic polynomials $\det(z\mathbf{I}_n - \mathbf{A}) \in \mathfrak{F}[z]$ and $\operatorname{mp}(z\mathbf{I}_n - \mathbf{A}) \in \mathfrak{F}[z]$ are called the *characteristic* and *minimal* polynomials of $\mathscr{S}$, respectively. We have $n = \partial \det(z\mathbf{I}_n - \mathbf{A})$. Further, the $\mathscr{S}$ is defined to be *stable* if $\det(z\mathbf{I}_n - \mathbf{A})$ is stable, or equivalently, if $\operatorname{mp}(z\mathbf{I}_n - \mathbf{A})$ is stable.

Defining

$$\Sigma_0 = \mathbf{D}\,,$$

$$\Sigma_k = \mathbf{CA}^{k-1}\mathbf{B}\,, \quad k = 1, 2, \ldots\,,$$

we can write

$$\boldsymbol{S} = \Sigma_0 + \Sigma_1 z^{-1} + \Sigma_2 z^{-2} + \ldots\,, \quad \Sigma_k \in \mathfrak{F}_{l,m}\,.$$

This is just another way of writing the impulse response matrix. The order of $\boldsymbol{S}$ can be interpreted as the *discrete-time delay* of $\mathscr{S}$.

Since $\boldsymbol{S} \in \mathfrak{F}_{l,m}\{z^{-1}\}$, it can be written as the ratio

$$\text{(2.3)} \qquad \boldsymbol{S} = \frac{\hat{B}}{\hat{a}}\,,$$

where $\hat{B} \in \mathfrak{F}_{l,m}[z]$ and $\hat{a} \in \mathfrak{F}[z]$ satisfy

$$(\hat{a}, \hat{B}) = 1\,,$$

$$\partial\hat{B} \leqq \partial\hat{a}$$

and where $\hat{a}$ is the least common denominator of all elements of $\boldsymbol{S}$.

The expression (2.3), however, tells very little about the structure and dynamical behavior of $\mathscr{S}$. We have to refine (2.3) as follows. Let

$$\hat{B} = \hat{E}_1 \operatorname{diag}\{\hat{g}_1, \hat{g}_2, \ldots, \hat{g}_r, 0, \ldots, 0\}\, \hat{E}_2$$

be the canonical representation of $\hat{B}$ and let

$$\frac{\hat{g}_k}{\hat{a}} = \frac{\hat{b}_k}{\hat{a}_k}, \quad k = 1, 2, \ldots, r,$$

after cancelling common factors, i.e. $(\hat{a}_k, \hat{b}_k) = 1$ up to a unit of $\mathfrak{F}[z]$. Then

$$\frac{\hat{B}}{\hat{a}} = \hat{E}_1 \operatorname{diag}\left\{\frac{\hat{b}_1}{\hat{a}_1}, \frac{\hat{b}_2}{\hat{a}_2}, \ldots, \frac{\hat{b}_r}{\hat{a}_r}, 0, \ldots, 0\right\} \hat{E}_2$$

and defining the matrices

(2.4)
$$\begin{aligned}
\hat{B}_1 &= \hat{E}_1 \operatorname{diag}\{\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_r, 0, \ldots, 0\} \in \mathfrak{F}_{l,m}[z], \\
\hat{A}_2 &= \hat{E}_2^{-1} \operatorname{diag}\{\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_r, 1, \ldots, 1\} \in \mathfrak{F}_{m,m}[z], \\
\hat{A}_1 &= \operatorname{diag}\{\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_r, 1, \ldots, 1\} \hat{E}_1^{-1} \in \mathfrak{F}_{l,l}[z], \\
\hat{B}_2 &= \operatorname{diag}\{\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_r, 0, \ldots, 0\} \hat{E}_2 \in \mathfrak{F}_{l,m}[z],
\end{aligned}$$

we can write

(2.5)
$$\boldsymbol{S} = \hat{B}_1 \hat{A}_2^{-1} = \hat{A}_1^{-1} \hat{B}_2 .$$

These factorizations of the rational matrix $\boldsymbol{S}$ into two polynomial matrices are fundamental ones and play the role analogous to the expression $\boldsymbol{S} = \hat{b}/\hat{a}$ for single-input single-output systems [30; 33]. The factorizations (2.5) enjoy the following properties.

(2.6) $\hat{A}_1$ and $\hat{B}_2$ are left coprime

while

$\hat{B}_1$ and $\hat{A}_2$ are right coprime.

(2.7) The $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_r$ and $\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_r$ are uniquely determined by $\boldsymbol{S}$ modulo units of $\mathfrak{F}[z]$ and $\hat{b}_k \mid \hat{b}_{k+1}$, $\hat{a}_{k+1} \mid \hat{a}_k$, $k = 1, 2, \ldots, r-1$.

(2.8) By construction, the $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_r$ and the nonunit invariant polynomials of the matrix $z\mathbf{I}_n - \mathbf{A} \in \mathfrak{F}_{n,n}[z]$ are associates in $\mathfrak{F}[z]$ provided $\mathbf{A}$ belongs to a minimal realization of $\boldsymbol{S}$. If the $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_r$ are chosen to be monic polynomials, they are called the *invariant polynomials* of $\mathscr{S}$.

The *degree of a rational matrix* $\boldsymbol{S}$ is defined as

$$\delta \boldsymbol{S} = \sum_{k=1}^{r} \partial \hat{a}_k = \partial \det \hat{A}_1 = \partial \det \hat{A}_2 .$$

There are other ways to define $\delta \boldsymbol{S}$; however, they are all equivalent [20].

All these properties justify the following terminology:

$\hat{B}_1$ = left matrix numerator of S
$\hat{A}_2$ = right matrix denominator of S
$\hat{A}_1$ = left matrix denominator of S
$\hat{B}_2$ = right matrix numerator of S.

The characteristic and minimal polynomials of $\mathscr{S}$ have been defined as $\det(z\mathbf{I}_n - \mathbf{A})$ and $\operatorname{mp}(z\mathbf{I}_n - \mathbf{A})$ respectively. If $\mathscr{S}$ is a minimal realization of S, then, by (2.4) and (2.8),

$$\det(z\mathbf{I}_n - \mathbf{A}) = \det \hat{A}_1 = \det \hat{A}_2 ,$$

$$\operatorname{mp}(z\mathbf{I}_n - \mathbf{A}) = \hat{a} = \hat{a}_1$$

modulo units of $\mathfrak{F}[z]$. Further, let ∂_k be the least common denominator of all minors of $S = \mathbf{C}(z\mathbf{I}_n - \mathbf{A})^{-1}\,\mathbf{B} + \mathbf{D}$ with order less than or equal to k. Then [10]

$$\det(z\mathbf{I}_n - \mathbf{A}) = \partial_l = \partial_m ,$$

$$\operatorname{mp}(z\mathbf{I}_n - \mathbf{A}) = \partial_1$$

modulo units of $\mathfrak{F}[z]$. It is to be noted that the two polynomials are different, in general, and both are also different from the denominator of $\det S$ (if S is a square matrix). Example:

$$S = \begin{bmatrix} \dfrac{z-1}{z-0{\cdot}5} & 0 \\ 0 & \dfrac{z-0{\cdot}5}{z-1} \end{bmatrix},$$

$$\det(z\mathbf{I}_n - \mathbf{A}) = (z - 0{\cdot}5)(z - 1) ,$$

$$\det S = 1 .$$

While the degree of the characteristic polynomial determines the dimension of the system, it is the minimal polynomial that determines the dynamical behavior of the system. Example:

$$S = \begin{bmatrix} \dfrac{1}{z} & \dfrac{1}{z} \\ 0 & \dfrac{1}{z} \end{bmatrix},$$

$$\det(z\mathbf{I}_n - \mathbf{A}) = z^2 ,$$

$$\operatorname{mp}(z\mathbf{I}_n - \mathbf{A}) = z ,$$

and, indeed, no dynamical mode that would correspond to z^2 can be excited in the system.

It will be instructive to make the relation between the impulse response matrix S and its minimal realization $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ explicit.

Given a quadruple $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$, the S can be determined simply by computing (2.1). The converse problem is much more difficult, however. Given an S, a minimal realization $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ of S can be obtained as follows [20; 23].

Consider the decomposition

$$S = \frac{\hat{B}}{\hat{a}}$$

defined in (2.3) and let

$$\hat{B} = \hat{B}^0 \bmod \hat{a} .$$

That is,

$$\hat{B} = \hat{B}^0 + \hat{a}\mathbf{D} , \quad \partial \hat{B}^0 < \partial \hat{a} \tag{2.9}$$

for a suitable matrix $\mathbf{D} \in \mathfrak{F}_{l,m}$.

Compute a canonical representation of $\hat{B}^0$,

$$\hat{B}^0 = \hat{E}_1^0 \operatorname{diag} \{\hat{g}_1^0, \hat{g}_2^0, \ldots, \hat{g}_r^0, 0, \ldots, 0\} \hat{E}_2^0 . \tag{2.10}$$

Let

$$\hat{d}_k = (\hat{a}, \hat{g}_k^0) , \quad k = 1, 2, \ldots, r$$

and let

$$\hat{a} = \hat{a}_k \hat{d}_k ,$$
$$\hat{g}_k^0 = \hat{d}_k b_k^0 .$$

It is necessary to choose $\hat{d}_k$ so that the $\hat{a}_k$ be a monic polynomial.

To each invariant polynomial

$$\hat{a}_k = \alpha_0^{(k)} + \alpha_1^{(k)} z + \ldots + \alpha_{n_k-1}^{(k)} z^{n_k-1} + z^{n_k} , \quad k = 1, 2, \ldots, r ,$$

we choose a cyclic matrix $\mathbf{A}_k$ such that $\det (z\mathbf{I}_{n_k} - \mathbf{A}_k) = \hat{a}_k$ e.g.

$$\mathbf{A}_k = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & & 1 \\ -\alpha_0^{(k)} & -\alpha_1^{(k)} & \ldots & & -\alpha_{n_k-1}^{(k)} \end{bmatrix} .$$

We recall that a matrix $\mathbf{A} \in \mathfrak{F}_{n,n}$ is said to be *cyclic* if the polynomials $\det (z\mathbf{I}_n - \mathbf{A})$ and $\operatorname{mp} (z\mathbf{I}_n - \mathbf{A})$ are identical. It means that the invariant polynomials of $z\mathbf{I}_n - \mathbf{A}$ are

$$\hat{a}_1 = \hat{a}_2 = \ldots = \hat{a}_{n-1} = 1 , \quad \hat{a}_n = \det (z\mathbf{I}_n - \mathbf{A}) .$$

By convention, a system whose matrix $\mathbf{A}$ is cyclic is called cyclic.

Write

$$\hat{E}_{1k} = k\text{th column of } \hat{E}_1^0, \quad k = 1, 2, \ldots, l,$$
$$\hat{E}'_{2k} = k\text{th row of } \hat{E}_2^0, \quad k = 1, 2, \ldots, m.$$

In view of our choice of $\mathbf{A}_k$, let us introduce the vectors

$$\hat{V}_k = \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n_k - 1} \end{bmatrix} \in \mathfrak{F}_{n_k,1}[z],$$

$$\hat{W}'_k = [\alpha_1^{(k)} + \ldots + \alpha_{n_k-1}^{(k)} z^{n_k-2} + z^{n_k-1}, \ldots, \alpha_{n_k-1}^{(k)} + z, 1] \in \mathfrak{F}_{n_k,1}[z],$$

and observe that

$$\hat{a}_k(z\mathbf{I}_{n_k} - \mathbf{A}_k)^{-1} = \hat{V}_k \hat{W}_k \bmod \hat{a}_k. \tag{2.11}$$

Consider the system of equations

$$\begin{aligned} \mathbf{C}_k \hat{V}_k &= \hat{E}_{1k} \hat{b}_k^0 \bmod \hat{a}_k, \quad k = 1, 2, \ldots, r, \\ \hat{W}'_k \mathbf{B}_k &= \hat{E}'_{2k} \bmod \hat{a}_k, \quad k = 1, 2, \ldots, r, \end{aligned} \tag{2.12}$$

for $\mathbf{B}_k \in \mathfrak{F}_{n_k,m}$ and $\mathbf{C}_k \in \mathfrak{F}_{l,n_k}$.

Since the right-hand sides of (2.12) can be taken to satisfy

$$\partial(\hat{E}_{1k}\hat{b}_k^0) < \partial\hat{a}_k, \quad \partial\hat{E}'_{2k} < \partial\hat{a}_k, \quad k = 1, 2, \ldots, r,$$

and since the elements of $\hat{V}_k$ and $\hat{W}'_k$ are linearly independent over $\mathfrak{F}$, equations (2.12) have unique solutions $\mathbf{B}_k$ and $\mathbf{C}_k$.

We get

$$\begin{aligned} \hat{E}_{1k}\hat{g}_k^0\hat{E}'_{2k} &= \hat{d}_k(\hat{E}_{1k}\hat{b}_k^0\hat{E}'_{2k}) = \\ &= \hat{d}_k(\mathbf{C}_k\hat{V}_k\hat{W}'_k\mathbf{B}_k) \bmod \hat{d}_k\hat{a}_k \end{aligned}$$

by virtue of (2.12) and

$$\begin{aligned} \hat{E}_{1k}\hat{g}_k^0\hat{E}'_{2k} &= \hat{d}_k[\hat{a}_k\mathbf{C}_k(z\mathbf{I}_{n_k} - \mathbf{A}_k)^{-1}\mathbf{B}_k] \bmod \hat{d}_k\hat{a}_k = \\ &= \hat{a}[\mathbf{C}_k(z\mathbf{I}_{n_k} - \mathbf{A}_k)^{-1}\mathbf{B}_k] \bmod \hat{a} \end{aligned}$$

on using (2.11). Hence, by (2.10),

$$\begin{aligned} \hat{B}^0 &= \sum_{k=1} \hat{E}_{1k}\hat{g}_k^0\hat{E}'_{2k} \bmod \hat{a} = \\ &= \hat{a}\sum_{k=1}^{r} \mathbf{C}_k(z\mathbf{I}_{n_k} - \mathbf{A}_k)^{-1}\mathbf{B}_k \bmod \hat{a}. \end{aligned}$$

Since $\partial \hat{B}^0 < \partial \hat{a}$, we have equality not only mod $\hat{a}$ but also in the ordinary sense. Using (2.9) we have

$$\hat{B} = \hat{a} \sum_{k=1}^{r} \mathbf{C}_k (z\mathbf{I}_{n_k} - \mathbf{A}_k)^{-1} \mathbf{B}_k + \hat{a}\mathbf{D}$$

and hence

$$\boldsymbol{S} = \sum_{k=1}^{r} \mathbf{C}_k (z\mathbf{I}_{n_k} - \mathbf{A}_k)^{-1} \mathbf{B}_k + \mathbf{D} .$$

Therefore, the quadruple $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ with

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & & \\ & \cdots\cdots & \\ & & \mathbf{A}_r \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_r \end{bmatrix},$$

$$\mathbf{C} = [\mathbf{C}_1 \dots \mathbf{C}_r]$$

and $\mathbf{D}$ given by (2.9) is a realization of $\boldsymbol{S}$; by construction, it is minimal. The dimension of the minimal realization is given by

$$n = \sum_{k=1}^{r} n_k = \sum_{k=1}^{r} \partial \hat{a}_k = \delta \boldsymbol{S} .$$

If $l = m = 1$ (single-input single-output system), we have

$$\boldsymbol{S} = \frac{\hat{b}}{\hat{a}} \in \mathfrak{F}\{z^{-1}\} ,$$

where $\hat{a}$ is a monic polynomial, and the problem of minimal realization greatly simplifies. We write

$$\hat{b} = \hat{b}^0 \bmod \hat{a} ,$$

that is,

$$\hat{b} = \hat{b}^0 + \hat{a}d , \quad \partial \hat{b}^0 < \partial \hat{a}$$

for a suitable $d \in \mathfrak{F}$ and hence

(2.13) $$\mathbf{D} = [d] .$$

There is just one invariant polynomial,

$$\hat{a} = \alpha_0 + \alpha_1 z + \dots + \alpha_{n-1} z^{n-1} + z^n ,$$

and we choose

(2.14) $$\mathbf{A} = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ \dots & \dots & \dots & \dots & \dots \\ & & & & 1 \\ -\alpha_0 & -\alpha_1 & \dots & & -\alpha_{n-1} \end{bmatrix} .$$

Equations (2.12) simplify to

$$\mathbf{C}\hat{V}_1 = \hat{b}^0 = \beta_0^0 + \beta_1^0 z + \ldots + \beta_{n-1}^0 z^{n-1},$$

$$\hat{W}_1'\mathbf{B} = 1,$$

the solution being

$$\mathbf{C} = [\beta_0^0 \ \beta_1^0 \ldots \beta_{n-1}^0], \quad \mathbf{B} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \tag{2.15}$$

The above quadruple $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ is a minimal realization of $\boldsymbol{S}$.

It is to be noted that the minimal realization has been obtained as a direct sum of cyclic subsystems, each subsystem being generated by one invariant polynomial of the impulse response matrix. This procedure makes clear the significance of the invariant polynomials. They fully describe the structure as well as dynamical behavior of the system and show how the system can be decomposed into a direct sum of subsystems which behave like single-input single-output systems.

Example 2.1. Find a minimal realization over $\mathfrak{R}$ of the impulse response matrix

$$\boldsymbol{S} = \frac{\begin{bmatrix} z+1 & z-1 \\ z^2+2z+1 & -1 \\ 1 & z-1 \end{bmatrix}}{z^2}.$$

We have

$$\begin{bmatrix} z+1 & z-1 \\ z^2+2z+1 & -1 \\ 1 & z-1 \end{bmatrix} = \begin{bmatrix} z+1 & z-1 \\ 2z+1 & -1 \\ 1 & z-1 \end{bmatrix} + z^2 \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}$$

and hence

$$\hat{B}^0 = \begin{bmatrix} z+1 & z-1 \\ 2z+1 & -1 \\ 1 & z-1 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Computing a canonical representation of $\hat{B}^0$ we obtain

$$\hat{E}_1^0 = \begin{bmatrix} z+1 & -z+1 & 0{\cdot}5 \\ 2z+1 & -2z+1 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \hat{E}_2^0 = \begin{bmatrix} 1 & z-1 \\ 0 & 1 \end{bmatrix},$$

$$\hat{g}_1^0 = 1, \quad \hat{g}_2^0 = z,$$

and

$$\hat{b}_1^0 = 1\,, \quad \hat{b}_2^0 = 1\,,$$

$$\hat{a}_1 = z^2\,, \quad \hat{a}_2 = z\,.$$

Therefore, we take

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{A}_2 = [0]\,.$$

Write

$$\hat{V}_1 = \begin{bmatrix} 1 \\ z \end{bmatrix}, \quad \hat{W}_1' = [z \;\; 1]\,.$$

Then $\mathbf{C}_1$ and $\mathbf{B}_1$ are given by

$$\mathbf{C}_1 \begin{bmatrix} 1 \\ z \end{bmatrix} = \begin{bmatrix} z + 1 \\ 2z + 1 \\ 1 \end{bmatrix} \bmod z^2\,,$$

$$[z \quad 1]\,\mathbf{B}_1 = [1 \quad z - 1] \bmod z^2\,,$$

which are also equalities in the ordinary sense and

$$\mathbf{C}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 2 & 0 \end{bmatrix}, \quad \mathbf{B}_1 = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}.$$

Further write

$$\hat{V}_2 = [1]\,, \quad \hat{W}_2' = [1]\,.$$

Then $\mathbf{C}_2$ and $\mathbf{B}_2$ are given by

$$\mathbf{C}_2\,[1] = \begin{bmatrix} -\;z + 1 \\ -2z + 1 \\ 0 \end{bmatrix} \bmod z\,,$$

$$[1]\,\mathbf{B}_2 = [0 \;\; 1] \bmod z\,,$$

which yields the equalities

$$\mathbf{C}_2\,[1] = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix},$$

$$[1]\,\mathbf{B}_2 = [0 \;\; 1]$$

in the ordinary sense, and

$$\mathbf{C}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{B}_2 = [0 \;\; 1]\,.$$

Therefore, a minimal realization of S is given as $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ where

$$\mathbf{A} = \left[\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 0 \end{array}\right], \quad \mathbf{B} = \left[\begin{array}{cc} 0 & 1 \\ 1 & -1 \\ \hline 0 & 1 \end{array}\right],$$

$$\mathbf{C} = \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 0 \end{array}\right], \quad \mathbf{D} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

This realization is shown in Fig. 1. Its dimension is $n = 3$ and it is exhibited as the direct sum of two cyclic subsystems $\mathscr{S}_1$ and $\mathscr{S}_2$.

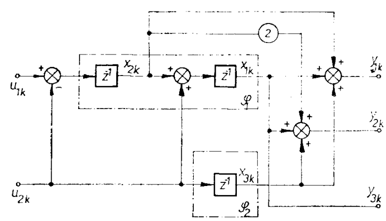

Fig. 1. A minimal realization of the system in Example 2.1.

Example 2.2. Find a minimal realization over $\mathfrak{Q}$ of the impulse response

$$S = \frac{z}{z^2 - z - 1}.$$

Since

$$S = z^{-1} + z^{-2} + 2z^{-3} + 3z^{-4} + 5z^{-5} + 8z^{-6} + 13z^{-7} + \dots,$$

we may think of this problem as of realizing the Fibonacci sequence $\{\sigma_k\}_{k=0}^{\infty}$, where

$$\sigma_0 = 0, \quad \sigma_1 = 1, \quad \sigma_k = \sigma_{k-2} + \sigma_{k-1}, \quad k = 2, 3, \dots.$$

We have

$$\hat{b} = z, \quad \hat{a} = z^2 - z - 1$$

and since

$$\partial\hat{b} < \partial\hat{a},$$

the quadruple $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ with

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} 0 & 1 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 0 \end{bmatrix}$$

is a minimal realization of S by virtue of (2.14) and (2.15), see Fig. 2.

Example 2.3. Find a minimal realization of the impulse response matrix

$$S = \frac{\begin{bmatrix} 1 & z+1 & z \\ z & 0 & z \\ z+1 & z+1 & z^2+z \end{bmatrix}}{z(z+1)}$$

describing a finite automaton over $\mathfrak{Z}_2$.

We compute

$$\hat{B}^0 = \begin{bmatrix} 1 & z+1 & z \\ z & 0 & z \\ z+1 & z+1 & 0 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

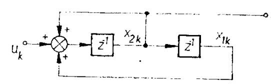

Fig. 2. A minimal realization of the Fibonacci sequence.

The canonical decomposition of $\hat{B}^0$ yields

$$\hat{E}_1^0 = \begin{bmatrix} 1 & 0 & 0 \\ z & 1 & 0 \\ z+1 & 1 & 1 \end{bmatrix}, \quad \hat{E}_2^0 = \begin{bmatrix} 1 & z+1 & z \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\hat{g}_1^0 = 1\,, \quad \hat{g}_2^0 = z(z+1)\,, \quad \hat{g}_3^0 = 0$$

and

$$\hat{b}_1^0 = 1\,, \qquad \hat{b}_2^0 = 1\,,$$

$$\hat{a}_1 = z^2 + z\,, \hat{a}_2 = 1\,,$$

since $r = 2$.

Therefore, the system itself is cyclic with dimension $n = 2$ and we take

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$

Write

$$\hat{V}_1 = \begin{bmatrix} 1 \\ z \end{bmatrix}, \quad \hat{W}_1' = \begin{bmatrix} z+1 & 1 \end{bmatrix}.$$

Then $\mathbf{C}_1$ and $\mathbf{B}_1$ are given by

$$\mathbf{C}_1 \begin{bmatrix} 1 \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ z \\ z+1 \end{bmatrix} \operatorname{mod}(z^2 + z)\,,$$

$$\begin{bmatrix} z+1 & 1 \end{bmatrix} \mathbf{B}_1 = \begin{bmatrix} 1 & z+1 & z \end{bmatrix} \operatorname{mod}(z^2 + z)\,,$$

which are also equalities in the ordinary sense and

$$\mathbf{C}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{B}_1 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Thus the $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$, where

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

is a minimal realization of $\boldsymbol{S}$. It is depicted in Fig. 3.

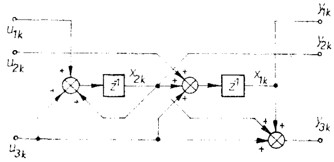

Fig. 3. A minimal realization of the system in Example 2.3.

By definition, the impulse response matrix $\boldsymbol{S}$ belongs to $\mathfrak{F}_{l,m}\{z^{-1}\}$. The elements of $\mathfrak{F}_{l,m}\{z^{-1}\}$ can be written as ratios of two polynomials, but in *two different ways*. If

$$\boldsymbol{S} = \Sigma_0 + \Sigma_1 z^{-1} + \Sigma_2 z^{-2} + \ldots \in \mathfrak{F}_{l,m}\{z^{-1}\},$$

then

(2.16) $$\boldsymbol{S} = \frac{\hat{B}}{\hat{a}}$$

where $\hat{B} \in \mathfrak{F}_{l,m}[z]$, $\hat{a} \in \mathfrak{F}[z]$ and

$$(\hat{a}, \hat{B}) = 1,$$
$$\partial \hat{B} \leqq \partial \hat{a},$$

as shown in (2.3). But also

(2.17) $$\boldsymbol{S} = \frac{\hat{B}}{\hat{a}} \frac{z^{-\partial \hat{a}}}{z^{-\partial \hat{a}}} = \frac{B}{a},$$

where $B \in \mathfrak{F}_{l,m}[z^{-1}]$, $a \in \mathfrak{F}[z^{-1}]$ and

$$(a, B) = 1,$$
$$(a, z^{-1}) = 1.$$

The above procedure implies that

$$a = z^{-\partial a}\hat{a}\,, \tag{2.18}$$
$$B = z^{-\partial a}\hat{B}$$

and a is the least common denominator of all elements of

$$S = \mathbf{C}z^{-1}(\mathbf{I}_n - z^{-1}\mathbf{A})^{-1}\,\mathbf{B} + \mathbf{D}\,.$$

To *analyze* the structure and dynamics of S we have employed representation (2.16). However, it proves more profitable to use representation (2.17) when *synthesizing* a system. The main advantage of this representation stems from the fact that any matrix in $\mathfrak{F}_{l,m}[z^{-1}]$ can be realized as a system $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$. This is not true of matrices in $\mathfrak{F}_{l,m}[z]$. To put it in other words, an impulse response matrix in the indeterminate z must be manipulated as a whole, not as a ratio of two polynomials. Otherwise the physical realizability of the synthesized system is at stake. If, however, the impulse response matrix is written in the indeterminate z^{-1}, we can take full advantage of manipulating the numerator and denominator polynomial matrices individually while the physical realizability of the synthesized product is inherently guaranteed. As a result, the algebra of rational matrices is reduced to much simpler algebra of polynomial matrices.

The polynomials $\det(\mathbf{I}_n - z^{-1}\mathbf{A}) \in \mathfrak{F}[z^{-1}]$ and $\mathrm{mp}(\mathbf{I}_n - z^{-1}\mathbf{A}) \in \mathfrak{F}[z^{-1}]$ will be called the *pseudocharacteristic* and *pseudominimal* polynomials of $\mathscr{S}$, respectively.

It is to be noted that even though

$$\det(\mathbf{I}_n - z^{-1}\mathbf{A}) = z^{-n}\det(z\mathbf{I}_n - \mathbf{A})\,,$$

the pseudocharacteristic (pseudominimal) polynomial may be quite different from the characteristic (minimal) polynomial. Example:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\det(z\mathbf{I}_n - \mathbf{A}) = z^3\,, \quad \det(\mathbf{I}_n - z^{-1}\mathbf{A}) = 1\,,$$

$$\mathrm{mp}(z\mathbf{I}_n - \mathbf{A}) = z^2\,, \quad \mathrm{mp}(\mathbf{I}_n - z^{-1}\mathbf{A}) = 1\,.$$

The notion of stability for the polynomials of $\mathfrak{F}[z^{-1}]$ can be defined as follows. Any polynomial $a \in \mathfrak{F}[z^{-1}]$ can be written in the form

$$a = z^{-d}a_0\,, \quad (a_0, z^{-1}) = 1$$

for some integer $d \geqq 0$. Then the a is *stable* if and only if

(a) $d = 0$,

(b) $\hat{a}_0 = z^{\partial a_0}a_0 \in \mathfrak{F}[z]$ is stable.

In view of this definition the characteristic polynomial of a system is stable if and only if its pseudocharacteristic polynomial is stable. Therefore, the pseudocharacteristic polynomial can be used to check for stability of a system.

The matrix $S = B/a \in \mathfrak{F}_{l,m}\{z^{-1}\}$ can also be put into the canonical form. Let

$$B = E_1 \operatorname{diag}\{g_1, g_2, \ldots, g_r, 0, \ldots, 0\} E_2$$

and let

$$\frac{g_k}{a} = \frac{b_k}{a_k}, \quad k = 1, 2, \ldots, r$$

after cancelling common factors, i.e. $(a_k, b_k) = 1$ up to a unit of $\mathfrak{F}[z^{-1}]$. Then

$$\frac{B}{a} = E_1 \operatorname{diag}\left\{\frac{b_1}{a_1}, \frac{b_2}{a_2}, \ldots, \frac{b_r}{a_r}, 0, \ldots, 0\right\} E_2$$

and defining the matrices

$$(2.19) \qquad \begin{aligned} B_1 &= E_1 \operatorname{diag}\{b_1, b_2, \ldots, b_r, 0, \ldots, 0\} \in \mathfrak{F}_{l,m}[z^{-1}], \\ A_2 &= E_2^{-1} \operatorname{diag}\{a_1, a_2, \ldots, a_r, 1, \ldots, 1\} \in \mathfrak{F}_{m,m}[z^{-1}], \\ A_1 &= \operatorname{diag}\{a_1, a_2, \ldots, a_r, 1, \ldots, 1\} E_1^{-1} \in \mathfrak{F}_{l,l}[z^{-1}], \\ B_2 &= \operatorname{diag}\{b_1, b_2, \ldots, b_r, 0, \ldots, 0\} E_2 \in \mathfrak{F}_{l,m}[z^{-1}], \end{aligned}$$

we can write

$$(2.20) \qquad S = B_1 A_2^{-1} = A_1^{-1} B_2 .$$

These factorizations of the rational matrix S are, as a rule, different from those obtained in (2.5). However, they are also very important, especially in the synthesis of optimum control systems, and play the role similar to the expression $s = b/a$ for single-input single-output systems [30; 33]. The factorizations (2.20) enjoy the following properties.

(2.21) A_1 and B_2 are left coprime while B_1 and A_2 are right coprime.

(2.22) The $a_1, a_2, \ldots, a_r$ and $b_1, b_2, \ldots, b_r$ are uniquely determined by S modulo units of $\mathfrak{F}[z^{-1}]$ and $b_k \mid b_{k+1}$, $a_{k+1} \mid a_k$, $k = 1, 2, \ldots, r + 1$.

(2.23) By construction, the $a_1, a_2, \ldots, a_r$ and the nonunit invariant polynomials of the matrix $\mathbf{I}_n - z^{-1}\mathbf{A} \in \mathfrak{F}_{n,n}[z^{-1}]$ are associates in $\mathfrak{F}[z^{-1}]$ provided $\mathbf{A}$ belongs to a minimal realization of S. The $a_1, a_2, \ldots, a_r$ will be called the *pseudoinvariant polynomials* of $\mathscr{S}$.

If $\mathscr{S}$ is a minimal realization of S, then by (2.19) and (2.23)

$$\det(\mathbf{I}_n - z^{-1}\mathbf{A}) = \det A_1 = \det A_2,$$

$$\operatorname{mp}(\mathbf{I}_n - z^{-1}\mathbf{A}) = a = a_1$$

modulo units of $\mathfrak{F}[z^{-1}]$. Further, let d_k be the least common denominator of all minors of $S = \mathbf{C}z^{-1}(\mathbf{I}_n - z^{-1}\mathbf{A})^{-1}\,\mathbf{B} + \mathbf{D}$ with order less than or equal to k. Then

$$\det(\mathbf{I}_n - z^{-1}\mathbf{A}) = d_l = d_m,$$

$$\operatorname{mp}(\mathbf{I}_n - z^{-1}\mathbf{A}) = d_1$$

modulo units of $\mathfrak{F}[z^{-1}]$. The two polynomials are different, in general, and both are also different from the denominator of det S (if S is a square matrix).

To motivate why the polynomials $a_k \in \mathfrak{F}[z^{-1}]$ have the prefix "pseudo", we note that they do not yield structural and dynamical invariants of the system, in general. Indeed,

$$z^{\partial a_k} a_k \neq \hat{a}_k$$

i.e. we cannot compute $\hat{a}_k$ when given a_k, and

$$\partial a_k \neq \partial \hat{a}_k = n_k .$$

As a consequence, we cannot find a minimal realization of S by manipulating the canonical decomposition of S over $\mathfrak{F}[z^{-1}]$ in the same way as the canonical decomposition of S over $\mathfrak{F}[z]$. Example:

$$S = \frac{\begin{bmatrix} z^2 & z^2 \\ z^2 & z^2 - 1 \end{bmatrix}}{z^3(z-1)} = \frac{\begin{bmatrix} z^{-2} & z^{-2} \\ z^{-2} & z^{-2} - z^{-4} \end{bmatrix}}{1 - z^{-1}},$$

$$\hat{b}_1 = 1, \qquad \hat{b}_2 = 1,$$

$$\hat{a}_1 = z^3(z-1), \quad \hat{a}_2 = z(z-1),$$

$$n_1 = 4, \quad n_2 = 2, \quad n = 6;$$

$$b_1 = z^{-2}, \qquad b_2 = z^{-2},$$

$$a_1 = 1 - z^{-1}, \quad a_2 = 1 - z^{-1}.$$

We shall see, however, that the knowledge of all invariants is unnecessary when synthesizing an optimal control system. It is just the information contained in the pseudoinvariant polynomials that is needed.

2.2. Valuation and Norms

The ground field $\mathfrak{F}$ is the first thing to be specified in the definition of a system. Roughly speaking, it defines the admissible numbers and the arithmetic operations on them. The second thing is to endow the ground field with a metrics so that we may investigate how a system response tends to zero, or to tell whether two responses are equal or which one is better. This is done by introducing the concept of valuation in a field [57; 66].

The *value* in a field $\mathfrak{F}$ is a mapping $\mathscr{V} : \mathfrak{F} \to \mathfrak{R}$ which with every element $\alpha \in \mathfrak{F}$ associates an element $\mathscr{V}(\alpha) \in \mathfrak{R}$, the ordered field of reals, satisfying the axioms

$$\mathscr{V}(0) = 0\,,$$
$$\mathscr{V}(\alpha) > 0\,, \quad \alpha \neq 0\,,$$
$$\mathscr{V}(\alpha\beta) = \mathscr{V}(\alpha)\,\mathscr{V}(\beta)\,,$$
$$\mathscr{V}(\alpha + \beta) \leqq \mathscr{V}(\alpha) + \mathscr{V}(\beta)\,.$$

It immediately follows that

$$\mathscr{V}(1) = 1\,,$$
$$\mathscr{V}(-\alpha) = \mathscr{V}(\alpha)\,,$$
$$\mathscr{V}\left(\frac{1}{\alpha}\right) = \frac{1}{\mathscr{V}(\alpha)}\,.$$

There is a trivial valuation, viz.

$$\mathscr{V}(0) = 0\,, \qquad \mathscr{V}(\alpha) = 1\,, \quad \alpha \neq 0\,, \tag{2.24}$$

which is valid in any field $\mathfrak{F}$.

If $\mathfrak{F} = \mathfrak{Z}_p$ or an algebraic extension $\mathfrak{Z}_p[z]_w$ of $\mathfrak{Z}_p$, where $\mathscr{W}$ is a polynomial prime in $\mathfrak{Z}_p[z]$, all nonzero elements of $\mathfrak{F}$ are roots of unity [66], i.e. if $\alpha \in \mathfrak{F}$, $\alpha \neq 0$ an integer n exists such that $\alpha^n = 1$. Then

$$1 = \mathscr{V}(\alpha^n) = \mathscr{V}^n(\alpha) \;\Rightarrow\; \mathscr{V}(\alpha) = 1$$

and no other valuation but the trivial one exists.

If $\mathfrak{F}$ is a subfield of $\mathfrak{C}$, we can always take

$$\mathscr{V}(\alpha) = |\alpha|\,, \tag{2.25}$$

the ordinary absolute value of possibly complex numbers. It is this valuation that is implicitely assumed in most control problems. However, other valuations may be taken.

Consider the quotient field $\mathfrak{F}(z^{-1})$ of the ring $\mathfrak{F}[z]$ for any field $\mathfrak{F}$. If $\alpha \in \mathfrak{F}(z^{-1})$, we can take

$$\mathscr{V}(\alpha) = \varrho^{-\partial\alpha} \tag{2.26}$$

for any real number $\varrho > 1$. In particular,

$$\mathscr{V}(0) = \varrho^{-\infty} = 0 .$$

We reiterate that valuation in a field induces a metrics relative to which the system behaviour is investigated. A polynomial $a \in \mathfrak{F}[z]$ has been defined stable if $1/a \in \mathfrak{F}^{+}\{z^{-1}\}$, i.e. when the sequence obtained by the formal long division of $1/a$ into ascending powers of z^{-1} is a "zero sequence in $\mathfrak{F}$". In fact, it should have been said "a zero sequence with respect to a valuation $\mathscr{V}$ in $\mathfrak{F}$", since it is the valuation $\mathscr{V}$ that determines the convergence of sequences in $\mathfrak{F}$. It is quite possible that a sequence converges to zero in one valuation and diverges from zero in another, see Example 2.9.

Given a vector space $\mathscr{W}$ over $\mathfrak{F}$, we define the *norm* in $\mathscr{W}$ as a mapping of $\mathscr{W}$ into $\mathfrak{R}$ which with every element $a \in \mathscr{V}$ associates an element $\|a\| \in \mathfrak{R}$ such that [57], [65]

$$\|0\| = 0 ,$$
$$\|a\| > 0 , \quad a \neq 0 ,$$
$$\|\alpha a\| = \mathscr{V}(\alpha)\, \|a\| , \quad \alpha \in \mathfrak{F} ,$$
$$\|a + b\| \leqq \| a \| + \| b\| .$$

Thus a valuation in $\mathfrak{F}$ induces a norm in a vector space $\mathscr{W}$ over $\mathfrak{F}$.

The set $\mathfrak{F}^{+}\{z^{-1}\}$ can be viewed as a vector space over $\mathfrak{F}$ and can be normed as follows. Let

$$a = \alpha_0 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \ldots \in \mathfrak{F}^{+}\{z^{-1}\} ,$$

then

$$\|a\|^2 = \sum_{k=0}^{\infty} \mathscr{V}^2(\alpha_k) .$$

More generally, the set $\mathfrak{F}^{+}_{l,m}\{z^{-1}\}$ is also a vector space over $\mathfrak{F}$ and it can be normed quite analogously. Let

$$A = A_0 + A_1 z^{-1} + A_2 z^{-2} + \ldots \in \mathfrak{F}^{+}_{l,m}\{z^{-1}\}$$

and let A_k have elements α_{ijk}, $i = 1, 2, \ldots, l$ and $j = 1, 2, \ldots, m$. Then

$$\|A\|^2 = \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=0}^{\infty} \mathscr{V}^2(\alpha_{ijk}) . \tag{2.27}$$

To distinguish this norm from other possible norms in $\mathfrak{F}^+_{l,m}\{z^{-1}\}$, we shall call it the *quadratic* norm. It will be used to compare system responses in the least squares control problems.

For any field $\mathfrak{F}$ valuated by (2.24) the norm (2.27) in $\mathfrak{F}^+_{l,m}\{z^{-1}\}$ can be interpreted as the number of nonzero elements in the sequences over $\mathfrak{F}_{l,m}$.

Now let $\mathfrak{F}$ be a subfield of the field $\mathfrak{C}$ of complex numbers valuated by (2.25). If $\alpha \in \mathfrak{F}$, then

$$\bar{\alpha} = \text{complex conjugate of } \alpha$$

and if

$$\boldsymbol{A} = A_n z^{-n} + A_{n+1} z^{-(n+1)} + \ldots \in \mathfrak{F}_{l,m}(z^{-1}),$$

then

$$\begin{aligned} \boldsymbol{A}' &= \text{transpose of } \boldsymbol{A}, \\ \operatorname{tr} \boldsymbol{A} &= \text{trace of } \boldsymbol{A}, \\ \langle \boldsymbol{A} \rangle &= A_0, \quad \text{the term of } A \text{ at } z^0, \\ \boldsymbol{A}^{=} &= \bar{A}_n z^n + \bar{A}_{n+1} z^{n+1} + \ldots. \end{aligned}$$

It can directly be verified that

$$\begin{aligned} (\boldsymbol{A} + \boldsymbol{B})^{=} &= \boldsymbol{A}^{=} + \boldsymbol{B}^{=}, \\ (\boldsymbol{AB})^{=} &= \boldsymbol{A}^{=}\boldsymbol{B}^{=}, \\ (\boldsymbol{A})^{==} &= \boldsymbol{A}, \end{aligned}$$

for any two suitably dimensioned matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ over $\mathfrak{F}(z^{-1})$.

The above definitions remain valid in the $\mathfrak{F}_{l,m}\{z^{-1}\}$, $\mathfrak{F}^+_{l,m}\{z^{-1}\}$, and $\mathfrak{F}_{l,m}[z^{-1}]$ since they are all subsets of $\mathfrak{F}_{l,m}(z^{-1})$. In particular, if

$$A = A_0 + A_1 z^{-1} + \ldots + A_n z^{-n} \in \mathfrak{F}_{l,m}[z^{-1}]$$

and if $\partial A = n \geqq 0$, we define the polynomial *reciprocal* to A as

$$A^{\sim} = z^{-n} A^{=} = \bar{A}_0 z^{-n} + \bar{A}_1 z^{-(n-1)} + \ldots + \bar{A}_n \in \mathfrak{F}_{l,m}[z^{-1}] \tag{2.28}$$

It follows that

$$\partial A^{\sim} \leqq \partial A,$$

the equality sign holding if and only if $A_0 \neq 0$.

Since $\mathfrak{F}$ is a subfield of $\mathfrak{C}$ and valuation (2.25) is taken, we can write

$$|\alpha|^2 = \bar{\alpha}\alpha.$$

Then the quadratic norm of A can be written as

$$\|A\|^2 = \operatorname{tr} \langle A'^{=} A \rangle . \tag{2.29}$$

Indeed,

$$A'^{=}A^{=} \dots + (\bar{A}'_1 A_0 + \bar{A}'_2 A_1 + \dots) z + + (\bar{A}'_0 A_0 + \bar{A}'_1 A_1 + \dots) + + (\bar{A}'_0 A_1 + \bar{A}'_1 A_2 + \dots) z^{-1} + \dots,$$

and since

$$\langle A'^{=} A \rangle = \bar{A}'_0 A_0 + \bar{A}'_1 A_1 + \dots$$

$$\operatorname{tr} \bar{A}'_k A_k = \sum_{i=1}^{l} \sum_{j=1}^{m} |\alpha_{ijk}|^2 ,$$

we obtain (2.29). In particular,

$$\|A\|^2 = \langle A'^{=} A \rangle$$

for any $A \in \mathfrak{F}^+_{l,1}\{z^{-1}\}$.

The assumption that $A \in \mathfrak{F}^+_{l,m}\{z^{-1}\}$ is essential. If A is not stable, $\|A\|^2$ goes to infinity while $\operatorname{tr} \langle A'^{=} A \rangle$ may remain finite. Example:

$$A = \frac{2 - z^{-1}}{1 - 2z^{-1}} = 2 + 3z^{-1} + 6z^{-2} + 12z^{-3} + \dots \in \mathfrak{R}\{z^{-1}\} ,$$

$$\|A\|^2 \to \infty .$$

$$\operatorname{tr} \langle A'^{=} A \rangle = \left\langle \frac{2 - z}{1 - 2z} \frac{2 - z^{-1}}{1 - 2z^{-1}} \right\rangle = \langle 1 \rangle = 1 .$$

2.3. Matrix Factorizations

Throughout all parts of the paper we shall frequently use the following concepts.

Given a field $\mathfrak{F}$ with valuation $\mathscr{V}$, consider a nonzero polynomial $p \in \mathfrak{F}[z^{-1}]$. We define the factorization

$$p = p^+ p^- ,$$

where p^+ is the stable (with respect to $\mathscr{V}$) factor of p having highest degree and belonging to $\mathfrak{F}[z^{-1}]$. Both factors are unique to within a unit $e \in \mathfrak{F}[z^{-1}]$, $p = (p^+ e)(e^{-1} p^-)$.

It should be noted that the same polynomial viewed over different fields can have different factorizations. Example:

$$p = 1 - 2z^{-1} - z^{-2} \in \mathfrak{Q}[z^{-1}] , \text{ value } (2.25); \ p^+ = 1 .$$

$$p = 1 - 2z^{-1} - z^{-2} \in \mathfrak{R}[z^{-1}], \text{ value } (2.25); \ p^+ = 1 - (1 - \sqrt{2}) z^{-1} .$$

The same is true of different valuations. Example:

$$p = 2 + z^{-1} \in \mathfrak{R}[z^{-1}], \text{ value (2.25)}; \; p^+ = 2 + z^{-1},$$

$$p = 2 + z^{-1} \in \mathfrak{R}[z^{-1}], \text{ value (2.24)}; \; p^+ = 1 .$$

Given a polynomial matrix $P \in \mathfrak{F}_{l,m}[z^{-1}]$, $P \neq 0$, and let

$$P = E_1 \operatorname{diag}\{p_1, p_2, \ldots, p_r, 0, \ldots, 0\} E_2$$

be the canonical representation of P. Then we define the factorizations

$$P = P_1^+ P_2^- = P_1^- P_2^+ , \tag{2.30}$$

where

$$P_1^+ = E_1 \operatorname{diag}\{p_1^+, p_2^+, \ldots, p_r^+, 1, \ldots, 1\} \in \mathfrak{F}_{l,l}[z^{-1}],$$

$$P_2^- = \operatorname{diag}\{p_1^-, p_2^-, \ldots, p_r^-, 0, \ldots, 0\} E_2 \in \mathfrak{F}_{l,m}[z^{-1}],$$

$$P_1^- = E_1 \operatorname{diag}\{p_1^-, p_2^-, \ldots, p_r^-, 0, \ldots, 0\} \in \mathfrak{F}_{l,m}[z^{-1}],$$

$$P_2^+ = \operatorname{diag}\{p_1^+, p_2^+, \ldots, p_r^+, 1, \ldots, 1\} E_2 \in \mathfrak{F}_{m,m}[z^{-1}].$$

Observe that P_1^+ and P_2^+ are nonsingular matrices by definition and that $\operatorname{mp} P_1^+ = \operatorname{mp} P_2^+ = (\operatorname{mp} P)^+$ up to units of $\mathfrak{F}[z^{-1}]$.

The factors P_2^+, P_2^- and P_1^+, P_1^- are determined by P uniquely to within their left and right associates respectively,

$$P = (P_1^+ E_3)(E_3^{-1} P_2^-) = (P_1^- E_4^{-1})(E_4 P_2^+),$$

where E_3 is a unit of $\mathfrak{F}_{l,l}[z^{-1}]$ and E_4 is a unit of $\mathfrak{F}_{m,m}[z^{-1}]$.

Example 2.4. Factorize the polynomial matrix

$$P = \begin{bmatrix} 1 & 1 \\ z^{-1} & -z^{-1} + z^{-2} \\ 1 & 1 + 2z^{-1} - z^{-2} \end{bmatrix}$$

over $\mathfrak{R}[z^{-1}]$ with valuation (2.25).

We compute

$$P = \begin{bmatrix} 1 & 0 & 0 \\ z^{-1} & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & z^{-1}(z^{-1} - 2) \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

and hence

$$P_1^+ = \begin{bmatrix} 1 & 0 & 0 \\ z^{-1} & z^{-1} - 2 & 0 \\ 1 & -z^{-1} + 2 & 1 \end{bmatrix}, \quad P_2^- = \begin{bmatrix} 1 & 1 \\ 0 & z^{-1} \\ 0 & 0 \end{bmatrix},$$

$$P_1^- = \begin{bmatrix} 1 & 0 \\ z^{-1} & z^{-1} \\ 1 & -z^{-1} \end{bmatrix}, \quad P_2^+ = \begin{bmatrix} 1 & 1 \\ 0 & z^{-1} - 2 \end{bmatrix}$$

to within left or right associates.

Now we turn our attention to another sort of matrix factorization, which is essential for the least squares optimization.

Let $\mathfrak{F}$ be a subfield of the field $\mathfrak{C}$ of complex numbers valuated by (2.25). Then an element $\Omega \in \mathfrak{F}_{m,m}$ is said to be *unitary* if $\bar{\Omega}'\Omega = \Omega\bar{\Omega}' = \mathrm{I}_m$. The unitary elements in $\mathfrak{C}_{m,m}$ are $m \times m$ unitary matrices while the unitary elements in $\mathfrak{R}_{m,m}$ are $m \times m$ orthogonal matrices. In particular, the unitary elements in $\mathfrak{R}$ are ± 1 and the unitary elements in $\mathfrak{C}$ are all complex units.

Given a nonzero polynomial $m \in \mathfrak{F}[z^{-1}]$, where $\mathfrak{F}$ is a subfield of $\mathfrak{C}$ valuated by (2.25), we define the polynomial

$$m^* = m^+ m^{-\sim},$$

which belongs again to $\mathfrak{F}[z^{-1}]$ and satisfies the relation

$$\begin{aligned} m^= m &= m^{+=} m^{-=} m^+ m^- \\ &= m^{+=}\left(z^{-\partial m^-} m^{-\sim}\right) m^+\left(z^{-\partial m^-} m^{-\sim =}\right) \\ &= \left(m^+ m^{-\sim}\right)^= \left(m^+ m^{-\sim}\right) \\ &= (\omega m^*)^= (\omega m^*) \end{aligned}$$

for any unitary element $\omega \in \mathfrak{F}$. In particular, if $\mathfrak{F} = \mathfrak{R}$, the field of reals, the polynomial m^* is the so-called minimum-phase spectral factor of $m^= m$ as defined by Wiener, see e.g. [55; 58; 60; 64]. It may not always be so over other fields. Example:

$$m = 1 - 2z^{-1} - z^{-2} \in \mathfrak{Q}[z^{-1}], \quad m^* = -1 - 2z^{-1} + z^{-2} \in \mathfrak{Q}[z^{-1}]$$

while the minimum-phase spectral factor f of $m^= m$ is

$$f = \pm\left(1 - (1 - \sqrt{2})\, z^{-1}\right)\left(z^{-1} - (1 + \sqrt{2})\right),$$

an element of $\mathfrak{R}[z^{-1}]$.

It is clear that

$$\partial m^* \leqq \partial m$$

the equality holding if an only if $(m, z^{-1}) = 1$.

Now consider a nonzero polynomial matrix $M \in \mathfrak{F}_{l,m}[z^{-1}]$, where $\mathfrak{F}$ is a subfield of $\mathfrak{C}$, and let

$$M'^{=}M = E_1'^{=} \operatorname{diag}\{p_1^{=}p_1, p_2^{=}p_2, \ldots, p_s^{=}p_s, 0, \ldots, 0\} E_1 ,$$
$$MM'^{=} = E_2 \operatorname{diag}\{q_1q_1^{=}, q_2q_2^{=}, \ldots, q_sq_s^{=}, 0, \ldots, 0\} E_2'^{=}$$

be the canonical representations of $M'^{=}M$ and $MM'^{=}$. Then we define the matrix $M_1^* \in \mathfrak{F}_{s,m}[z^{-1}]$ by

$$\operatorname{diag}\{p_1^*, p_2^*, \ldots, p_s^*, 0, \ldots, 0\} E_1 = \begin{bmatrix} M_1^* \\ 0 \end{bmatrix}$$

and the matrix $M_2^* \in \mathfrak{F}_{l,s}[z^{-1}]$ by

$$E_2 \operatorname{diag}\{q_1^*, q_2^*, \ldots, q_s^*, 0, \ldots, 0\} = [M_2^* \quad 0] .$$

It is clear that the M_1^* and M_2^* satisfy the relations

$$M'^{=}M = (\Omega_1 M_1^*)'^{=} (\Omega_1 M_1^*) ,$$
$$MM'^{=} = (M_2^* \Omega_2)(M_2^* \Omega_2)'^{=}$$

for arbitrary unitary elements Ω_1, $\Omega_2 \in \mathfrak{F}_{s,s}$. In particular, if $\mathfrak{F} = \mathfrak{R}$, the field of reals, the M_1^* and M_2^* are the so-called minimum-phase matrix spectral factors of $M'^{=}M$ and $MM'^{=}$ respectively, see [55; 60; 64].

It is clear that

$$\operatorname{rank} M_1^* = s , \quad \operatorname{rank} M_2^* = s ,$$
$$\operatorname{mp} M_1^* = p_s^* , \quad \operatorname{mp} M_2^* = q_s^*$$

and

$$\partial M_1^* \leqq \partial M , \quad \partial M_2^* \leqq \partial M .$$

Example 2.5. Consider the matrix

$$M = \begin{bmatrix} 1 \\ \sqrt{2}\,(1 - z^{-1}) \end{bmatrix} \in \mathfrak{R}_{2,1}[z^{-1}] .$$

$$M'^{=}M = [-2z + 5 - 2z^{-1}] ,$$

$$MM'^{=} = \begin{bmatrix} 1 & \sqrt{2}\,(1 - z) \\ \sqrt{2}\,(1 - z^{-1}) & 2(1 - z^{-1})(1 - z) \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & 0 \\ \sqrt{2}\,(1 - z^{-1}) & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & \sqrt{2}\,(1 - z) \\ 0 & 1 \end{bmatrix}$$

and

$$s = 1\,,$$

$$p_1^{=} p_1 = -2z + 5 - 2z^{-1}\,, \quad q_1 q_1^{=} = 1\,.$$

Hence

$$p_1^* = z^{-1} - 2\,, \quad q_1^* = 1$$

and

$$M_1^* = [z^{-1} - 2]\,, \quad M_2^* = \begin{bmatrix} 1 \\ \surd 2\backslash(1 - z^{-1}) \end{bmatrix}$$

to within unitary elements. (Note that square roots are closed in $\surd$ and $\backslash$.)

Example 2.6. Consider the matrix

$$M = \begin{bmatrix} 1 & z^{-2} \\ 2 & z^{-3} \end{bmatrix} \in \mathfrak{R}_{2,2}[z^{-1}]\,.$$

Then

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

$$MM'^{=} = \begin{bmatrix} 1 & 0 \\ \dfrac{2 + z^{-1}}{2} & \dfrac{1}{2} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 2(-2z^{-1} + 5 - 2z) \end{bmatrix} \begin{bmatrix} 1 & \dfrac{2 + z}{2} \\ 0 & \dfrac{1}{2} \end{bmatrix}$$

and

$$s = 2\,,$$

$$p_1^{=} p_1 = 5\,, \qquad q_1 q_1^{=} = 2\,,$$

$$p_2^{=} p_2 = 5(-2z + 5 - 2z^{-1})\,, \quad q_2 q_2^{=} = 2(-2z^{-1} + 5 - 2z)\,.$$

Hence

$$p_1^* = \surd 5\,, \qquad q_1^* = \surd 2\,,$$

$$p_2^* = \surd 5\backslash(z^{-1} - 2)\,, \quad q_2^* = \surd 2\backslash(z^{-1} - 2)$$

and

$$M_1^* = \begin{bmatrix} \sqrt{5} & \dfrac{z^{-2}(1 + 2z^{-1})}{\sqrt{5}} \\ 0 & \dfrac{z^{-1} - 2}{\sqrt{5}} \end{bmatrix},$$

$$M_2^* = \begin{bmatrix} \sqrt{2} & 0 \\ \dfrac{2 + z^{-1}}{\sqrt{2}} & \dfrac{z^{-1} - 2}{\sqrt{2}} \end{bmatrix}$$

to within unitary elements.

Observe that $\partial M_1^* = \partial M$ while $\partial M_2^* < \partial M$.

2.4. Some Computational Aspects

In the previous section we have defined stable polynomials and certain matrix factorization which may be nontrivial to obtain. This section provides an effective check for stability and a simple iterative procedure to compute these factorizations and, in turn, it demonstrates the power and elegance of the algebraic approach.

A stable polynomial has been defined via formal long division. This is impractical from the computational point of view. There is a well-known check [2; 55] for stability of real polynomials. We now state without proof its generalization to polynomials defined over an arbitrary field with arbitrary valuation. To this effect we introduce the following convention. If $\alpha \in \mathfrak{F}$, then

$\bar{\alpha}$ = complex conjugate of α if $\mathfrak{F}$ is a subfield of $\mathfrak{C}$ valuated by (2.25),
$= \alpha$ otherwise, i.e. the macron is a void symbol.

Given a polynomial $a \in \mathfrak{F}[z]$ of degree $n \geqq 0$. Introduce polynomials

$$a_k = \alpha_{k,0} + \alpha_{k,1}z + \ldots + \alpha_{k,n-k}z^{n-k}$$

which are defined recursively by

$$za_{k+1} = a_k - \frac{\alpha_{k,0}}{\bar{\alpha}_{k,n-k}}\, a_k^{\sim}\,, \quad k = 0, 1, \ldots, n-1\,,$$

$$a_0 = a\,.$$

Then the polynomial a is stable with respect to a valuation $\mathscr{V}$ in $\mathfrak{F}$ if and only if

$$\mathscr{V}\left(\frac{\alpha_{k,0}}{\bar{\alpha}_{k,n-k}}\right) < 1\,, \quad k = 0, 1, \ldots, n-1.$$

The above recursive steps can be arranged in a table as follows.

(2.31)

$$
\begin{array}{ccccc|c|}
\alpha_{0,n} & \alpha_{0,n-1} & \cdots & \alpha_{0,1} & \alpha_{0,0} & \\
\bar{\alpha}_{0,0} & \bar{\alpha}_{0,1} & \cdots & \bar{\alpha}_{0,n-1} & \bar{\alpha}_{0,n} & \dfrac{\alpha_{0,0}}{\bar{\alpha}_{0,n}} \\
\alpha_{1,n-1} & \alpha_{1,n-2} & \cdots & \alpha_{1,0} & 0 & \\
\bar{\alpha}_{1,0} & \bar{\alpha}_{1,1} & \cdots & \bar{\alpha}_{1,n-1} & 0 & \dfrac{\alpha_{1,0}}{\bar{\alpha}_{1,n-1}} \\
 & & \cdots & & & \\
\alpha_{n-1,1} & \alpha_{n-1,0} & \cdots & 0 & 0 & \\
\bar{\alpha}_{n-1,0} & \bar{\alpha}_{n-1,1} & \cdots & 0 & 0 & \dfrac{\alpha_{n-1,0}}{\bar{\alpha}_{n-1,1}} \\
\alpha_{n,0} & 0 & \cdots & 0 & 0 &
\end{array}
$$

Example 2.7. Check whether the polynomials

$$m = -2z^2 + 2z - 0.5 \in \Re[z],$$

$$p = z^3 - z + 1 \in \Re[z],$$

are stable with respect to valuation (2.25) by the ordinary absolute value.

Table (2.31) for m becomes

$$
\begin{array}{cccc}
-2 & 2 & -\frac{1}{2} & \\
-\frac{1}{2} & 2 & 2 & \frac{1}{4} \\
-\frac{15}{8} & \frac{3}{2} & 0 & \\
\frac{3}{2} & -\frac{15}{8} & 0 & -\frac{4}{5} \\
-\frac{27}{40} & 0 & 0 &
\end{array}
$$

and $|\frac{1}{4}| < 1$, $|-\frac{4}{5}| < 1$ implies that m is stable.

Table (2.31) for p becomes

$$
\begin{array}{ccccc}
1 & 0 & -1 & 1 & \\
1 & -1 & 0 & 1 & 1 \\
0 & 1 & -1 & 0 & \\
-1 & 1 & 0 & 0 & ?
\end{array}
$$

and the computations must be stopped. Since $|1| \nless 1$, however, the p is not stable.

Example 2.8. Check whether the polynomials

$$a = 4 - 4z^{-1} + z^{-2} \in \Re[z^{-1}],$$
$$b = z^{-1} \in \Re[z^{-1}],$$

are stable with respect to (2.25).

We can write

$$a = z^{-d}a_0,$$

where

$$d = 0, \quad a_0 = 4 - 4z^{-1} + z^{-2},$$
$$\hat{a}_0 = z^2 a_0 = 1 - 4z + 4z^2.$$

Hence

$$\begin{array}{cccc} 4 & -4 & 1 & \\ 1 & -4 & 4 & \frac{1}{4} \\ \frac{15}{4} & -3 & 0 & \\ -3 & \frac{15}{4} & 0 & -\frac{12}{15} \\ \frac{27}{20} & 0 & 0 & \end{array}$$

implies that $\hat{a}_0$, and in turn a, is stable.

We write

$$b = z^{-d}b_0,$$

where

$$d = 1, \quad b_0 = 1,$$
$$\hat{b}_0 = z^0 b_0 = 1.$$

Since $d \neq 0$, b is not stable.

Example 2.9. Check if the polynomials

$$m = z - 0{\cdot}5 \in \Re[z],$$
$$p = z^n \in \Re[z], \quad n \text{ natural},$$

which are both stable with respect to valuation (2.25), are stable with respect to valuation (2.24).

Table (2.31) for m becomes

$$\begin{array}{ccc} 1 & -0{\cdot}5 & \\ -0{\cdot}5 & 1 & -0{\cdot}5 \\ 0{\cdot}75 & 0 & \end{array}$$

and since $\mathscr{V}(-0{\cdot}5) \nless 1$, the m is not stable.

Table (2.31) for p becomes

$$\begin{array}{l} 1\ 0\ \ldots\ 0\ 0 \\ 0\ 0\ \ldots\ 0\ 1\ 0 \\ 1\ 0\ \ldots\ 0\ 0 \\ 0\ 0\ \ldots\ 1\ 0\ 0 \\ \quad \ldots \\ 1\ 0\ \ldots\ 0\ 0 \\ 0\ 1\ \ldots\ 0\ 0\ 0 \\ 1\ 0\ \ldots\ 0\ 0 \end{array}$$

and since $\mathscr{V}(0) < 1$, the p is stable.

Example 2.10. Check the polynomial

$$a = \mathrm{i}z + (1 - \mathrm{i}) \in \mathfrak{C}[z]$$

for stability with respect to (2.25).

Table (2.31) reads

$$\begin{array}{lll} \mathrm{i} & 1-\mathrm{i} & \\ 1+\mathrm{i} & -\mathrm{i} & 1+\mathrm{i} \\ -\mathrm{i} & 0 & \end{array}$$

and since $|1 + \mathrm{i}| = 2$, the a is not stable.

Example 2.11. Check the polynomial

$$a = z^3 + 3z \in \mathfrak{Z}_5[z]$$

for stability (with respect to (2.24), of course, as no other valuation is possible).

Table (2.31) yields

$$\begin{array}{l} 1\ 3\ 0 \\ 0\ 3\ 1\ 0 \\ 1\ 3\ 0 \\ 3\ 1\ 0\ 3 \\ 2\ 0\ 0 \end{array}$$

and since $\mathscr{V}(3) = 1 \nless 1$, the a is not stable.

Example 2.12. Consider $\mathfrak{F} = \mathfrak{R}(w^{-1})$, the field of rational functions over $\mathfrak{R}$, and check whether the polynomial

$$a = z - w^{-1} \in \mathfrak{F}[z]$$

is stable with respect to valuation (2.26). Here

$$w^{-1} = 0 + 1w^{-1} + 0w^{-2} + \ldots \in \mathfrak{R}(w^{-1}).$$