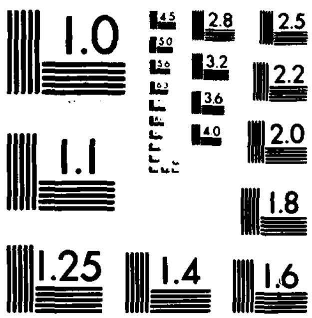
AD-A089 635 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
SPECTRAL THEORY OF MATRICES. I. GENERAL MATRICES.(U)
MAY 80 S FRIEDLAND DAAG29-75-C-0024
UNCLASSIFIED MRC-TSR-2082 NL

1 of 2



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL II

MRC Technical Summary Report #2082

SPECTRAL THEORY OF MATRICES

I. GENERAL MATRICES

Shmuel Friedland

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

May 1980

(Received February 12, 1980)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

SPECTRAL THEORY OF MATRICES
I. GENERAL MATRICES

Shmuel Friedland

Technical Summary Report #2082
May 1980

ABSTRACT

This is the first of our survey reports on the spectral theory of matrices. The report is self contained from the matrix point of view. The main subject of this paper is the theory of analytic similarity of matrices. This topic of matrix theory is very important in the study of systems of ordinary differential equations having singularities either in time or a parameter.

AMS (MOS) Subject Classifications: 15A21, 15A54, 34A25, 34E99

Key Words: Analytic similarity of matrices, canonical forms.

Work Unit Number 1 - Applied Analyis

Sponsored by the United States Army under Contract Nos. DAAG29-75-C-0024 and DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

In this survey paper we summarize the recent progress on the problem of analytic similarity of matrices. That is, given $n \times n$ complex values matrices $A(x) = (a_{ij}(x))$, $B(x) = (b_{ij}(x))$, whose entries $a_{ij}(x)$, $b_{ij}(x)$, $i,j = 1,\ldots,n$ are analytic functions in some domain Ω, when $B(x) = X(x)A(x)X^{-1}(x)$, where $X(x)$ and $X^{-1}(x)$ are analytic in x? What is the canonical form of A under the analytic similarity? These problems are related closely to the study of systems of ordinary differential equations having singularities either in time or a parameter. To make this survey paper to be self contained we had to recall some basic facts in theory of rings, functions of one and several compplex variables. Also we did repeat and extend some basic facts in theory of matrices in order to apply them for the analytic similarity problem and the related questions.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

SPECTRAL THEORY OF MATRICES

I. GENERAL MATRICES

Shmuel Friedland

Hebrew University at Jerusalem
and

Mathematics Research Center*
University of Wisconsin - Madison

*Sponsored by the United States Army under Contract Nos. DAAG29-75-C-0024 and DAAG29-80-C-0041.

CONTENTS

Introduction

Matrix theory is constantly gaining popularity in pure and applied mathematics and as well as in other branches of sciences. Perhaps the best book on the subject is the well known book of Gantmacher - "The Theory of Matrices" which was printed in the beginning of the fifties in U.S.S.R. . Since then the literature on the subject expanded enormously. There are recent books on the subject e.g., Berman-Plemmons [1979] which usually treat special topics in theory of matrices. It is my personal belief that the time is ripe for writing a comprehensive treatise on the main developments in theory of matrices.

I decided to write a series of survey reports on the most attractive and important subjects in theory of matrices - Spectral Theory. This paper is the first one in the series. It deals with general types of matrices. The climax of this paper are Section 1.29 - 1.36 which are dealing with the concept of analytic similarity of matrices. This subject arises naturally in the study of ordinary differential equations having singularities either in time or a parameter. See for example Wasow [1963], [1977], [1978] and the references therein. As the reader can see, the subject of analytic similarities of matrices is far from being completed. The main reason for the difficulties in this problem is the non-existence of a simple canonical form. Clearly the topic of the analytic similarity of matrices is a part of a more general algebraic problem of similarity of matrices over the integral domains. That is the reason I started the book with several sections on rings, domains and fields and their properties.

I tried to make this paper (and the following ones) to be self contained as much as possible from the matrix point of view. However, in dealing with some problems in matrix theory one needs to use various kinds of techniques - theory of functions of one and several complex variables, methods of algebraic geometry and non-linear analysis. Whenever these tools are used the reader is referred to appropriate references. The basic knowledge for this paper are basic results in matrix theory (e.g., a few first chapters in Gantmacher [1959]) and basic knowledge in function of one complex variable (e.g., Rudin [1974]). Since I tried to make these papers self contained from the matrix point of view, I did repeat some standard facts in theory of matrices as the Jordan canonical form. In that case I tried to make the exposition short and concise. Note that the problems appearing in the end of each section are an integral part of the paper and sometime they are used in the main text. Finally let me apologize to those authors whose results were not mentioned or improperly cited.

The four other papers in this survey series are planned to be as follows:

2. Cones, convex sets and norms,
3. Nonnegative matrices,
4. Symmetric and hermitian matrices,
5. Inverse eigenvalue problems.

Also these reports will be eventually collected to a book.

Special Notation

$\mathbf{D}$ - an integral domain.

$\mathbf{F}$ - a field, sometimes the division field of $\mathbf{D}$.

$\mathbf{C}$ - complex numbers.

$\mathbf{R}$ - real numbers.

$\mathbf{Z}$ - integers.

$\mathbf{Z}_+$ - non-negative integers.

BD - Bezout domain.

GCCD - greatest common divisor domain.

EDD - elementary divisor domain.

UFD - unique factorization domain.

PID - principal ideal domain.

I - ideal in $\mathbf{D}$.

$\mathbf{D}^n$ - a set of column vectors with n coordinates in $\mathbf{D}$.

Ω - a set of points in $\mathbf{C}^n$.

$H(\Omega)$ - the class of all analytic functions in Ω.

H_ζ - $H(\Omega)$ for $\Omega = \{\zeta\} \subseteq \mathbf{C}^n$.

$M(\Omega)$ - the quotient field of $H(\Omega)$, for connected sets Ω.

$\mathbf{D}[x_1,\ldots,x_n]$ - the ring of all polynomials in n variables with the coeffients in $\mathbf{D}$.

$a|b$ - a divides b.

$(a_1,\ldots,a_k)$ - the greatest common divisor of $a_1,\ldots,a_k$.

$\mathbf{M}$ - a D-module.

$[x^1,\ldots,x^k]$ - a D-module generated by the elements $x^1,\ldots,x^k \in \mathbf{M}$.

dim M - the dimension of a D-module.

$\mathrm{Hom}(\mathbf{M},\mathbf{N})$ - the set of homomorphismes $T : \mathbf{M} \to \mathbf{N}$.

V - a vector space over $\mathbf{F}$.

$L(\mathbf{V})$ - the set of linear operators $T : \mathbf{V} \to \mathbf{V}$.

$M_{mn}(\mathbf{D})$ - $m \times n$ matrices with entries in $\mathbf{D}$.

$M_n(D)$ - $M_{nn}(D)$.

$r(A)$ - the rank of $A \in M_{mn}(D)$.

$|A|$ - the determinant of $A \in M_n(\mathbf{D})$.

$\mathrm{tr}(A)$ - the trace of $A \in M_n(\mathbf{D})$.

$UM_n(D)$ - the set of all invertible matrices in $M_n(\mathbf{D})$.

$A \underset{\ell}{\sim} B$ - A and B are left equivalent.

$A \underset{r}{\sim} B$ - A and B are right equivalent.

$A \sim B$ - A and B are equivalent.

$\mathrm{diag}(A_1,\ldots,A_k)$, $\sum_{i=1}^{k} \oplus A_i$ - the direct sum of $A_1,\ldots,A_k$.

$A \otimes B$ - the tensor (Kronecker product).

A^t - the transpose of A.

$Q_{k,n}$ - totality of strictly increasing sequences of k integers chosen from $1,\ldots,n$.

$A[\alpha|\beta]$ - submatrix of A using row numbered α and columns numbered β.

$I, I(n), I_n$ - identity matrix of order n.

$\delta_k(A)$ - the k-th determinant invariant of A.

$i_k(A)$ - the k-th invariant factor of A.

$\eta(A), \eta(A,B)$ - the indices of $A(x)$ and $I_n \otimes A - B^t \otimes I_m$ respectively, $A \in M_m(H_0)$, $B \in M_n(H_0)$.

$\kappa_p(A)$, $\kappa_p(A,B)$ - the number of local invariant polynomials of degree p of A and $I_n \otimes A - B^t \otimes I_m$ respectively.

$r(A,B)$ - the rank of $I_n \otimes A - B^t \otimes I_m$, $A \in M_m(D)$, $B \in M_n(D)$.

$\nu(A,B)$ - the nullity of $I_n \otimes A - B^t \otimes I_m$.

$A \approx B$ - A and B are similar.

$A(x) \underset{s}{\sim} B(x)$ - $A(x)$ and $B(x)$ are strictly equivalent.

$A(x) \underset{s}{\approx} B(x)$ - $A(x)$ and $B(x)$ are strictly similar.

$A(x) \underset{a}{\approx} B(x)$ - $A(x)$ and $B(x)$ are analytically similar

$A(x) \underset{p}{\approx} B(x)$ - $A(x)$ and $B(x)$ are pointwise similar.

$A(x) \underset{r}{\approx} B(x)$ - $A(x)$ and $B(x)$ are rationally similar.

$\deg p$ - the degree of a polynomial p.

$C(p)$ - the companion matrix of p.

$H(m)$ - the matrix $(\delta_{(i+1)j})$ $i,j = 1,\ldots,m$.

$C(A,B)$ - the set of matrices X, $AX = XB$.

$C(A)$ - the set of matrices commuting with A.

$D(A,\rho)$ - a ρ neighborhood of A.

$Z_{ij}(A)$ - the components of A.

$\rho(A)$ - the spectral radius of A.

$\sigma(A)$, $\sigma_d(A)$ - the spectrum and the distinct spectrum of A.

$\sigma_d(A)$, $\sigma_{dp}(A)$ - the peripherial and the distinct peripheral spectrum of A.

index (λ) - the index of λ, $\lambda \in \sigma(A)$.

index (A) - the index of A.

$\|A\|$ - the ℓ_∞ norm of A.

$R(\lambda,A)$ - the resolvent of A.

$F(A_0,\ldots,A_{s-1})$ - Toeplitz upper triangular matrices.

$F(A_0,\ldots,A_{s-1}) \approx F(B_0,\ldots,B_{s-1})$ - strong similarity of Toeplitz upper triangular matrices.

o(1) - quantities which tend to zero as $r \to 0$.

0(1) - quantities which are uniformly bounded.

SPECTRAL THEORY OF MATRICES
I. GENERAL MATRICES

Shmuel Friedland

1.1 Rings, Domains and Fields.

Definition 1.1.1. A non-empty set S is said to be a ring if there are defined two operations addition and multiplication such that for all a,b,c in S

(1.1.2) $a + b \in S$;

(1.1.3) $a + b = b + a$ (the commutative law);

(1.1.4) $(a+b) + c = a + (b+c)$ (the associative law);

(1.1.5) there exists an element 0 in S such that $a + 0 = 0 + a = a$ for every $a \in S$;

(1.1.6) there exists an element $-a$ such that $a + (-a) = 0$;

(1.1.7) $ab \in S$;

(1.1.8) $a(bc) = (ab)c$, (the associative law);

(1.1.9) $a(b+c) = ab + ac$, $(b+c)a = ba + ca$ (the distributive laws).

S is said to have an identity element 1 if $a1 = 1a$ for all $a \in S$. S is called commutative if

(1.1.10) $ab = ba$, for all $a,b \in S$.

Note that the properties (1.1.3)-(1.1.9) imply that $a0 = 0a = 0$. It may happen that

(1.1.11) $ab = 0$

without a or b equal to 0. In that case we say that a and b are zero divisors.

Definition 1.1.12. $\mathbb{D}$ is called an integral domain if $\mathbb{D}$ is a commutative ring without zero divisors containing identity 1.

The classical example of an integral domain is $\mathbb{Z}$ - set of integers. In what follows we shall use frequently another example of integral domains.

Example 1.1.13. Let $\Omega \subseteq \mathbb{C}^n$ be a set of points. Denote by $H(\Omega)$ the class of all analytic functions $f(z_1,\ldots,z_n)$ which are analytic in the neighborhood of any point $\zeta = (\zeta_1,\ldots,\zeta_n) \in \Omega$. If Ω is open then we assume that f is defined only in Ω. In case that Ω consists of one point ζ denote $H(\Omega)$ by H_ζ.

Sponsored by the United States Army under Contract Nos. DAAG29-75-C-0024 and DAAG29-80-C-0041.

The properties of analytic functions imply that $H(\Omega)$ is an integral domain under the addition and multiplication of functions, provided that Ω is connected. We shall always assume that Ω is connected except where otherwise stated. The element 0 is the zero function and the identity is the function $f = 1$. For properties of analytic functions of one or several variables, consult for example Rudin [1974] and Gunning-Rossi [1965].

Let $a,b \in \mathbb{D}$. We say that a divides b $(a|b)$ if $b = ab_1$, $b_1 \in \mathbb{D}$. An element a is called invertible if $a|1$. For an invertible a denote by a^{-1} the element such that

$$aa^{-1} = a^{-1}a = 1 \quad . \tag{1.1.14}$$

In $H(\Omega)$ the invertible elements are only those functions which do not vanish at any point of Ω.

Definition 1.1.15. A field $\mathbb{F}$ is an integral domain $\mathbb{D}$ such that any non-zero element is invertible.

The familiar examples of fields are the set of rational numbers $\mathbb{Q}$, the set of real numbers $\mathbb{R}$, and the set of complex numbers $\mathbb{C}$. Given an integral domain $\mathbb{D}$ there is a standard way to obtain the field of its quotients. That is we consider the set of quotients $\frac{a}{b}$, $b \neq 0$ such that

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \quad \left(\frac{a}{b}\right)\left(\frac{c}{d}\right) = \frac{ac}{bd}, \quad b,d \neq 0 \quad . \tag{1.1.16}$$

Thus the set of the rational numbers $\mathbb{Q}$ is the quotient field of the integers $\mathbb{Z}$.

Definition 1.1.17. Let $\Omega \subseteq \mathbb{C}^n$. Denote by $M(\Omega)$ the quotient field of $H(\Omega)$. By M_ζ denote the quotient field of H_ζ. That is $M(\Omega)$ is the set of meromorphic functions in Ω.

Definition 1.1.18. By $\mathbb{D}[x_1,\ldots,x_n]$ denote the ring of all polynomials in n variables with the coefficients in $\mathbb{D}$

$$p(x_1,\ldots,x_n) = \sum_{|\alpha| \leq m} a_\alpha x^\alpha \quad , \tag{1.1.19}$$

$$\alpha = (\alpha_1,\ldots,\alpha_n) \in \mathbb{Z}_+^n, \quad |\alpha| = \sum_{i=1}^{n} \alpha_i, \quad x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n} \quad . \tag{1.1.20}$$

Thus $\mathbb{C}[x]$ is the ring of polynomials in one variable with complex coefficients. We say that the degree of $p(x_1,\ldots,x_n)$ (deg p) is m if there exists $a_\alpha \neq 0$ such that $|\alpha| = m$. A polynomial p is called homogeneous if $a_\alpha = 0$ for $|\alpha| < \deg p$. It is a standard fact that $\mathbb{D}[x_1,\ldots,x_n]$ is an integral domain (see Problems 1.1.22-1.1.23).

Problems

(1.1.21) Let $C[a,b]$ denote the set of real valued continuous functions on the interval $[a,b]$. Show that $C[a,b]$ is a commutative ring with an identity and zero divisors.

(1.1.22) Prove that $\mathbb{D}[x]$ is an integral domain

(1.1.23) Prove that $\mathbb{D}[x_1,\ldots,x_n]$ is an integral domain. (Use the previous problem and the identity $\mathbb{D}[x_1,\ldots,x_n] = (\mathbb{D}[x_1,\ldots,x_{n-1}])\,[x_n]$.

(1.1.24) Let $p(x_1,\ldots,x_n) \in \mathbb{D}[x_1,\ldots,x_n]$. Show that

(1.1.25) $$p = \sum_{\alpha \leq \deg p} p_\alpha ,$$

where either $p_\alpha = 0$ or p_α is a homogeneous polynomial of degree α. Moreover, if $m = \deg p \geq 1$ then $p_m \neq 0$. The polynomial $p_m(x_1,\ldots,x_n)$ is called the principal part of p and is denoted by p_π. (In case that p is a constant polynomial $p_\pi = p$.)

(1.1.26) Let $p,q \in \mathbb{D}[x_1,\ldots,x_n]$. Show

(1.1.27) $$(pq)_\pi = p_\pi q_\pi \ .$$

1.2 Bezout Domains

An element $d \in \mathbf{D}$ is called the greatest common divisor (g.c.d.) of $a_1,\ldots,a_n$ if $d|a_i$, $i = 1,\ldots,n$, and for any d' such that $d'|a_i$, $i = 1,\ldots,n$ we have $d'|d$. We denote $d = (a_1,\ldots,a_n)$ if g.c.d. of $a_1,\ldots,a_n$ exists. Clearly $(a_1,\ldots,a_n)$ is unique up to a multiple of an invertible element. The elements $a_1,\ldots,a_n$ are called co-prime if $(a_1,\ldots,a_n) = 1$.

Definition 1.2.1. $\mathbf{D}$ is called a greatest common divisor domain (G$\mathbb{C}$ $\mathbf{D}$ $\mathbf{D}$) if any two elements in $\mathbf{D}$ have g.c.d.

A trivial example of G$\mathbb{C}\mathbf{D}\mathbf{D}$ is $\mathbf{Z}$.

Definition 1.2.2. A subset $I \subseteq \mathbf{D}$ is called ideal if for any $a,b \in I$ and $p,q \in \mathbf{D}$ the element $pa + qb$ belongs to I.

In $\mathbf{Z}$ any ideal is the set of all numbers divisible by an integer $k \neq 0$. In $H(\Omega)$, the set of functions which vanish on a prescribed set $U \subseteq \Omega$, i.e.

$$I(U) = \{f | f(\zeta) = 0,\ \zeta \in U,\ f \in H(\Omega)\} \tag{1.2.3}$$

form an ideal. An ideal I is called prime if the fact $ab \in I$ implies that either $a \in I$ or $b \in I$. In $\mathbf{Z}$ the prime ideals are the set of all integers divided by a certain prime p. An ideal I is said to be maximal if the only ideals which contain I are I and $\mathbf{D}$. I is called finitely generated if there exists k elements (generators) $p_1,\ldots,p_k \in I$ such that any $i \in I$ is of the form

$$i = a_1p_1 + a_2p_2+\cdots+a_kp_k \tag{1.2.4}$$

for some $a_1,\ldots,a_k$ in $\mathbf{D}$. For example, in $\mathbf{D}[x,y]$ the set of all polynomials $p(x,y)$ such that

$$p(0,0) = 0 \tag{1.2.5}$$

is an ideal which is generated by the polynomials x and y. An ideal I is called principal ideal if it is generated by one element p.

Definition 1.2.6. $\mathbf{D}$ is called a Bezout domain ($\mathbf{B}$ $\mathbf{D}$) if any two elements $a,b \in \mathbf{D}$ have a g.c.d. (a,b) such that

(1.2.7) $$(a,b) = pa + qb$$

for some $p,q \in \mathbb{D}$.

It is easy to show by induction that for $a_1,\ldots,a_n \in \mathbb{BD}$ one has

(1.2.8) $$(a_1,\ldots,a_n) = \sum_{i=1}^{n} p_i a_i \quad .$$

We give another characterization of $\mathbb{BD}$.

Lemma 1.2.9. An integral domain is a Bezout domain if and only if any finitely generated ideal is principal.

Proof. Assume that an ideal I of $\mathbb{BD}$ is generated by $a_1,\ldots,a_n$. Then (1.2.8) implies that $(a_1,\ldots,a_n) \in I$. Clearly $(a_1,\ldots,a_n)$ is the generator of I. Assume now that any finitely generated ideal of $\mathbb{D}$ is principal. For a given $a,b \in \mathbb{D}$ let I be generated by a and b. Let d be the generator of I. So

(1.2.10) $$d = pa + qb \quad .$$

Also $d|a$ and $d|b$ since d generates I. Obviously if $d'|a$ and $d'|b$ then (1.2.10) implies that $d'|d$. Thus $d = (a,b)$. So $\mathbb{D}$ is $\mathbb{BD}$.

□

Consider the ideal $I \subseteq \mathbb{D}[x,y]$ given by (1.2.5). Obviously $(x,y) = 1$. As $1 \notin I$, I is not principal. Since x,y generate I we showed that $\mathbb{D}[x,y]$ is not $\mathbb{BD}$. In particular $\mathbb{F}[x_1,\ldots x_n]$ are not $\mathbb{BD}$ for $n \geq 2$. The same arguments show that $H(\Omega)$ are not $\mathbb{BD}$ in case that $\Omega \subseteq \mathbb{C}^n$ for $n \geq 2$. It is a standard fact that $\mathbb{F}[x]$ is a Bezout domain (e.g. Lang, [1967]) and we shall demonstrate it later on. For $\Omega \subseteq \mathbb{C}$, $H(\Omega)$ is $\mathbb{BD}$. This result is implied by the following interpolation of open sets in $\mathbb{C}$ (e.g. Rudin [1974, Thrs 15.11,15.15]).

Theorem 1.2.11. Suppose that Ω is an open set in $\mathbb{C}$, $A \subset \Omega$, A has no limit point in Ω, and to each $\zeta \in A$ then are associated a non-negative integer $m(\zeta)$ and complex numbers $w_{n,\zeta}$, $0 \leq n \leq m(\zeta)$. Then there exists $f \in H(\Omega)$ such that

$$f^{(n)}(\zeta) = n!\, w_{n,\zeta}, \quad \zeta \in A, \ 0 \leq n \leq m(\zeta) \quad .$$

Moreover of all $w_{n,\zeta} = 0$ then it is possible to choose f such that all zeros of f are in A and f has a zero of order $m(\zeta)$ at each $\zeta \in A$.

Theorem 1.2.12. *Let* Ω *be an open connected set in* $\mathbb{C}$.

For given functions $a,b \in H(\Omega)$ *there exists a function* $p \in H(\Omega)$ *such that*

$$c = pa + b \quad , \tag{1.2.13}$$

and

$$a = a_1 c,\ b = b_1 c,\ a_1, b_1 \in H(\Omega) \quad . \tag{1.2.14}$$

That is $c = (a,b)$.

Proof.

In case that a or b are zero functions choose $p = 1$ and the theorem trivially holds.

Assume that $ab \neq 0$. Let A be the set of common zeros of $a(z)$ and $b(z)$. Thus A is at most countable. For each $\zeta \in A$ let $m(z)$ be the common multiplicity of zero in $a(z)$ and $b(z)$ at $z = z$. *Let* $f(z) \in H(\Omega)$ *whose only zeros* ζ *are in* A *such that* ζ *has* multiplicity $m(z)$. The existence of such a function implied by Theorem 1.2.11. Put

$$a = \hat{a}f,\ b = \hat{b}f,\ \hat{a},\hat{b} \in H(\Omega) \quad .$$

Thus $\hat{a}$ and $\hat{b}$ do not have common zeros. Let B be the set of zeros of $\hat{a}$ such that $\zeta \in B$ has the multiplicity $n(\zeta)$. According to Theorem 1.2.11 there exists $g \in H(\Omega)$ such that

$$\frac{d^k}{dz^k}(e^{g(z)})\Big|_{z=\zeta} = \hat{b}^{(k)}(\zeta),\ k = 0,\ldots,n(\zeta)-1,\ \zeta \in B \quad . \tag{1.2.15}$$

Since $b(\zeta) \neq 0$ for $\zeta \in B$. Put

$$p = (e^g - \hat{b})|\hat{a},\ c = fe^g \tag{1.2.16}$$

and (1.2.13) holds. So $c|a$, $c|b$ and in view of (1.2.13) $c = (a,b)$.

Corollary 1.2.17. *Let* $\Omega \subseteq \mathbb{C}$. *Then* $H(\Omega)$ *is a Bezout domain*.

Proof. Let $a,b \in H(\Omega)$. By the definition of $H(\Omega)$ there exists an open domain $\Omega_0 \supseteq \Omega$ such that $a,b \in H(\Omega_0)$. (In case that Ω is open $\Omega_0 = \Omega$.) Consider the functions $p,c \in H(\Omega_0)$ which are constructed in Theorem 1.2.12. Clearly $p,c \in H(\Omega)$ and (1.2.13)-(1.2.14) implies that $H(\Omega)$ is $\mathbb{B}\,\mathbb{D}$.

□

Problems

(1.2.18) Let $a,b,c \in \mathbb{B}\,\mathbb{D}$. Assume that $(a,b) = 1$ and $(a,c) = 1$. Shows that $(a,bc) = 1$.

(1.2.19) Let I be a prime ideal in $\mathbb{D}$. Show that $\mathbb{D}/I$ (that is the set of all cosets of the form $I + a$) is an integral domain.

(1.2.20) Let I be an ideal in $\mathbb{D}$. Denote by $I(p)$ the following set

$$I(p) = \{a \mid a = bp + q,\ b \in \mathbb{D},\ q \in I\} .$$

Show that $I(p)$ is an ideal. Prove that I is a maximal ideal in $\mathbb{D}$ if and only if for any $p \notin I$, $I(p) = \mathbb{D}$.

(1.2.21) Show that an ideal I is maximal if and only if $\mathbb{D}/I$ is a field.

1.3 $\mathbb{UFD}$, $\mathbb{PID}$ and $\mathbb{ED}$ domains

A non-zero, non-invertible element $p \in \mathbb{D}$ is called *irreducible* (*prime*) if the only elements which divide p are p itself and the invertible elements. For example a positive integer $p \in \mathbb{Z}$ is irreducible if and only if p is prime. A linear polynomial is irreducible in $\mathbb{D}[x_1,\dots,x_n]$. For $\Omega \subseteq \mathbb{C}$ it is possible to determine all irreducible elements in $H(\Omega)$.

Lemma 1.3.1. *Let $\Omega \subseteq \mathbb{C}$. Then all the irreducible elements of $H(\Omega)$ are of the form $z - \zeta$, $\zeta \in \Omega$.*

Proof. Let $f \in H(\Omega)$ be not invertible. Then there exists $\zeta \in \Omega$ such that $f(\zeta) = 0$. So $z - \zeta$ divides $f(z)$. Therefore the only irreducible elements (up to multiplication by the invertible elements) are $z - \zeta$, $\zeta \in \Omega$. Clearly $\frac{z-\eta}{z-\zeta}$ is analytic in Ω if and only if $\eta = \zeta$ ($\zeta \in \Omega$). This proves the lemma.

□

In particular if $\zeta \in \mathbb{C}$ then H_ζ has one prime element $z - \zeta$.

Definition 1.3.2. $\mathbb{D}$ *is called unique factorization domain* ($\mathbb{UFD}$) *if any non-zero, non-invertible element* a *can be factored as a product of irreducible elements*

$$a = p_1 \cdots p_r \tag{1.3.3}$$

and those primes are uniquely determined up to invertible factors.

Again the ring of integers is obviously $\mathbb{UFD}$. Another example of unique factorization domain is $\mathbb{F}[x_1,\dots,x_n]$ for any n. (e.g. Lang [1967]).

Lemma 1.3.4. *Let $\Omega \subseteq \mathbb{C}$ be an open set. Then $H(\Omega)$ is not unique factorization domain.*

Proof. Let $a(z) \in H(\Omega)$ be a non-zero function which has an infinite number of zeros in Ω. Such functions exist in view of Theorem 1.2.10. If (1.3.3) was holding then according to Lemma 1.3.1 $a(z)$ would have a finite number of zeros which contradicts the choice of a.

□

A straightforward consequence of Lemma 1.3.4 that for an open set $\Omega \subseteq \mathbb{C}^n$, $H(\Omega)$ is not $\mathbb{UFD}$. See Problem 1.3.17.

Definition 1.3.5. $\mathbb{D}$ *is called a principal ideal domain* ($\mathbb{PID}$) *if any ideal of* $\mathbb{D}$ *is principal.*

The standard examples of **PID** are the ring of integers and the ring of polynomials in one variable over a field. It is known that any **PID** is **UFD** (e.g. Lang [1967] or v.d. Waerden [1959]). Thus $H(\Omega)$ for an open set Ω is not **PID**. A very useful and even more restrictive class of **PID** is the class of Euclidean domains.

Definition 1.3.6. $\mathbb{D}$ is called Euclidean domain ($\mathbb{ED}$) if for every $a \neq 0$ there is defined a non-negative integer $d(a)$ such that

(1.3.7) for all $a,b \in \mathbb{D}$, $ab \neq 0$ $d(a) \le d(ab)$;

(1.3.8) for any $a,b \in \mathbb{D}$, $ab \neq 0$, there exists $t,r \in \mathbb{D}$ such that

(1.3.9) $a = tb + r$, where either $r = 0$ or $d(r) < d(b)$.

The ring of integers is $\mathbb{ED}$ with

(1.3.10) $d(a) = |a|$.

The ring $\mathbb{F}[x]$ is $\mathbb{ED}$ with $d(p)$ - the degree of the polynomial $p(x)$. It is well known that any $\mathbb{ED}$ is $\mathbb{PID}$. Indeed, consider an ideal $I \subseteq \mathbb{ED}$. Choose $a \in I$ with the minimal $d(a)$. In view of (1.3.9) $a|b$ for any $b \in I$. Thus a generates I. This argument show that $\mathbb{F}[x]$ is $\mathbb{PID}$. We also have

Lemma 1.3.11. Let Ω be a compact connected domain in $\mathbb{C}$. Then $H(\Omega)$ is $\mathbb{ED}$. Here $d(a)$ is the number of zeros of $a(z)$ in Ω counted with their multiplicities.

Proof. Let $a \neq 0$. Then $a(z)$ must have a finite number of zeros in Ω. Otherwise there will be a sequence of zeros $\{\zeta_k\}$ of $a(z)$ which converge to some point $\zeta \in \Omega$. Since a is analytic in the neighborhood of ζ a is the zero function in the neighborhood. The connectivity of Ω implies $a \equiv 0$ contrary to our assumptions. Let $p_a(z)$ be a polynomial such that $\frac{a}{p_a}$ does not vanish in Ω. By the definition $d(a) = d(p_a)$. Let $a,b \in H(\Omega)$ $ab \neq 0$. Since $\mathbb{C}[z]$ is $\mathbb{ED}$

(1.3.12) $$p_a(z) = t(z)p_b(z) + r(z), \quad r = 0 \text{ or } d(r) < d(p_b) \ .$$

So

(1.3.13) $$a = \left(\frac{a}{p_a}\right)p_a = \left(\frac{a}{p_a}\frac{p_b}{b}t\right) + \left(\frac{a}{p_a}\right)r,\ d\left(\frac{ar}{p_a}\right) = d(r) \quad .$$

$\square$

Let $a_1, a_2 \in \mathbb{ED}$. Assume that $d(a_1) \geq d(a_2)$. The Euclid algorithm consists of a sequence $\{a_i\}$ which is defined recursively as follows:

(1.3.14) $$a_i = t_i a_{i+1} + a_{i+2},\ a_{i+2} = 0 \ \text{ or } \ d(a_{i+2}) < d(a_{i+1}) \quad .$$

Since $d(a) \geq 0$ the Euclid algorithm must terminate. That is

(1.3.15) $$a_1, \ldots, a_k \neq 0, \quad a_{k+1} = 0 \quad .$$

It is a standard fact that

(1.3.16) $$(a_1, a_2) = a_k \quad .$$

See Problem 1.3.18. That is the g.c.d. of a_1 and a_2 can be found explicitly in a finite number of steps. While for an open set $\Omega \subseteq \mathbb{C}$ the construction of (a,b) may involve infinite number of steps, i.e. limits, which appear in the proof of Theorem 1.2.11.

Problems

(1.3.17) Let $\Omega \subseteq \mathbb{C}^n$ be an open set in $\mathbb{C}^n$. Construct a function f depending on one variable in Ω which has an infinite number of zeros in $\Omega(f \not\equiv 0)$. Prove that f cannot be decomposed to a finite product of irreducible elements. That is $H(\Omega)$ is not $\mathbb{UFD}$.

(1.3.18) Consider the equality (1.3.9) for $r \neq 0$. Show that $(a,b) = (b,r)$. Using this result prove (1.3.16).

1.4 Factorizations in $\mathbb{D}[x]$.

Let $\mathbb{F}$ be the field of quotients of $\mathbb{D}$. Assume that $p(x) \in \mathbb{D}[x]$. Suppose that

$$p(x) = p_1(x)p_2(x), \; p_i(x) \in \mathbb{F}[x], \; i = 1,2 \; . \tag{1.4.1}$$

Our problem is to find conditions which yield that $p_i(x) \in \mathbb{D}[x]$, $i = 1,2$. Clearly that if (1.4.1) holds then we can multiply $p_1(x)$ by $a \in \mathbb{F}$, $a \neq 0$, and divide $p_2(x)$ by a. So we must compose some normalizations on p_1 and p_2. Clearly for any $q(x) \in \mathbb{F}[x]$

$$q(x) = p(x)/a, \; p(x) \in \mathbb{D}[x], \; a \in \mathbb{D} \; . \tag{1.4.2}$$

In case that $\mathbb{D}$ is a $G\mathbb{C}\mathbb{D}\mathbb{D}$ the decomposition (1.4.2) can be unique (up to multiplication of invertible elements in $\mathbb{D}$).

Definition 1.4.3. Let $p(x)$ be a polynomial of degree m in $\mathbb{D}[x]$

$$p(x) = a_0x^m + \cdots + a_m \; . \tag{1.4.4}$$

The polynomial $p(x)$ is called normalized if $a_0 = 1$. If $\mathbb{D}$ is the greatest common divisor domain then let $c(p) = (a_0,\ldots,a_m)$. The polynomial $p(x)$ is said to be primitive if $c(p) = 1$.

The following result is obvious.

Lemma 1.4.5. Let $\mathbb{F}$ be the field of quotients of $G\mathbb{C}\mathbb{D}\mathbb{D}$. Then for any $q(x) \in \mathbb{F}[x]$ we have the decomposition (1.4.2) when $c(p)$ and a are co-prime. The polynomial $p(x)$ is uniquely defined up to an invertible factor in $G\mathbb{C}\mathbb{D}\mathbb{D}$. Moreover $q(x)$ decomposes to

$$q(x) = \frac{b}{a} r(x), \quad r(x) \in G\mathbb{C}\mathbb{D}\mathbb{D}[x], \quad b,a \in G\mathbb{C}\mathbb{D}\mathbb{D} \tag{1.4.6}$$

when $(b,a) = 1$ and $r(x)$ is primitive.

The crucial step in proving that $\mathbb{U}\mathbb{F}\mathbb{D}[x]$ is $\mathbb{U}\mathbb{F}\mathbb{D}$ is the Gauss lemma.

Lemma 1.4.7. Let $p(x),q(x) \in \mathbb{U}\mathbb{F}\mathbb{D}[x]$ be primitive then $p(x)q(x)$ is primitive.

Using this lemma one easily gets

Lemma 1.4.8. Let $p(x) \in \mathbb{U}\mathbb{F}\mathbb{D}[x]$ be primitive. Assume that $p(x)$ is irreducible in $\mathbb{F}[x]$, where $\mathbb{F}$ is the quotient field of $\mathbb{U}\mathbb{F}\mathbb{D}$. Then $p(x)$ is irreducible in $\mathbb{U}\mathbb{F}\mathbb{D}[x]$. See Lang [1967] and Problems 1.4.17-1.4.18.

Thus any $p(x) \in \mathbf{UFD}[x]$ has the unique decomposition

$$p(x) = a\, q_1(x) \cdots q_s(x) \tag{1.4.9}$$

where $q_1(x),\ldots,q_s(x)$ are primitive and irreducible in $\mathbb{F}[x]$ and a has the decomposition (1.3.3). I fact (1.4.9) is the decomposition of $p(x)$ to irreducible factors in $\mathbb{F}[x]$. Thus we proved (e.g. Lang [1967]).

Theorem 1.4.10. $\mathbf{UFD}[x]$ is a unique factorization domain.

Normalization 1.4.11. Let $\mathbb{F}$ be a field and assume that $p(x)$ is a normalized non-constant polynomial in $\mathbb{F}[x]$. Let (1.4.9) be a decomposition of $p(x)$ to irreducible factors. Then normalize the decomposition (1.4.9) by the assumption that $q_1(x),\ldots,q_s(x)$ are normalized irreducible polynomials in $\mathbb{F}[x]$. (This implies that $a = 1$.)

It is not difficult to show that Lemmas 1.4.7-1.4.8 yield

Theorem 1.4.12. Let $p(x)$ be a normalized non-constant polynomial in $\mathbf{UFD}[x]$. Let (1.4.9) be a normalized decomposition of $p(x)$ in $\mathbb{F}[x]$, where $\mathbb{F}$ is the quotient field $\mathbf{UFD}$. Then each $q_j(x)$ is an irreducible polynomial belonging to $\mathbf{UFD}[x]$.

See Problem 1.4.19 It turns out that Theorem 1.4.12 holds for any $H(\Omega)$, $\Omega \subseteq \mathbb{C}^n$.

Theorem 1.4.13. Let $p(x)$ be a normalized non-constant polynomial in $H(\Omega)[x]$. Let (1.4.9) be a normalized decomposition in $M[x]$, where M is the field of meromorphic functions in Ω. Then each $q_j(x)$ is an irreducible polynomial in $H(\Omega)$.

Proof. By the definition of $H(\Omega)$ we may assume that $p(x) \in H(\Omega_0)$ for some open domain Ω_0 containing Ω. So $q_j(x) \in M(\Omega_0)[x]$, $j = 1,\ldots,s$. Let

$$q(x,z) = x^t + \sum_{r=1}^{t} \frac{\alpha_{t-r}(z)}{\beta_{t-r}(z)} x^{t-r}, \quad \alpha_i(z),\beta_i(z) \in H(\Omega_0) \quad , i = 1,\ldots,t \quad . \tag{1.4.14}$$

Thus $q(x,z)$ is analytic on $\Omega_0 - \Gamma$ where Γ is given by

$$\Gamma = \{z \mid z \in \Omega_0, \prod_{r=1}^{t} \beta_r(z) = 0\} \quad . \tag{1.4.15}$$

It is known that Γ is a closed set of zero measure in $\mathbb{C}^n$. See for example Gunning-Rossi [1965].

Consider $x_1(z),\dots,x_t(z)$ the roots of

$$q(x,z) = 0 \quad . \tag{1.4.16}$$

Clearly $x_1(z),\dots,x_t(z)$ are well defined functions on $\Omega_0 - \Gamma$. Suppose that each $x_k(z)$ is bounded when z tends to any $\zeta \in \Gamma$. Then each $\frac{\alpha_j(z)}{\beta_j(z)}$ is bounded on any compact domain in Ω_0. This would imply that $\alpha_j(z)/\beta_j(z) \in H(\Omega_0)$, $j=1,\dots,t$, i.e. $q(x,z) \in H(\Omega_0)$ (e.g. Gunning-Rossi [1965]). Thus, if $q(x,z) \notin H(\Omega_0)$ there exists a sequence $z^{(k)} \to \zeta \in \Gamma$ such that $x_r(z^{(k)}) \to \infty$. To this end assume that $q_j(x) \notin H(\Omega_0)$ for some j. Then, we have $q_j(x_r(z^{(k)})) = 0$, $z^{(k)} \to \zeta \in \Omega_0$, $x(z^{(k)}) \to \infty$, as $k \to \infty$.

The assumption that $p(x) \in H(\Omega_0)$ implies that all the coefficients of the normalized *polynomial* $p(x,z)$ are bounded in the neighborhood of ζ, so the roots of the equation $p(x(z),z) = 0$ are bounded in the neighborhood of ζ. This contradicts the equality $p(x(z^{(k)}),z^{(k)}) = 0$. That is $q_j(x,z) \in H(\Omega_0)$ for $j=1,\dots,s$.

Problems

(1.4.17) Let $p(x)$ be given by (1.4.4) and put $q(x) = b_0x^n+\cdots b_n$, $r(x) = p(x)q(x) = c_0x^{m+n}+\cdots+c_{m+n}$. Assume that $p(x),q(x) \in \mathrm{UFD}[x]$. Let π be an irreducible element in UFD such that $\pi|a_i$, $i=0,\dots,\alpha$, $\pi|b_0$, $j=0,\dots,\beta$, $\pi|c_k$, $k = 0, \alpha+\beta+2$. Then either $\pi|a_{\alpha+1}$ or $\pi|b_{\beta+1}$.

(1.4.18) Prove that if $p(x),q(x) \in \mathrm{UFD}[x]$, then $c(pq) = c(p)c(q)$.

Deduce from the above inequality Lemma 1.4.7. Also if $p(x)$ and $q(x)$ are normalized polynomials then $p(x)q(x)$ is primitive.

(1.4.19) Prove Theorem 1.4.12.

(1.4.20) Using the equality

$$(D[x_1,\dots,x_{n-1}])[x_n] = D[x_1,\dots,x_n]$$

prove that $\mathrm{UFD}[x_1,\dots,x_n]$ is UFD. Deduce that $F[x_1,\dots,x_n]$ is UFD.

1.5 Elementary Divisor Domain

Definition 1.5.1. D *is called elementary divisor domain* (EDD) *if for any three elements* $a,b,c \in D$ *there exist* $p,q,x,y \in D$ *such that*

$$(a,b,c) = (px)a + (py)b + (qy)c \quad . \tag{1.5.2}$$

By letting $c = 0$ we obtain that (a,b) is a linear combination of a and b. So elementary divisor domain is Bezout domain.

Theorem 1.5.3. *Let* D *be a principal ideal domain. Then* D *is elementary divisor domain.*

Proof. Without the loss of generality we may assume that $abc \neq 0$, $(a,b,c) = 1$. Let $(a,c) = d$. Since D is UFD (e.g. Lang [1967]) decompose $a = a'a''$, where in the prime decomposition (1.3.3) of a, a' contains all the irreducible elements of a which appear in the decomposition of d to irreducible factors. So

$$a = a'a'',\ (a',a'') = 1,\ (a',c) = (a,c),\ (a'',c) = 1 \quad , \tag{1.5.4}$$

and if a',f are not co-prime then c,f are not co-prime. Thus, there exists q and α such that

$$b - 1 = -qc + \alpha a'' \quad . \tag{1.5.5}$$

Let $d' = (a,b+qc)$. The above equality implies that $(d',a'') = 1$. Suppose that d' is not co-prime with a'. So, there exists a non-invertible f such that f divides d' and a'. According to (1.5.4) $(f,c) = f'$ and f' is not invertible. Thus $f'|b$ which implies that f' divides $a,c,$ and b. This contradicts our assumption $(a,b,c) = 1$. So $(d',a') = 1$ which means $(d',a) = 1$. Therefore there exists $x,y \in D$ such that $xa + y(b+qc) = 1$.

This establishes (1.5.2) with $p = 1$.

Theorem 1.5.6. Let $\Omega \subseteq \mathbb{C}$. Then $H(\Omega)$ is elementary divisor domain.

Proof. Given $a,b,c, \in H(\Omega)$ we may assume that $a,b,c, \in H(\Omega_0)$ for some open set $\Omega_0 \supseteq \Omega$. According to Theorem 1.2.12

$$(1.5.7) \qquad (a,b,c) = (a,(b,c)) = xa + (b,c) = xa + (b+yc) \quad .$$

Problems

(1.5.8) D is called adequate if for any $a,c \in D$, $ac \neq 0$, (1.5.4) holds. Use the proof of Theorem 1.5.6 to show that any adequate BD domain is EDD.

(1.5.9) Prove that $H(\Omega)$, $\Omega \subseteq \mathbb{C}$, is an adequate domain (Helmer [1943]).

1.6 Algebraically closed fields

Definition 1.6.1. A field F is called algebraically closed if any polynomial $p(x) \in F[x]$ of the form (1.4.4) splits to linear factors in F

$$p(x) = a_0 \prod_{i=1}^{m} (x-\xi_i), \quad \xi_i \in F, \ i=1,\dots,m, \ a_0 \neq 0 \ . \tag{1.6.2}$$

The classical example of an algebraically closed field is the field of the complex numbers C. The field of the real numbers R is not algebraically closed.

Definition 1.6.3. Let K and F be fields. Assume that $K \subseteq F$. Then K is called an extension field of F. K is called a finite extension of F if K is a finite dimensional vector space over F.

Thus C is a finite extension of R of the dimension 2. It is known (e.g. Lang [1967]).

Theorem 1.6.4. Let $p(x) \in F[x]$. Then there exists a finite extension K of F such that $p(x)$ splits to linear factors in $K[x]$.

The classical Weierstrass preparation theorem in two complex variable is an explicit example of the above theorem. We state the Weierstrass preparation theorem in the form needed later. (See for example Gunning-Rossi [1965].)

Theorem 1.6.5. Let H_0 be the ring of analytic functions in one variable analytic in the neighborhood of the origin. Let $p(\lambda) \in H_0[\lambda]$ be a normalized polynomial of degree n

$$p(\lambda,z) = \lambda^n + \sum_{i=1}^{n} a_i(z)\lambda^{n-i}, \quad a_i(z) \in H_0, \ i=1,\dots,n \ . \tag{1.6.6}$$

Then there exists a positive integer $s \leq n!$, such that

$$p(\lambda,w^s) = \prod_{i=1}^{n} (\lambda-\lambda_i(w)), \quad \lambda_i(w) \in H_0, \ i=1,\dots,n \ . \tag{1.6.7}$$

Thus, in that particular case, the extension field of H_0 is the set of multivalued functions which are analytic in $z^{1/s}$ in the neighborhood of the origin. Algebraically speaking, K is $H_0[w]$ together with the identity

$$w^s = z \quad . \tag{1.6.8}$$

The vector dimension of K over F is s.

1.7 The ring $\mathbf{F}[x_1,\ldots,x_n]$

We already pointed out in Section 1.2 that for $n \geq 2$ $\mathbf{F}[x_1,\ldots,x_n]$ is not BD. However $\mathbf{F}[x_1,\ldots,x_n]$ has some nice properties. An important property is that any $\mathbf{F}[x_1,\ldots,x_n]$ is Noetherian ring (e.g. Lang [1967]).

Definition 1.7.1. $\mathbf{D}$ is said to be Noetherian (ND) if any ideal of $\mathbf{D}$ is finitely generated.

In what follows we shall assume that $\mathbf{F}$ is an algebraically closed. Let $p_1,\ldots,p_k \in F[x_1,\ldots,x_n]$. Denote by $U(p_1,\ldots,p_k)$ the common set of zeros of $p_1,\ldots,p_k$.

$$(1.7.2) \qquad U(p_1,\ldots,p_k) = \{z | z = (x_1,\ldots,x_n),\ p_j(z) = 0,\ j=1,\ldots,k\} \quad .$$

$U(p_1,\ldots,p_k)$ may be an empty set. $U(p_1,\ldots,p_k)$ is called an algebraic variety in $\mathbf{C}^n$. It is known (e.g. Lang [1958]) that any non-empty algebraic variety U in $\mathbf{F}^n$ splits

$$(1.7.3) \qquad U = \bigcup_{i=1}^{k} V_i \quad ,$$

where each V_i is an irreducible algebraic variety. Over the complex numbers each irreducible algebraic variety V is a closed connected set which almost everywhere (in V) is an analytic manifold in $\mathbf{C}^n$ of a fixed (complex) dimension d which is called the dimension of V. If $d = 0$ then V consists of one point. For any set $U \subseteq \mathbf{F}^n$ let $I(U)$ be the ideal of polynomials in $\mathbf{F}[x_1,\ldots,x_n]$ vanishing on U

$$(1.7.4) \qquad I(U) = \{p | p \in F[x_1,\ldots,x_n],\ p(z) = 0,\ z \in U\} \quad .$$

Consider an ideal $I \subseteq \mathbf{F}[x_1,\ldots,x_n]$. Suppose that the ideal generated by $p_1,\ldots,p_k$. Clearly $I \subseteq I(U(p_1,\ldots,p_k))$.

The celebrated Hilbert Nullstellensatz gives the other relation between the above two ideals (e.g. Lang [1967]).

Theorem 1.7.5. *Let F be an algebraically closed field. Consider an ideal $I \subseteq F[x_1,\ldots,x_n]$ generated by $p_1,\ldots,p_k$. Let $g \in F[x_1,\ldots,x_n]$. Then $g^j \in I$ for some positive integer j if and only if*

$$g \in I(U(p_1,\ldots,p_k)) \quad . \tag{1.7.6}$$

Corollary 1.7.7. *Let $p_1,\ldots,p_k \in F[x_1,\ldots,x_n]$, where F is an algebraically closed field. Then $p_1,\ldots,p_k$ generate $F[x_1,\ldots,x_n]$ if and only if*

$$U(p_1,\ldots,p_k) = \emptyset \quad . \tag{1.7.8}$$

1.8 Modules

Definition 1.8.1. Let S be a ring with identity. An abelian group M (which has an operation $+$ satisfying the conditions (1.1.2) - (1.1.6)) is called a (left) S-module if for each $r \in S$, $u \in M$ the product ru is defined to be an element in M such that the following properties hold.

(1.8.2) $$r(v_1+v_2) = rv_1+rv_2,\ (r_1+r_2)v = r_1v+r_2v,\ (rs)v = r(sv),\ 1v = v \ .$$

A standard example of a S-module is

(1.8.3) $$S^m = \{v,\ v = (v_1,\ldots,v_m)^t,\ v_i \in S,\ i=1,\ldots,m\}$$

where

(1.8.4) $$u + v = (u_1+v_1,\ldots,u_m+v_m)^t \ ,$$

(1.8.5) $$ru = (ru_1,\ldots,ru_m)^t,\ r \in S \ .$$

M is said to be finitely generated if there exist n-elements (generators) $v^1,\ldots,v^n$ such that any $v \in M$ is of the form

(1.8.6) $$v = \sum_{i=1}^{n} a_i v^i,\ a_i \in S,\ i=1,\ldots,n \ .$$

If each v can be expressed uniquely in the above form then $v^1,\ldots,v^n$ is called a basis in M and M is said to have a finite basis. N is called a submodule if $N \subseteq M$ and N is a S-module.

For example, consider a linear homogeneous system

$$(1.8.7) \qquad \sum_{j=1}^{n} a_{ij}x_j = 0, \ a_{ij}, x_j \in D, \ i=1,\ldots,m, \ j=1,\ldots,n \ .$$

Thus the set of all $x = (x_1,\ldots,x_n)^t$ is a submodule of D^n. In what follows we shall assume that S is an integral domain D. Let F be a field. Then a F-module is called a vector space (V) over F. It is a standard fact in linear algebra (e.g. Gantmacher [1959]) that if V is finitely generated then V has a finite basis. In that case V is called <u>finite dimensional</u> vector space. The number of vectors in any basis of V is called the <u>dimension</u> of V and by dim V. A submodule of V is called a <u>subspace</u> of V. Let M be a D-module with a finite basis. Let F be the quotient ring of D. It is possible to imbed M in a vector space V over F by considering all vectors v of the form (1.8.6) when $a_i \in F$, $i=1,\ldots,n$. Thus $\dim V = n$. Using this fact we get

<u>Lemma</u> 1.8.8. <u>Any two finite basis of a D-module contain the same number of elements.</u>

This number is called the dimension of M - and denoted by dim M.

<u>**Problems**</u>

(1.8.9) Let M be a - D module with a finite basis. Let N be a submodule of M. Prove that if D is PID then N has a finite basis.

(1.8.10) Let M be a - D module with a finite basis. Assume that N is a finitely generated submodule of M. Prove that if D is BD then N has a finite basis.

1.9 Matrices and homomorphismes

Notation 1.9.1. Denote by $M_{mn}(D)$ the set of all $m \times n$ matrices $A = (a_{ij})$, where $a_{ij} \in D$, $i=1,\dots,m$, $j=1,\dots,n$. The set $M_{nn}(D)$ is denoted by $M_n(D)$. Let M and N be D-modules. Let $T : M \to N$. T is called a homomorphism if

$$(1.9.2) \qquad T(au+bv) = aTu + bTv, \; u,v \in M, \; a,b \in D \;,$$

for all u,v and a,b. As usual, let

$$(1.9.3) \qquad \text{range } (T) = \{v | v = Tu, \; u \in M, \; v \in N\}, \; \ker (T) = \{u | Tu = 0, \; u \in M\} \;,$$

be the range and the kernel of T. Denote by $\mathrm{Hom}(M,N)$ the set of all homomorphismes of M to N. It is a standard fact that $\mathrm{Hom}(M,N)$ is a D-module by letting

$$(aS+bT)u = aSu + bTu$$

for any $S,T \in \mathrm{Hom}(M,N)$, $a,b \in D$, $u \in M$. Assume that M and N have finite bases. Let $u^1,\dots,u^m$ and $v^1,\dots,v^n$ be bases in M and N respectively. Then we can set a natural isomorphism between $\mathrm{Hom}(M,N)$ and $M_{mn}(D)$. Namely, for each $T \in \mathrm{Hom}(M,N)$ let $A = (a_{ij}) \in M_{mn}(D)$ be defined as follows

$$(1.9.4) \qquad Tu^i = \sum_{j=1}^{n} a_{ij} v^j, \; i=1,\dots,m \;.$$

Conversely, for any $A \in M_{mn}(D)$ there exists a unique $T \in \mathrm{Hom}(M,N)$ which satisfies (1.9.4). The matrix A is called a representation of T in bases $u^1,\dots,u^m$ and $v^1,\dots,v^n$. The rank of A - denoted by $r(A)$ - is defined as the size of the largest minor of $A(|A[\alpha|\beta]|, \; \alpha \in Q_{k,m} \; \beta \in Q_{k,n})$ which do not vanish. Thus if A is the representation matrix of $T \in Hom(M,N)$ then $r(A) = \dim TM$ if TM has a finite basis. Let $A \in M_{mn}(D)$. We shall view A as an element in $Hom(D^n,D^m)$ by letting $A(x) = Ax$ for $x \in D^n$.

We now study the relations between the representations of a fixed homomorphism $T \in Hom(M,N)$ with respect to different choices of bases in M and N.

Definition 1.9.5. A matrix $U \in M_n(D)$ is called unimodular if $|U|$ (the determinant of U) is an invertible element in D.

The above definition is equivalent to the existence of $V \in M_n(D)$ such that

$$UV = VU = I \quad , \tag{1.9.6}$$

when I is the identity matrix. Indeed $|U|$ is invertible then U^{-1} exists in the division ring F. Moreover the standard formula for U^{-1} in terms of the minors of U implies that $V = U^{-1} \in M_n(D)$. Vice versa if (1.9.6) holds then $|U||V| = 1$ so $|U|$ is invertible. Also U is unimodular if and only if the transpose of $U - U^t$ is unimodular.

Notation 1.9.7. Denote by $UM_n(D)$ the set of unimodular matrices in $M_n(D)$.

Clearly $UM_n(D)$ is a multiplicative group under the ordinary multiplication of the matrices. Unimodular matrices appear naturally when we change bases in D-module.

Lemma 1.9.8. Let M be a D-module with a finite basis $u^1,\ldots,u^m$. Then $\{\tilde{u}^1,\ldots,\tilde{u}^m\}$ is a basis in M if and only if the matrix $Q = (q_{ij}) \in M_m(D)$ given by the equalities

$$\tilde{u}^i = \sum_{j=1}^{m} q_{ij}u^j, \quad i=1,\ldots,m \tag{1.9.9}$$

is a unimodular matrix.

Proof. Suppose first that $\{\tilde{u}^1,\ldots,\tilde{u}^m\}$ is a basis in M. Then

$$u^j = \sum_{k=1}^{m} r_{jk}\tilde{u}^k, \quad j=1,\ldots,m \quad . \tag{1.9.10}$$

Put $R = (r_{jk})$. Insert (1.9.10) to (1.9.8) to get $QR = I$ - as $\{\tilde{u}^1,\ldots,\tilde{u}^m\}$ is a basis. This shows that Q is unimodular. Assume now that $Q \in UM_n(D)$. Let $R = A^{-1}$. So (1.9.10) holds. Also $\tilde{u}^1,\ldots,\tilde{u}^m$ linearly independent over D, i.e. 0-cannot be written as a non-trivial combination of $\tilde{u}^1,\ldots,\tilde{u}^m$, since otherwise we deduce the linear dependence of $u^1,\ldots,u^m$. But this is impossible as $u^1,\ldots,u^m$ is a basis in M. So $\{\tilde{u}^1,\ldots,\tilde{u}^m\}$ is also a basis in M.

Definition 1.9.11. *Let* $A,B \in M_{mn}(D)$. *We say that* A *is right equivalent to* $B(A \underset{r}{\sim} B)$ *if*

$$B = AP \tag{1.9.12}$$

for some $P \in UM_n(D)$; A *is left equivalent to* $B(A \underset{\ell}{\sim} B)$ *if*

$$B = QA \tag{1.9.13}$$

for some $Q \in UM_m(D)$; A *is equivalent to* $B(A \sim B)$ *if*

$$B = QAP \quad , \tag{1.9.14}$$

for some $Q \in UM_m(D)$, $P \in UM_n(D)$.

Obviusly, all the above relations are equivalence relations.

Theorem 1.9.15. *Let* M *and* N *be* D-*modules with finite bases having* m *and* n *elements respectively. Then* $A,B \in M_{mn}$ *are* (i) *left equivalent;* (ii) *right equivalent;* (iii) *equivalent; if correspondingly there exists* $T \in Hom(M,N)$ *such that* A *and* B *are the representation matrices of* T *in the following bases*

(i)' $\{u^1,\ldots,u^m\}, \{v^1,\ldots,v^n\}$ *and* $\{\tilde{u}^1,\ldots,\tilde{u}^m\}, \{v^1,\ldots,v^n\}$;

(ii)' $\{u^1,\ldots,u^m\}$, $\{v^1,\ldots,v^n\}$ and $\{u^1,\ldots,u^m\}$, $\{\tilde v^1,\ldots,\tilde v^n\}$;

(iii)' $\{u^1,\ldots,u^m\}$, $\{v^1,\ldots,v^n\}$ and $\{\tilde u^1,\ldots,\tilde u^m\}$, $\{\tilde v^1,\ldots,\tilde v^n\}$.

Proof. Let A be the representation matrix of T in the bases $u^1,\ldots,u^m$ and $v^1,\ldots,v^n$. Assume that the relation between the bases $u^1,\ldots,u^m$ and $\{\tilde u^1,\ldots,\tilde u^m\}$ is given by (1.9.9). Then the reresentation matrix B in bases $\{\tilde u^1,\ldots,\tilde u^m\}$ and $\{v^1,\ldots,\tilde v^n\}$ is given by (1.9.13). Indeed

$$T\tilde u^i = \sum_{j=1}^{m} q_{ij} T u^j = \sum_{j,k=1}^{m,n} q_{ij} a_{jk} v^k,\ i=1,\ldots,m \ ,$$

which proves (1.9.13). On the other hand if we change the basis $\{v^1,\ldots,v^n\}$ to $\{\tilde v^1,\ldots,\tilde v^n\}$ according to $v^j = \sum_{\ell=1}^{n} p_{j\ell}\tilde v^\ell$, $j=1,\ldots,n$, $P = (p_{j\ell}) \in UM_n(D)$ then a similar computation shows that T is represented in $\{u^1,\ldots,u^m\}$ and $\{\tilde v^1,\ldots,\tilde v^n\}$ by AP. Combine the above two results to deduce that the representation matrix B of T is bases $\{\tilde u^1,\ldots,\tilde u^m\}$ and $(\tilde v^1,\ldots,\tilde v^m)$ is given by (1.9.14). The proof of the theorem is concluded.

□

1.10 Hermite normal form

The notions of equivalence of $A, B \in M_{mn}(D)$ give rise to the following problems.

Problem 1.10.1. *Given* $A, B \in M_{mn}(D)$. *When* A *and* B *are* (i) *left equivalent;* (ii) *right equivalent;* (iii) *equivalent.*

Problem 1.10.2. *For a given* $A \in M_{mn}(D)$ *characterize the equivalence classes corresponding to* A *for* (i) *left equivalence;* (ii) *right equivalence;* (iii) *equivalence.*

Clearly a satisfying solution of Problem 1.10.2 would automatically solve Problem 1.10.1. However, if the solution of Problem 1.10.2 is unknown or is complicated there is a point to solve Problem 1.10.1 separately.

We first note that for **GCDD** the equivalence relations have certain obvious invariants.

Lemma 1.10.3. *For* $A \in M_{mn}(\mathrm{GCDD})$ *let*

$$\begin{aligned}
\mu(\alpha,A) &= \text{g.c.d.}\{|A[\alpha|\theta]|,\ \theta \in Q_{k,n}\},\ \alpha \in Q_{k,m}\ ,\\
\nu(\beta,A) &= \text{g.c.d.}\{|A[\varphi|\beta]|,\ \varphi \in Q_{k,m}\},\ \beta \in Q_{k,n}\ ,\\
\delta_k(A) &= \text{g.c.d.}\{|A[\varphi|\theta]|,\ \varphi \in Q_{k,m},\ \theta \in Q_{k,n}\}\ .
\end{aligned} \tag{1.10.4}$$

($\delta_k(A)$ *is called the k-th determinant invariant of* A.) *Then*

$$\begin{aligned}
\mu(\alpha,A) &= \mu(\alpha,B),\quad \forall\alpha \in Q_{k,m} \quad \text{if} \quad A \underset{r}{\sim} B\ ,\\
\nu(\beta,A) &= \nu(\beta,B),\quad \forall\beta \in Q_{k,n} \quad \text{if} \quad A \underset{\ell}{\sim} B\ ,\\
\delta_k(A) &= \delta_k(B) \quad \text{if} \quad A \sim B\ ,
\end{aligned} \tag{1.10.5}$$

$k=1,\ldots,\min(m,n)$.

Proof. Suppose that (1.9.12) holds. The Cauchy-Binet formula (e.g. Gantmacher [1959]) implies $|B[\alpha,\gamma]| = \sum_{\theta\in Q_{k,n}} |A(\alpha,\theta)|\ |P[\theta,\gamma]|$. So $\mu(\alpha,A)$ divides $\mu(\alpha,B)$. As $A = BP^{-1}$ $\mu(\alpha,B)|\mu(\alpha,A)$. Thus $\mu(\alpha,A) = \mu(\alpha,B)$. (Recall that $\mu(\alpha,A)$ and $\mu(\alpha,B)$ are determined up to invertible elements). The other equalities in (1.10.5) are established in a similar way.

Note that

$$A \underset{\ell}{\sim} B \Longleftrightarrow A^t \underset{r}{\sim} B^t,\quad A,B \in M_{mn}(D) \quad . \tag{1.10.6}$$

Thus it is enough to consider the left equivalence relation. In what follows we characterize the equivalence class for left (right) equivalence relation in case that D is a Bezout domain. To do so we need a few notations.

Recall that $P \in M_n(D)$ is called a permutation matrix if P is a matrix having at each row and column one non-zero element which is the identity element 1. Clearly $P \in UM_n(D)$ since $P^{-1} = P^t$.

Definition 1.10.7. *A unimodular matrix* $U \in M_n(D)$ *is called simple if there exist permutation matrices* P,Q *such that*

$$U = P(V \oplus I_{n-2})Q \quad , \tag{1.10.8}$$

where V *is a unimodular* 2×2 *matrix*

$$V = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in UM_2(D) \quad , \tag{1.10.9}$$

i.e. $\alpha\delta - \beta\gamma$ is invertible. U is said to be elementary if V is of the form

$$V = \begin{pmatrix} \alpha & \beta \\ 0 & \delta \end{pmatrix}, \quad \alpha,\delta - \text{invertible} \quad . \tag{1.10.10}$$

Definition 1.10.11. Let $A \in M_{mn}(D)$. The following row (column) operations are called elementary

(i) interchange any two rows (columns) of A;

(ii) multiply row (column) i by an invertible element a;

(iii) add to row (column) j b times row (column) $i (i \neq j)$.

The following row (column) operation is called simple.

(iv) replace row (column) i by a times row (column) i plus b times row (column) j; and row (column) j by c times row (column) i plus d times row (column) j; where $ad-bc$ is an invertible element in D.

It is not difficult to see that elementary row (column) operations can be carried out by multiplication of A by a suitable elementary matrix U from left (right), and the simple row (column) operations are carried out by multiplication of A by a simple matrix U from left (right).

Theorem 1.10.12. Let D be a Bezout domain. Consider $A \in M_{mn}(D)$. Assume that $r(A) = r$. Then there exists $B = (b_{ij}) \in M_{mn}(D)$ which is left equivalent to A and satisfies

$$b_{ij} = 0 \quad \text{for} \quad i > r \quad ; \tag{1.10.13}$$

if b_{in_i} is the first non-zero entry in i^{th} row then

$$1 \leq n_1 < n_2 < \cdots < n_r \leq n \quad . \tag{1.10.14}$$

The numbers $n_1,\ldots,n_r$ *are uniquely determined and the elements* b_{in_i}, $i=1,\ldots,r$ *are uniquely determined, apart from arbitrary invertible factors, by the conditions*

$$(1.10.15)\qquad \begin{aligned} &\nu((n_1,\ldots,n_i),A) = b_{1n_1}\cdots b_{in_i},\ i=1,\ldots,r\ ,\\ &\nu(\alpha,A) = 0,\quad \alpha \in Q_{i,(n_i-1)},\ i=1,\ldots,r\ . \end{aligned}$$

The elements b_{jn_i} $(j < i)$ *are then successively uniquely determined apart from the addition of arbitrary multiples of* b_{in_i}. *The remaining elements* b_{ik} *are now uniquely determined. Moreover, the unimodular matrix* Q *which satisfied* (1.9.13) *can be given as a product of finite number of simple matrices.*

Proof. Our proof is by induction on n and m. For $n = m = 1$ the theorem is obvious. Let $n = 1$ and assume that for a given $m \geq 1$ there exists a matrix Q which is a finite product of simple matrices such that the entries $(i,1)$ of QA are equal to zero for $i = 2,\ldots,m$. Let $A_1 \in M_{(m+1)1}(D)$ and assume that $A = (a_{i1})$, $i=1,\ldots,m$. Put $Q_1 = Q \oplus I_1$. Then $A_2 = Q_1A_1$ and the $(i,1)$ entries of A_2 are zero for $i=2,\ldots,m$. Interchange the second and the last row of A_2 to obtain A_3. Clearly, $A_3 = (a_{i1}^{(3)}) = Q_2A_2$ where Q_2 is an appropriate permutation matrix. Let $A_4 = (a_{11}^{(3)},a_{21}^{(3)})^t$. Since D was assumed to be Bezout domain, there exists $\alpha,\beta \in D$ such that

$$(1.10.16)\qquad \alpha a_{11}^{(3)} + \beta a_{21}^{(3)} = (a_{11}^{(3)},a_{21}^{(3)}) = d\ .$$

As $(\alpha,\beta) = 1$ there exists $\gamma,\delta \in D$ such that

$$(1.10.17)\qquad \alpha\delta - \beta\gamma = 1\ .$$

Let V be a 2×2 unimodular matrix given by (1.10.9). But $A_5 = VA_4 = \binom{d}{d'}$.

According to Lemma 1.10.3 $\nu((1),A_4) = \nu((1),A_5)$. As $\nu((1),A_4) = d$ we must have $d|d'$. That is $d' = pd$. Thus

$$\binom{d}{0} = WA_5, \quad W = \begin{pmatrix} 1 & 0 \\ -p & 1 \end{pmatrix} \in UM_2(D) \quad .$$

Let

$$Q_3 = (W \oplus I_{m-1})(V \oplus I_{m-1}) \quad .$$

Then the last m rows of $A_6 = (a_{i1}^{(6)}) = Q_3A_3$ are zero. So $a_1^{(6)} = \nu((1),A_6) = \nu((1),A)$ and the theorem is proved in this case.

Assume now that we proved the theorem for all $A_1 \in M_{mn}(D)$ where $n \leq p$. Let $n = p+1$ and $A \in M_{m(p+1)}(D)$. Call $A_1 = (a_{ij})$, $i=1,\ldots,m$, $j=1,\ldots,p$. The induction assumption implies the existence of $Q_1 \in UM_m(D)$ which is a product of simple matrices such that $B_1 = (b_{ij}^{(1)}) = Q_1A_1$ satisfies the assumptions of the theorem. Let $n_1',\ldots,n_s'$ be the integers defined by A_1. If $b_{in}^{(1)} = 0$ for $i > s$ then $n_i = n_i'$, $i=1,\ldots,s$ and $B = Q_1A$ is in the right form. Suppose now that $b_{in}^{(1)} \neq 0$ at least for some i, $s < i \leq m$. Let $B_2 = (b_{in}^{(1)})$, $i=s+1,\ldots,m$. According to what we proved, there exists $Q_2 \in UM_{m-s}(D)$ such that $Q_2B_2 = (c,0,\ldots,0)^t$, where c is the g.c.d. of $b_{(s+1)m}^{(1)},\ldots, b_{mn}^{(1)}$. So $B_3 = (I_s \oplus Q_2)B_1$ is in the right form, where $s = r-1$, $n_i = n_i'$, $i=1,\ldots,r-1$, $n_r = n$. Next we prove (1.10.15). First if $\alpha \in Q_{i,(n_i-1)}$ then any matrix $B[\beta|\alpha]$, $\beta \in Q_{i,m}$ has at least one zero row, so $|B[\beta|\alpha]| = 0$. Thus $\nu(\alpha,B) = 0$. Lemma 1.10.3 yields that $\nu(\alpha,A) = 0$. Let $\alpha = (n_1,\ldots,n_i)$. Then $B[\beta|\alpha]$ has at least one zero row unless β equals to $\gamma = (1,2,\ldots,i)$. Therefore $\nu(\alpha,A) = \nu(\alpha,B) = |B[\gamma|\alpha]| = b_{1n_1}\cdots b_{in_i} \neq 0$. This establishes (1.10.15). So $n_1,\ldots,n_r$ are determined by (1.10.15). It is obvious that $b_{1n_1},\ldots,b_{rn_r}$ are determined up invertible elements. Suppose that $n = n_r$. Then we can add to $j^{\underline{th}}$ row $(j < r)$ any multiple of $r^{\underline{th}}$ row without changing the first $n-1$ columns of B while the last $n-r$ rows of B remain zero rows. This argument shows that $b_{jn_i}(j < i)$ can be only determined up to the addition of multiples of b_{in_i}. It is left to prove that once b_{jn_i} are chosen within the above freedom for $j=1,\ldots,i$, $i=1,\ldots,r$, then all the remaining elements are uniquely determined. Let $C \in M_{mn}(D)$ such that the $n_1,\ldots,n_r$ columns of C and B are the same and

$$c_{kj} = 0 \quad \text{for} \quad n_i < j < n_{i+1},\ i < k,\ i=0,\ldots,r\ (n_0 = -1 \ \text{if} \ n_1 > 1, \ \text{and}$$

$$n_{r+1} = n \ \text{if} \ n_r < n) \ .$$

Assume that

$$C = UB,\ U \in M_m(D) \ .$$

We claim that

$$u_{ij} = \delta_{ij},\ i=1,\ldots,m,\ j=1,\ldots,r \ . \tag{1.10.18}$$

We prove this result by induction on n. For $n = 1$ it is obvious. Suppose that the assertion holds for $n = p$. Let $n = p+1$. Put $\tilde{B} = (b_{ij})$, $\tilde{C} = (c_{ij})$, $i=1,\ldots,m$, $j=1,\ldots p$. So $\tilde{C} = U\tilde{B}$. If $n_r < p+1$ then the induction hypothesis implies (1.10.18). Assume that $n = n_r$. Then (1.10.18) holds for $j=1,\ldots,p$. The equality $C = UB$ implies $c_{rn_r} = u_{rr}b_{rn_r}$. As $b_{rn_r} = c_{rn_r} \neq 0$ we have $u_{rr} = 1$. Now

$$c_{in_r} = b_{in_r} + u_{ir}b_{rn_r},\ i=1,\ldots,r-1,\ c_{in_r} = u_{ir}b_{rn_r},\ i=r+1,\ldots,m \ .$$

By our assumption $c_{in_r} = b_{in_r}$ for $i \leq r-1$ so the above equalities yield $u_{ir} = 0$ for $i = r-1$. Also the assumption that $c_{in_r} = 0$ for $i > r$ implies $u_{ir} = 0$. Since the last $m - r$ rows of C and B are equal to zero we finally deduce that $C = B$. This establishes the uniqueness of B provided that the elements b_{jn_i}, $j=i,\dots,1$ were chosen as above.

□

A matrix $B \in M_{mn}(\mathbf{D})$ is said to be in the Hermite normal form if it satisfies the conditions (1.10.13) - (1.10.14). In what follows we shall always assume the specified normalizations.

Normalization 1.10.19. *If* b_{in_i} *is an invertible element we choose* $b_{in_i} = 1$ *and* $b_{jn_i} = 0$ *for* $j < i$.

Theorem 1.10.20. *Let* U *be a unimodular matrix over a Bezout domain. Then* U *is a finite product of simple matrices.*

Proof. Since $|U|$ is invertible according to Theorem 1.10.12 each b_{ij} is an invertible element. Thus the Normalization 1.10.19 implies that Hermite normal form of U is the identity matrix. Thus the inverse of U is a finite product of simple matrices. Therefore U itself is a finite product of simple matrices.

Normalization 1.10.21. *For Euclidean domains assume*

$$\text{(1.10.22)} \qquad \text{either } b_{jn_i} = 0 \text{ or } d(b_{jn_i}) < d(b_{in_i}) \text{ for } j < i \; .$$

For $F[x]$ *we choose* b_{in_i} *tc be a normalized polynomial.*

Combining Normalization 1.10.21 with Theorem 1.10.12 we get

Corollary 1.10.23. *Over the ring* $F[x]$ *the Hermite normal form of* $A \in M_{mn}(F[x])$ *is unique provided that the Normalization* 1.10.21 *holds.*

It is a well known fact that over Euclidean domains Hermite normal form can be achieved by performing elementary rows operations. This result follows by considering 2×2 matrices.

Theorem 1.10.24. *Let* $A \in M_2(\mathbf{ED})$. *Then* A *can be brought to its Hermite normal form by a finite number of elementary rows operations.*

Proof. Suppose that

$$A_i = \begin{pmatrix} a_i & b_i \\ a_{i+1} & b_{i+1} \end{pmatrix}, \; A_1 = A \;.$$

Let us compute a_{i+2} by (1.3.14). So A_i is equivalent to

$$\tilde{A}_i = \begin{pmatrix} a_{i+2} & b_{i+2} \\ a_{i+1} & b_{i+1} \end{pmatrix}, \; a_{i+2} = a_i - t_i a_{i+1}, \; b_{i+2} = b_i - t_i b_{i+1},$$

$$a_{i+2} = 0 \quad \text{or} \quad d(a_{i+2}) < d(a_{i+1}) \;.$$

Thus A_i is left equivalent to A_{i+1}. As the Euclid algorithm terminates after a finite number of steps we obtain that $a_{k+1} = 0$. If $a_k \neq 0$, A_k is in the Hermite normal form. Otherwise, $a_1 = a_2 = 0$ and we perform the Euclid algorithm on b_1, b_2 to obtain the Hermite form.

Corollary 1.10.25. Let $U \in UM_2(\mathbb{D})$. Then U is a finite product of elementary unimodular matrices.

Corollary 1.10.26. Let $U \in UM_n(\mathbb{D})$. Then U is a finite prouct of elementary unimodular matrices.

Problems

(1.10.27) Let $T \in \mathrm{Hom}(M,N)$ where M and N are $\mathbb{D}$ modules. Assume that $M = \mathbb{D}^m$. Let $\mathrm{Im}(T)$ be the image of M in N. Then the module $\mathrm{Im}(T)$ has a basis $Tu^1,\ldots,Tu^k$ such that

(1.10.28) $$u^i = \sum_{j=1}^{i} c_{ij} v^j, \; c_{ii} \neq 0, \; i=1,\ldots,k \;,$$

where $v^1,\ldots,v^m$ is a permutation of the standard basis $(\delta_{i1},\ldots,\delta_{im})^T$, $i=1,\ldots,m$ in $\mathbb{D}^m$.

(1.10.29) Let $A \in M_{mn}(BD)$ and assume the B is the Hermite normal form. Let $n_i < j < n_{i+1}$. Prove that for $\alpha = (n_1,\dots,n_{i-1},j)$, $\nu(\alpha,A) = b_{1n_1}\dots b_{(i-1)n_{i-1}} b_{ij}$ $(n_0 = 0)$.

1.11 Systems of linear equations over Bezout domains

Consider a system of m linear equations in n unknowns.

$$\sum_{j=1}^{n} a_{ij}x_j, \; i=1,\dots,m, \; a_{ij},b_i \in D, \; i=1,\dots,m, \; j=1,\dots,n \quad . \tag{1.11.1}$$

In a matrix notation (1.11.1) is equivalent to

$$Ax = b, \; A \in M_{mn}(D), \; x \in M_{n1}(D), \; b \in M_{m1}(D) \quad . \tag{1.11.2}$$

Let

$$\hat{A} = (A,b) \in M_{m(n+1)}(D) \quad . \tag{1.11.3}$$

The matrix A is called the coefficient matrix and the matrix $\hat{A}$ is called the augmented coefficient matrix. In case that D is a field the classical Kronecker-Capelli theorem states (e.g. Gantmacher [1959]) that (1.11.1) is solvable if and only if

$$r(A) = r(\hat{A}) \quad . \tag{1.11.4}$$

Let F be the quotient field of D. Thus if (1.11.1) is solvable over D it is also solvable over F. Therefore (1.11.4) is a necessary condition for the solvability of (1.11.1) over D. Clearly, even in the case where $m = n = 1$ this condition is not sufficient.

In this section we give necessary and sufficient condition on $\hat{A}$ for the solvability of (1.11.1) over D in case that D is a Bezout domain. To do so we need the following lemma.

Lemma 1.11.5. *Let* $A \in M_{mn}(BD)$, $A \neq 0$. *Then there exists an* $m \times m$ *permutation matrix* P *and an* $n \times n$ *unimodular matrix* U *such that*

$$(1.11.6) \qquad C = (c_{ij}) = PAU, \; c_{ij} = 0 \text{ for } j > i, \; c_{ii} \neq 0, \; i=1,\dots,r \;, \quad c_{ij} = 0 \text{ for } j > r, \; r = r(A) \;.$$

Proof. Consider the matrix A^t. By interchanging the columns of A^t, i.e. multiplying A^t from right by some permutation matrix P^t, we can assume that in the Hermite normal form of A^tP^t, $n_i = i$, $i=1,\dots,r$. This establishes (1.11.6)

Theorem 1.11.7. *Consider the system* (1.11.1). *If* D *is a Bezout domain. Then* (1.11.1) *is solvable if and only if*

$$(1.11.8) \qquad r = r(A) = r(\hat{A}), \; \delta_r(A) = \delta_r(\hat{A}) \;.$$

Proof. Assume first the existence of $x \in M_{n1}(D)$ which satisfies 1.11.2. As we pointed out already this assumption implies the equality (1.11.4). Also from (1.11.2) we deduce that b is a linear combination of the columns of A. Consider any $r \times r$ minor of A which contains the column b. Since b is a linear combination of the other columns of A we deduce that this minor is a linear combination of $r \times r$ minors of A. So $\delta_r(A)$ divides the minor in question. In view of the definition (1.10.4) of $\delta_r(A)$ clearly $\delta_r(\hat{A}) | \delta_r(A)$. Thus we proved that the condition (1.11.8) is necessary for the solution of (1.11.1) over Bezout domain (in fact over GCDD). Suppose now that the condition (1.11.8) holds. By changing the order of the equations in (1.11.1) and considering a new set of variables

(1.11.9) $$y = U^{-1}x, \quad U \in UM_n(D) \quad ,$$

we may assume that $A = C$ where C is given in Lemma 1.11.5. In view of (1.11.6) and the condition (1.11.4) the last $m - r$ are linear combinations of the first r equations (possibly the coefficients of the linear combinations are in the quotient field of D). Therefore it is enough to show the solvability of the system

(1.11.10) $$\sum_{j=1}^{i} c_{ij}x_j = b_i, \ i=1,\dots,m, \ c_{ii} \neq 0, \ i=1,\dots,m \quad .$$

Let $m = 1$. Clearly, in this case $\delta_1(\hat{C}) = c_{11}$ and $\delta_1(C) = (c_{11},b_1)$. The second equality in (1.11.8) means $c_{11}|b_1$ so (1.11.10) is solvable over D. Assume that (1.11.8) holds for the system (1.11.10) $(r = m)$. Consider an $m \times m$ minor of $\hat{C}$ composed from the m rows and the columns $2,\dots,m+1$. This minor is equal to $(-1)^{m-1}b_1c_{22}\cdots c_{mm}$. Since this minor is divided by $\delta_1(C) = c_{11}\cdots c_m$ we have that c_{11} divides b_1. So $x_1 = b_1/c_{11} \in D$. Thus it is left to show that the system

(1.11.11) $$\sum_{j=2}^{i} c_{ij}x_j = b_i - c_{i1}b_1/c_{11}, \ i=2,\dots,m$$

is solvable over D. Put $\bar{C} = (c_{ij})$, $i=2,\dots,m$, $j=2,\dots,m$, $\bar{b} = (b_i - c_{i1}b_{11}/c_{11})$, $i=2,\dots,m$. The induction hypothesis would imply the solvability of (1.11.11) if $\delta_{m-1}(\bar{C}) = \delta_{m-1}(\hat{\bar{C}})$. That is, it is enough to show that $c_{22}\cdots c_{mm}$ divides any $(m-1) \times m-1)$ minor $|\bar{J}|$ of $\hat{\bar{C}}$ which is composed of $2,\dots,m$ rows and $2,\dots,i-1,$ $i+1,\dots,m+1$ columns of $\hat{C}$. Consider the minor $|J|$ of $\hat{C}$ which is composed of the rows $1,\dots,m$ and the columns $1,2,\dots,i-1,\ i+1,\dots,m+1$. Substract from the last column the first column times b_1/c_{11}. So $|\mathcal{J}| = c_{11}|\bar{\mathcal{J}}|$. Since $\delta_m(C) = \delta_m(\hat{C})$ we have that $c_{11}\cdots c_m | c_{11}|\bar{\mathcal{J}}|$. Therefore $\delta_{m-1}(\bar{C}) \Big| |\bar{\mathcal{J}}|$ which finally implies that (1.11.11) is solvable over D. This completes the proof of the theorem.

$\square$

Theorem 1.11.12. *Let* $A \in M_{mn}(BD)$. *Then* range (A) *and* ker (A) *are modules in* D^m *and* D^n *having finite bases with* $r(A)$ *and* $n-r(A)$ *elements respectively. Moreover, the base of* ker (A) *can be completed to the base in* D^n.

Proof. As in the proof of Theorem 1.11.7 we may assume that $A = C$ where C is given by (1.11.6). Let $\varepsilon^i = (\delta_{i1},\ldots,\delta_{in})^t$. Clearly $A\varepsilon^1,\ldots,A\varepsilon^r$ is a basis in range (A) and $\varepsilon^{r+1},\ldots,\varepsilon^n$ is a basis for ker (A).

$\square$

Let $A \in M_{mn}(GCDD)$. If we expand any $q \times q$ minor of A by any $q - p$ rows, $1 \leq p < q$, we immediately deduce that

$$\delta_p(A) | \delta_q(A) \quad \text{for} \quad 1 \leq p \leq q \leq \min(m,n) \quad . \tag{1.11.13}$$

Definition 1.11.14. *For* $A \in (GCDD)$ *denote* $i_1(A),\ldots,i_r(A)$ $(r=r(A))$ *the invariant factors of* A.

$$i_j(A) = \delta_j(A)/\delta_{j-1}(A), \ j=1,\ldots,r(A), (\delta_0(A) = 1) \quad . \tag{1.11.15}$$

Then invariant factor $i_j(A)$ *is called trivial if* $i_j(A)$ *is invertible.*

From (1.11.14) and (1.11.12) we deduce

$$i_{j-1}(A) | i_j(A), \ j=2,\ldots,r(A) \quad . \tag{1.11.16}$$

Suppose that (1.11.1) is solvable over D. Using the fact that b is a linear combination of the columns of A and Theorem 1.11.7 we get a weaker version of Theorem 1.11.7.

Corollary 1.11.17. *Let* $A \in M_{mn}(BD)$. *Then the system* (1.11.1) *is solvable over* BD *if and only if*

(1.11.18) $$r = r(A) = r(A),\ i_k(A) = i_k(\hat{A}),\ k=1,\dots,r\ .$$

Problems

(1.11.19) Let $A \in M_n(GCDD)$. Prove that

(1.11.20) $|A| = i_1(A)\dots i_n(A)$.

1.12 Smith normal form

According to Lemma 1.10.3 and Definition 1.11.14 for $A \in M_{mn}(GCDD)$ the rank of $A - r$ and the invariant factors $i_1(A),\ldots,i_r(A)$ are the invariants with respect to the equivalence relation. It turns out the if D is elementary divisor domain then the above invariants characterize the equivalence class of A with respect to the relation $A \sim B$.

<u>Theorem</u> 1.12.1. <u>Let</u> $0 \neq A \in M_{mn}(D)$. <u>Assume that</u> D <u>is elementary divisor domain. The</u> A <u>is equivalent to a diagonal matrix of the form</u>

$$B = \mathrm{diag}(i_1(A),\ldots,i_r(A),\ 0,\ldots,0) \tag{1.12.2}$$

<u>where</u> r <u>is the rank of</u> A.

<u>Proof.</u> Recall that if D is EDD then D is BD. For $n = 1$ the Hermite normal form of A is a diagonal matrix where $i_1(A) = \delta_1(A)$. So the theorem is established in this case. Next we show that the theorem holds for $m = n = 2$. We may assume that A is in the Hermite normal form

$$A = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} .$$

According to the Definition 1.5.1 there exist $p,q,x,y \in D$ which satisfy (1.5.2). Clearly $(p,q) = (x,y) = 1$. So there exist $\bar{p},\bar{q},\ \bar{x},\bar{y}$ such that

$$p\bar{p} - q\bar{q} = x\bar{x} - y\bar{y} = 1.$$

Let

$$V = \begin{pmatrix} p & q \\ \bar{q} & \bar{p} \end{pmatrix} , \quad U = \begin{pmatrix} x & \bar{y} \\ y & \bar{x} \end{pmatrix} .$$

Then (1.5.2) and the above equalities imply

$$G = VAU = \begin{pmatrix} \delta_1(A) & g_{12} \\ g_{21} & g_{22} \end{pmatrix} .$$

Since $\delta_1(G) = \delta_1(A)$ we deduce that $\delta_1(A)$ divides g_{12} and g_{21}. By applying appropriate elementary row and column operations we deduce that A is equivalent to a diagonal matrix

$$C = \mathrm{diag}(i_1(A), d_2) .$$

Since $\delta_2(C) = i_1(A)d_2 = \delta_2(A)$ we obtain that C is in the form (1.12.2). We now prove the case $m > 3$, $n = 2$ by the induction on m. Let $\bar{A} = (a_{ij})$, $i=1,\ldots,m-1$, $j=1,2$. We can assume that $\bar{A}$ is in the form (1.12.2). In particular a_{11} is the g.c.d. of all elements of $\bar{A}$. Interchange the second row with the last one to obtain A_1. Apply simple row and column operations on the first two rows and columns of A_1 to obtain $A_2 = (a_{ij}^{(2)})$, where $a_{11}^{(2)} = i_1(A)$. Now use the elementary row and column operations to obtain A_3 of the form

$$A_3 = i_1(A) \oplus A_4 . \tag{1.12.3}$$

Since A_4 has one column we bring it to a diagonal form. So A is equivalent to a matrix

$$C = \mathrm{diag}(i_1(A), i_1(A_4), 0,\ldots,0) .$$

We claim that C is in the form (1.12.2). Indeed

$$\delta_2(C) = i_1(A)i_1(A_4) = \delta_2(A) \quad .$$

So

$$i_1(A_4) = \frac{\delta_2(A)}{i_1(A)} = \frac{\delta_2(A)}{\delta_1(A)} = i_2(A) \quad .$$

For $n \geq 3$ we prove the theorem by the induction. Thus we may assume that $\bar{A} = (a_{ij})$, $i=1,\ldots,m$, $j=1,\ldots,n-1$ is already in the form (1.12.2). So a_{11} is the g.c.d..of all the elements of $\bar{A}$. Interchange the second column with the last column in A to obtain $A_1 = (a_{ij}^{(1)})$. Then A is equivalent to a matrix $A_2 = (a_{ij}^{(2)})$ such that the first two columns of A_2 form the canonical form of $\bar{A}_1$. This in particular implies that $a_{11}^{(2)} = i_1(A)$. Perform elementary row and column operations to bring A_2 to a matrix A_3 of the form (1.12.3). As $i_1(A_3) = i_1(A)$ we obtain that $i_1(A)$ divides all entries of A_4. The induction hypothesis implies that A is equivalent to a diagonal matrix

$$C = \operatorname{diag}(i_1(A), i_1(A_4), \ldots, i_{r-1}(A_4), 0, \ldots, 0) \quad .$$

It is left to show that C is the matrix (1.12.2). Indeed, as $i_1(A)|i_j(A_4)$ and $i_j(A_4)|i_{j+1}(A_4)$ we immediately deduce that

$$\delta_k(C) = i_1(A)i_1(A_4)\cdots i_{k-1}(A_4) \quad .$$

So

$$i_1(C) = i_1(A), \; i_k(C) = i_{k-1}(A_4), \; k=2,\ldots,r \quad .$$

This shows that A is equivalent to the matrix B given by (1.12.2). The proof of the theorem is completed.

The matrix (1.12.2) is called Smith normal form of A.

Corollary 1.12.4. Let $A, B \in M_{mn}(\mathbf{EDD})$. Then A and B are equivalent if and only if A and B have the same rank and the same invariant factors.

Over elementary divisor domain the system (1.11.2) is equivalent to very simple system

$$(1.12.5) \qquad i_k(A)y_k = c_k,\ k=1,\dots,r(A),\ 0 = c_k,\ k=1,\dots,m \quad ,$$

$$(1.12.6) \qquad y = P^{-1}x,\ c = Qb$$

where P and Q are the unimodular matrices appearing in (1.9.14) and B is of the form (1.12.2). For the system (1.12.5) Theorems 1.11.7 and 1.11.12 are quite obvious. We also have

Theorem 1.12.7. Let $A \in M_{mn}(\mathbf{EDD})$. Assume that all the invariant factors of A are trivial. The the basis of range (A) can be completed to a basis of D^n.

In what follows we adopt

Normalization (1.12.8). Let $A \in M_{mn}(\mathbf{F}[x])$. Then the invariant polynomials (the invariant factors) of $A(x)$ are assumed to be normalized polynomials.

Problems

(1.12.9) Let $A = \mathrm{diag}(p,q) \in M_2(\mathbf{BD})$. Then A is equivalent to $\mathrm{diag}((p,q), \frac{pq}{(p,q)})$.

(1.12.10) Let $A \in M_{mn}(\mathbf{D})$, $B \in M_{pq}(\mathbf{D})$ and assume that $\mathbf{D}$ is $\mathbf{GCDD}$. Suppose that either $i_s(A)|i_t(B)$ or $i_t(B)|i_s(A)$ for $s=1,\dots,r(A) = \alpha$, $t=1,\dots,r(B) = \beta$. Show the set of the invariant factors of $A \oplus B$ is $\{i_1(A),\dots,i_\alpha(A),\ i_1(B),\dots,i_\beta(B)\}$.

(1.12.11) Let $\mathbf{N} \subseteq \mathbf{M}$ be $\mathbf{D}$ modules with finite bases. Assume that $\mathbf{D}$ is $\mathbf{EDD}$. Prove that there exists a basis $u_1,\dots,u_n$ in $\mathbf{M}$ such that $i_1u_1,\dots,i_ru_r$ is a basis in $\mathbf{N}$ when $i_1,\dots,i_r \in D$ and $i_j|i_{j+1}$, $j=1,\dots,r-1$.

(1.12.12) Let M be a D module and $N_1, N_2 \subseteq M$ be submodules. N_1 and N_2 are called equivalent if there exists $T \in \mathrm{Hom}(M,M)$, T isomorphism (i.e. $T^{-1} \in \mathrm{Hom}(M,M)$) such that $TN_1 = N_2$. Suppose that M, N_1 and N_2 have bases $\{u_1,\ldots,u_n\}$, $\{v_1,\ldots,v_m\}$ and $\{w_1,\ldots,w_m\}$ respectively. Let

$$(1.12.13) \qquad v_i = \sum_{j=1}^{n} a_{ij}u_j, \; w_i = \sum_{j=1}^{n} b_{ij}u_j, \; i=1,\ldots,m, \; A = (a_{ij}), \; B = (b_{ij}) \; .$$

Show that N_1 and N_2 are equivalent if and only if $A \sim B$.

(1.12.14) Let $N \subseteq M$ be D modules with bases. Assume that N has a division property. That is if $ax \in N$ for $0 \neq a \in D$, $x \in M$ then $x \in N$. Show that if D is EDD and N has a division property then any basis in N can be completed to a basis in M.

1.13 Applications to the ring of local anayltic functions in one variable

In this section we consider the applications of the Smith normal form to the system of linear equations over the ring of local analytic functions in one variable in the neighborhood of the origin. According to (1.1.13) this ring is denoted by H_0. In Section 1.3 we showed that the only irreducible element in H_0 is z. Let $A \in M_{mn}(H_0)$. Then $A = A(z)$ and $A(z)$ has the McLaurin expansion

$$A(z) = \sum_{k=0}^{\infty} A_k z^k \tag{1.13.1}$$

which converges in some disc $|z| < R(A)$. Here $R(A)$ is a positive number which depends on A. That is each entry $a_{ij}(z)$ of A has convergent McLaurin series at least for $|z| < R(A)$.

Notations and Definitions. Let $A \in M_{mn}(H_0)$. Then the local invariant polynomials (the invariant factors) of A are normalized to be

$$i_k(A) = z^{i_k}, \; 0 \leq i_1 \leq \ldots \leq i_r, \; r = r(A) \quad . \tag{1.13.3}$$

The number i_r is called the index of A and is denoted by $\eta = \eta(A)$. For a non-negative p denote by $\kappa_p = \kappa_p(A)$ - the number oflocal invariant polynomials of A whose degree equals to p.

We start with the following perturbation result.

Lemma 1.13.4. Let $A, B \in M_{mn}(H_0)$. Consider the matrix

$$C(z) = A(z) + z^{k+1} B(z) \quad , \tag{1.13.5}$$

where k is a non-negative integer. Then A and B have the same local invariant polynomials up to the degree k. Moreover if k is equal to the index of A, and A and B have the same ranks then A is equivalent to B.

Proof. Without restriction in generality we may assume that A is in the diagonal form

$$(1.13.6) \qquad A(z) = \mathrm{diag}(z^{i_1},\dots,z^{i_r},0,\dots,0) \ .$$

Let $s = \sum_{j=0}^{k} \kappa_j(A)$. For $t \le s$ any $t \times t$ minor of $C(z) = (c_{ij}(z))$ which does not contain the first t diagonal elements $c_{11}(z),\dots,c_{tt}(z)$ is divisible at least by $z^{i_1+\cdots+i_{t-1}+k+1}$. On the other hand the minor of $C(z)$ which is composed of the first t rows and columns of $C(z)$ is of the form $z^{i_1+\cdots i_t}(1+z\,0(z))$.

So

$$(1.13.7) \qquad \delta_t(C) = \delta_t(A),\ t = 1,\dots s \ .$$

This proves

$$(1.13.8) \qquad i_t(C) = i_t(A),\ t = 1,\dots s \ .$$

As $s = \sum_{j=1} \kappa_j(A)$ and $i_s = k$ it follows

$$\kappa_j(C) = \kappa_j(A),\ j < k,\ \kappa_k(A) \le \kappa_k(C) \ .$$

Interchange the roles of A and C to deduce

$$\kappa_j(C) = \kappa_j(A),\ j = 0,\ldots,k \quad . \tag{1.13.9}$$

This shows that $A(z)$ and $C(z)$ have the same local invariant polynomials up to the degree k. Suppose that $r(A) = r(C)$. Then (1.13.8) implies that $A(z)$ and $C(z)$ have the same invariant polynomials. That proves that $A \sim C$.

□

Consider a system of linear equations over H_0.

$$A(z)u = b(z),\ A(z) \in M_{mn}(H_0),\ b(z) \in M_{m1}(H_0) \quad , \tag{1.13.10}$$

where we look for a solution $u(z) \in M_{n1}(H_0)$. According to Theorem 1.11.7 the system (1.13.10) is solvable if and only if $r(A) = r(\hat{A})$ and the g.c.d. of all $r \times r$ minors of A and $\hat{A}$ are the same. In theory of analytic functions it is common to try to solve (1.3.10) by the method of power series. That is assume that $A(z)$ has an expansion (1.13.1) and $b(z)$ has an expansion

$$b(z) = \sum_{k=0}^{\infty} b^{(k)} z^k. \tag{1.13.11}$$

Then one looks for a formal solution

$$u(z) = \sum_{k=0}^{\infty} u^{(k)} z^k,\ u^{(k)} \in M_{mn}(C),\ k = 0,1,\ldots, \tag{1.13.12}$$

which satisfies

(1.13.13) $$\sum_{j=0}^{k} A_{k-j}u^{(j)} = b^{(k)}$$

for $k = 0,1,2,\ldots$. A vector $u(z)$ is called a formal solution of (1.13.10) if (1.13.13) holds for any non-negative k. A vector $u(z)$ is called (analytic) solution if $u(z)$ is a formal solution and the series (1.13.12) converge in some neighborhood of the origin, i.e. $u(z) \in M_{n1}(H_0)$. We now give the precise conditions under which the system (1.13.13) is solvable for $k = 0,1,\ldots,q$.

Theorem 1.13.14. *Consider the system* (1.13.13) *for* $k = 0,1,\ldots,q$. *Then this system is solvable if and only if* $A(z)$ *and* $\hat{A}(z)$ *have the same local invariant polynomials up to the degree* q, *that is*

(1.13.15) $$\kappa_j(A) = \kappa_j(\hat{A}),\ j = 0,\ldots,q \quad .$$

Assume that the system (1.3.10) *is solvable over* H_0. *Let* $q = \eta(A)$ *and suppose that* $u^{(0)},\ldots,u^{(q)}$ *satisfies* (1.13.13) *for* $k = 0,\ldots,q$. *Then there exists* $u(z) \in M_{n1}(H_0)$ *satisfying* (1.13.10) *such that* $u(0) = u^{(0)}$.

Let $W_q \subseteq \mathbb{C}^n$ *be the subspace of all vectors* $w^{(0)}$, *such that* $w^{(0)},\ldots,w^{(q)}$ *is a solution of the homogeneous system.*

(1.13.16) $$\sum_{j=0}^{k} A_{kj}w^{(j)} = 0 \quad .$$

Then

(1.13.17) $$\dim W_q = n - \sum_{j=0}^{q} \kappa_j(A) \quad .$$

In particular if $\eta = \eta(A)$ *then for any* $w^{(0)} \in W_\eta$ *there exists* $w(x) \in M_{n1}(H_0)$ *such that*

$$A(x)w(x) = 0,\ w(0) = w^{(0)} \quad . \tag{1.13.18}$$

Proof. We first establish the theorem when $A(z)$ is in the Smith normal form (1.13.6). In that case the system (1.13.13) reduces to

$$u_s^{(k-i_s)} = b_s^{(k)} \ \text{ if } \ i_s \le k,\ 0 = b_s^{(k)} \ \text{ if either } \ i_s > k \ \text{ or } \ s > r(A) \quad . \tag{1.13.19}$$

The the above equations are solvable for $k = 0,\ldots,q$ if and only if z^{i_s} divides $b_s(z)$ for all $i_s \le q$ and for $i_s > q$ z^{q+1} divides $b_s(z)$. If $i_s \le q$ then substract from the last column of $\hat{A}$ the s - the column times $b_s(z)|z^{i_s}$. So $\hat{A}$ is equivalent to the matrix

$$A_1(z) = \operatorname{diag}(z^{i_1},\ldots,z^{i_\ell}) \oplus z^{q+1}A_2(z),\ \ell = \sum_{j=0}^{q} \kappa_j(A),\ A_1 \in M_{(m-\ell)(n+1-\ell)}(H_0) \quad .$$

According to Problem (1.12.10) the local invariant polynomials of $A_1(z)$ whose degree does not exceed q are $z^{i_1},\ldots,z^{i_\ell}$. So $A(z)$ and $A_1(z)$ have the same local invariant polynomials up to degree q. Thus we proved that (1.13.19) is solvable for $k = 0,\ldots,q$ if and only if $A(z)$ and $\hat{A}(z)$ have the same local invariant polynomials up to degree q. Assume next that (1.13.10) is solvable. Since $A(z)$ is of the form (1.13.6) the general solution of (1.13.10) in that case is

$$u_j(z) = b_i(z)/z^{i_j},\ j=1,\ldots,r(A),\ u_j(z) \text{ - arbitrary for } j = r(A)+1,\ldots,n \quad .$$

So

$$u_j(0) = b^{(i_j)},\ j=1,\ldots,r(A),\ u_j(0) - \text{arbitrary for } j=r(A)+1,\ldots,n\ .$$

Clearly (1.13.19) implies $u_s^{(0)} = u_s(0)$ for $k = i_s$. The solvability of (1.3.10) implies that $b_s(z) = 0$ for $s > r(A)$. So $u_s^{(0)}$ is not determined from (1.13.19) for $s > r(A)$. This proves the existence of $u(z)$ satisfying (1.13.10) such that $u(0) = u^{(0)}$. Consider the homogeneous system corresponding to (1.13.19) for $k = 0,\ldots,q$. So $u_s^{(0)} = 0$ for $i_s \leq q$ and otherwise $u_s^{(0)}$ is a free variable. This verifies (1.13.17). Finally as the homogeneous system (1.13.18) is solvable then for $q = \eta(A)$ if $w^{(0)} \in W_q$, that is we have a solution of the homogeneous system corresponding to (1.13.19) of the form $w^{(0)},\ldots,w^{(q)}$, then as we proved above (1.13.18) follows.

It is left to show that the general case can be reduced to the special one discussed above. According to Theorem 1.12.1 there exist matrices $P \in M_n(H_0)$, $Q \in M_m(H_0)$ such that

$$Q(z)A(z)P(z) = B(z) = \operatorname{diag}(z^{i_1},\ldots,z^{i_r}, 0,\ldots,0),\ r = r(A),\ P(z) = \sum_{k=0}^{\infty} P_k z^k,$$

$$Q(z) = \sum_{k=0}^{\infty} Q_k z^k,\ |P_0| \neq 0,\ |Q_0| \neq 0.$$

Introduce a new set of variables

$v(z),\ v^{(0)},\ v^{(1)},\ldots,\ u(z) = P(z)v(z),\ u^{(k)} = \sum_{j=0}^{k} P_{k-j}v^{(j)},\ k = 0,1,\ldots\ .$ Since $|P_0| \neq 0$ we can express $v(z)$ and $v^{(0)}, v^{(1)},\ldots,$ in terms of $u(z)$ and $u^{(0)}, u^{(1)},\ldots,$ correspondingly. Thus (1.13.10) and (1.13.13) is equivalent to

$$B(z)v(z) = c(z),\ c(z) = Q(z)b(z),\ \sum_{j=0}^{k} B_{k-j}v^{(j)} = c^{(k)},\ k = 0,1,\ldots,q.$$

Now the theorem follows since $A \sim B$ and $\hat{A} \sim \hat{B} = (B,c)$ as $\hat{B} = Q\hat{A}(P \oplus I_1)$.

□

Problems

(1.13.20) Consider the system (1.13.10). This system is said to be solvable in the punctured disc if the system.

$$A(z_0)u(z_0) = b(z_0) \tag{1.13.21}$$

is solvable for any point $0 < |z_0| < R$ (as a linear system over C) for some positive R. Prove that (1.13.10) is solvable in the punctured disc if and only if

$$r(A(z)) = r(\hat{A}(z)) \quad . \tag{1.13.22}$$

That is (1.13.10) is solvable over the quotient field M_0.

(1.13.23) Consider the system (1.13.10). This system is said to be pointwise solvable if the system (1.13.21) is solvable for all $|z_0| < R$ for some positive R. Prove that (1.13.10) is pointwise solvable if and only if in addition to (1.13.22) the equality

$$r(A(0)) = r(\hat{A}(0)) \tag{1.13.24}$$

holds.

(1.13.25) Let $A(z) \in M_{mn}(H_0)$. $A(z)$ is call generic if whenever the system (1.13.10) is pointwise solvable then it is analytically solvable (i.e. there exists $u(z) \in M_{n1}(H_0)$ such that (1.13.10) holds). Prove that $A(z)$ is generic if and only if $\eta(A) \leq 1$.

(1.13.26) Let $A(z) \in M_{mn}(H(\Omega))$, $b(z) \in M_{m1}(H(\Omega))$, $\Omega \subseteq C$. Consider the equation (1.13.10). Show that (1.13.10) has a solution $u(z) \in M_{n1}(H(\Omega))$ if and only if for any $\zeta \in \Omega$ the equation (1.13.10) is solvable over H_ζ, i.e. there exists $u(\zeta) \in M_{n1}(H_\zeta)$ which satisfies (1.13.10). (Use the fact that $H(\Omega)$ is EDD so as in Section 1.12 one may assume that A is in Smith normal form.)

(1.13.27) Let $A(z)$ and $b(z)$ satisfy the assumptions of Problem (1.13.26). $A(z)$ is called generic if whenever the system (1.13.10) is pointwise solvable, i.e. 1.13.21 is solvable for any $z_0 \in \Omega$, then there exists a solution $u(z) \in M_{n1}(H(\Omega))$. Prove that $A(z)$ is generic if and only if the invariant functions (factors) of $A(z)$ have only simple zeros. (ζ is called a simple zero of $f \in H(\Omega)$ if $f(\zeta) = 0$ and $f'(\zeta) \neq 0$).

(1.13.28) Let $A \in M_{mn}(H(\Omega))$, $\Omega \subseteq C$. Prove that all the invariant factors of A are trivial if and only if

(1.13.29) $$r(A(\zeta)) = r(A) \quad \text{for all} \quad \zeta \in \Omega \ .$$

(1.13.30) Let $A \in M_{mn}(H(\Omega))$, $\Omega \subseteq C$. Assume that (1.13.29) holds. Using Theorem 1.12.7 prove the existence of n vectors $x^1,\ldots,x^m \in M_{m1}(H(\Omega))$, such that $|x^1(\zeta),\ldots,x^m(\zeta)| \neq 0$ for all $\zeta \in \Omega$ and $AC^n = [x^1,\ldots,x^r]$, $r = r(A)$.

1.14 Strict equivalence of pencils

Definition 1.14.1. *A matrix* $A(x) \in M_{mn}(\mathbf{D}[x])$ *is called pencil if*

$$(1.14.2) \qquad A(x) = A_0 + xA_1,\ A_0, A_1 \in M_{mn}(\mathbf{D}) \quad .$$

A pencil $A(x)$ *is called regular if*

$$(1.14.3) \qquad A(x) \in M_n(\mathbf{D}[x]),\ |A(x)| \neq 0 \quad .$$

Otherwise the pencil is called singular. Two pencils $A(x), B(x) \in M_{mn}(\mathbf{D}[x])$ *are called strictly equivalent if*

$$(1.14.4) \qquad B(x) = QA(x)P,\ P \in UM_n(\mathbf{D}),\ Q \in UM_m(\mathbf{D}) \quad .$$

We denote this relation by $A(x) \underset{s}{\sim} B(x)$.

The classical works of Weierstrass [1867] and Kronecker [1890], see also Gantmacher [1959], classify the equivalence classes of pencils under the strict equivalence relation in case that $\mathbf{D}$ is a field $\mathbf{F}$.

We now give a short account of their main results. First we note that if $A(x) \underset{s}{\sim} B(x)$ then $A(x) \sim B(x)$ over the domain $\mathbf{D}[x]$. In fact we have little more. Put

$$(1.14.5) \qquad B(x) = B_0 + xB_1 \quad .$$

Then the condition (1.14.4) is equivalent to

$$(1.14.6) \qquad B_0 = QA_0P,\ B_1 = QA_1P,\ P \in UM_n(\mathbf{D}),\ Q \in UM_m(\mathbf{D}) \quad .$$

So we can interchange A_0 with A_1 and B_0 with B_1 without affecting the strict equivalence relation. Thus it is natural to consider a *homogeneous* pencil

(1.14.7) $$A(x_0,x_1) = x_0A_0 + x_1A_1 \quad .$$

Suppose that D is a unique factorization domain. So (e.g. Lang [1967]) $D[x_0,x_1]$ is UFD which implies that $D[x_0,x_1]$ is GCDD. So we can define the invariant determinants $\delta_k(x_0,x_1)$ and the invariant factors $i_k(x_0,x_1)$, $k = 1,\ldots,r(A)$, for the homogeneous pencil $A(x_0,x_1)$.

Lemma 1.14.8. Let $A(x_0,x_1)$ be a homogeneous pencil over UFD$[x_0,x_1]$. Then the invariant determinants $\delta_k(x_0,x_1)$ and the invariants factors $i_k(x_0,x_1)$, $k = 1,\ldots,r(A)$ are homogeneous polynomials. Moreover, if $\delta_k(x)$ and $i_k(x)$ are the invariant determinants and factors of the pencil $A(x)$, $k = 1,\ldots,r(A)$, then

(1.14.9) $$\delta_k(x) = \delta_k(1,x),\ i_k(x) = i_k(1,x),\ k = 1,\ldots,r(A) \quad .$$

Proof. Clearly any $k \times k$ minor of $A(x_0,x_1)$ is either zero or a homogeneous polynomial of degree k. Thus, in view of Problem 1.14.24 we deduce that the g.c.d. of all non-vanishing $k \times k$ minors is a homogeneous polynomial $\delta_k(x_0,x_1)$. As $i_k(x_0,x_1) = \delta_k(x_0,x_1)/\delta_{k-1}(x_0,x_1)$ Problem 1.14.24 implies that $i_k(x_0,x_1)$ is a homogeneous polynomial. Consider the pencil $A(x)$ which is given in terms of the homogeneous pencil $A(x_0,x_1)$ as

(1.14.10) $$A(x) = A(1,x) \quad .$$

So $\delta_k(x)$ - the g.c.d. of $k \times k$ minors of $A(x)$ is obviously divisible by $\delta_k(1,x)$. On the other hand we have the following relation between the minors of $A(x_0,x_1)$ and $A(x)$

(1.14.11) $$A(x_0,x_1)[\alpha|\beta] = x_0^k A(\frac{x_1}{x_0})[\alpha|\beta],\ \alpha,\beta \in Q_{k,n}$$

This shows that $x_0^{\rho_k}\delta_k(\frac{x_1}{x_0})$ (ρ_k = deg $\delta_k(x)$) divides any $k \times k$ minor of $A(x_0,x_1)$. So $x^{\rho_k}\delta_k(\frac{x_1}{x_0}) \Big| \delta_k(x_0,x_1)$. This proves the first part of (1.14.9). So

$$\delta_k(x_0,x_1) = x_0^{\varphi_k}[x_0^{\rho_k}\delta_k(\frac{x_1}{x_0})], \; \rho_k = \deg \delta_k(x), \; \varphi_k \geq 0. \tag{1.14.12}$$

Now the equality

$$i_k(x_0,x_1) = \delta_k(x_0,x_1)/\delta_{k-1}(x_0,x_1)$$

implies

$$i_k(x_0,x_1) = x_0^{\psi_k}[x_0^{\sigma_k}i_k(\frac{x_1}{x_0})], \; \sigma_k = \deg i_k(x), \; \psi_k \geq 0 \;. \tag{1.14.13}$$

This establishes the lemma.

□

We call $\delta_k(x_0,x_1)$ and $i_k(x_0,x_1)$ the invariant homogeneous determinants and the invariant homogeneous polynomials (fac ors) respectively.

The classical result due to Weierstrass [1867] states:

Theorem 1.14.14. Let $A(x) \in M_n(\mathbb{F}[x])$ be a regular pencil. Then a pencil $B(x)$ is strictly equivalent to $A(x)$ if and only if $A(x)$ and $B(x)$ have the same invariant homogeneous polynomials.

Proof. The necessary part of the theorem holds for any $A(x)$, $B(x)$ which are strictly equivalent. Suppose now that $A(x)$ and $B(x)$ have the same invariant homogeneous polynomials. According to (1.14.9) the pencils $A(x)$ and $B(x)$ have the same invariant polynomials. So $A(x) \sim B(x)$ over $\mathbb{F}[x]$. Therefore

$$W(x)B(x) = A(x)U(x), \; U(x),W(x) \in M_n(\mathbb{F}[x]), \quad |U(x)| = \text{Const} \neq 0, \; |W(x)| = \text{Const} \neq 0 \;. \tag{1.14.15}$$

Assume now that A_1 and B_1 are non-singular. Then (see Problem 1.14.25) it is possible to divide $W(x)$ by $A(x)$ from right and $U(x)$ by $B(x)$ from left

$$W(x) = A(x)W_1(x) + R,\ U(x) = U_1(x)B(x) + P \quad , \tag{1.14.16}$$

where P and R are constant matrices. So

$$A(x)(W_1(x) - U_1(x))B(x) = A(x)P - RB(x) \quad .$$

As $|A_1|\ |B_1| \neq 0$ we must have $W_1(x) = U_1(x)$ otherwise the left-hand side of the above equality would be of degree at least 2 (see Definition 1.14.19) while the right-hand side of this equality is at most of degree 1. So

$$W_1(x)U_1(x),\quad RB(x) = A(x)P \quad . \tag{1.14.17}$$

It is left to show that P and R are non-singular. As $W(x)$ is unimodular there exists $V(x) \in UM_n(\mathbb{F}[x])$ such that $I = W(x)U(x)$.

Let

$$V(x) = B(x)V_1(x) + S \quad .$$

So

$$\begin{aligned} I &= (A(x)W_1(x) + R)V(x) = A(x)W_1(x)V(x) + RV(x) = \\ &= A(x)W_1(x)V(x) + RB(x)V_1(x) + RS = \\ &= A(x)W_1(x)V(x) + A(x)PV_1(x) + RS = \\ &= A(x)[W_1(x)V(x) + PV_1(x)] + RS \ , \end{aligned}$$

where we used the second equality in (1.14.17). Since $|A_1| \neq 0$ the above equality implies

$$W_1(x)V(x) + PV_1(x) = 0, \quad RS = I \quad .$$

So R is invertible. The same arguments show that P is invertible. Thus $A(x)$ and $B(x)$ are strictly equivalent if $|A_1B_1| \neq 0$.

Consider now the general case. Introduce a new variables y_0, y_1

$$y_0 = ax_0 + bx_1, \; y_1 = cx_0 + dx_1, \; ad - cb \neq 0 \quad .$$

Then

$$A(y_0,y_1) = y_0A_0' + y_1A_1', \; B(y_0,y_1) = y_0B_0' + y_1B_1' \quad .$$

Clearly $A(y_0,y_1)$ and $B(y_0,y_1)$ have the same invariant homogeneous polynomials. Also $A(y_0,y_1) \underset{s}{\sim} B(y_0,y_1)$ if and only if $A(x_0,x_1) \underset{s}{\sim} B(x_0,x_1)$. Since $A(x_0,x_1)$ and $B(x_0,x_1)$ are regular pencils it is possible to choose a,b,c,d such that A_1' and B_1' are non-singular. This shows that $A(y_0,y_1) \underset{s}{\sim} B(y_0,y_1)$ according to the previous case. So $A(x) \underset{s}{\sim} B(x)$.

In fact, we also proved

Corollary 1.14.18. Let $A(x),B(x) \in M_n(\mathbf{F}[x])$. Assume that A_1 and B_1 are non-singular. Then the pencils $A(x)$ and $B(x)$ are strictly equivalent if and only if $A(x)$ and $B(x)$ are equivalent.

For singular pencils the invariant homogeneous polynomials alone do not determine the class of strictly equivalent pencils as in the case of regular pencils.

We now introduce the notion of column and row indices for $A(x) \in M_{mn}(\mathbf{F}[x])$. Consider the system (1.13.18). The set of all solutions $w(x)$ is a $\mathbf{F}[x]$-module $\mathbf{M}$ with a finite basis $w_1(x),\ldots,w_s(x)$. (Theorem 1.11.12).

To specify a choice of basis we need the following definition.

Definition 1.14.19. Let $A \in M_{mn}(\mathbf{D}[x_1,\ldots,x_k])$. So

(1.14.20)
$$A(x_1,\ldots,x_k) = \sum_{|\alpha| \le d} A_\alpha x^\alpha, \quad A_\alpha \in M_{mn}(\mathbf{D})$$
$$\alpha = (\alpha_1,\ldots,\alpha_k) \in Z_+^k, \quad |\alpha| = \sum_{i=1}^{k} \alpha_i, \quad x^\alpha = x_1^{\alpha_1} \cdots x_k^{\alpha_k} .$$

Then the degree of $A(x_1,\ldots,x_k)$ (deg A) is d if there exists $A_\alpha \ne 0$ with $|\alpha| = d$.

Definition 1.14.21. Let $A \in M_{mn}(\mathbf{F}[x])$ and consider the module $M \subseteq \mathbf{F}[x]^n$ of all solutions of (1.13.18). Choose a basis $w_1(x),\ldots w_s(x)$, $s = n - r(A)$ in M such that $w_k(x) \in M$ has the lowest degree among $w(x) \in M$ which are linearly independent (over the quotient field of $\mathbf{F}[x]$) of $w_1(x),\ldots,w_{k-1}(x)$ for $k = 1,\ldots,s$. Then the column indices $\alpha_1 \le \alpha_2 \le \ldots \le \alpha_s$ of $A(x)$ are given as

(1.14.22)
$$\alpha_k = \deg w_k(x), \quad k = 1,\ldots,s .$$

The row indices $0 \le \beta_1 \le \ldots \le \beta_t$, $t = m-r(A)$, of $A(x)$ are the column indices of the transposed matrix $A^t(x)$.

It can be shown (e.g. Gantmacher [1959]) that the column (row) indices are independent of a particular choice of a basis $w_1(x),\ldots,w_s(x)$. We now state the Kronecker result [1890].

Theorem 1.14.23. The pencils $A(x), B(x) \in M_{mn}(\mathbf{F}[x])$ are strictly equivalent if and only if they have the same invariant homogeneous polynomials and the same row and column indices.

See for example Gantmacher [1959] for a proof of this theorem.

Problems

(1.14.24) Using the fact $\mathbf{UFD}[x_1,\ldots,x_n]$ is $\mathbf{UFD}$ and the equality (1.1.26) show that if $a \in \mathbf{UFD}[x_1,\ldots,x_n]$ is a homogeneous polynomials then in the decomposition (1.3.3) each p_i is a homogeneous polynomial.

(1.14.25) Let

(1.14.26) $$W(x) = \sum_{k=0}^{q} W_k x^k, \; U(x) = \sum_{k=0}^{p} U_k x^k \quad .$$

Assume that $A(x)$ is a pencil (1.14.2) such that A_1 is a square non-singular matrix. Show that if $p,q \geq 1$ then

$$W(x) = A(x)A_1^{-1}[W_q x^{q-1}] + \tilde{W}(x), \; U(x) = [U_p x^{p-1}]A_1^{-1}A(x) + \tilde{U}(x) \quad ,$$

where

$$\deg \tilde{W}(x) < q, \; \deg \tilde{U}(x) < p \quad .$$

Prove the equalities (1.14.16) where R and P are constant matrices. Suppose that $A_1 = I$. Show that R and P in (1.14.16) can be given as

(1.14.27) $$R = \sum_{k=0}^{q} (-A_0)^k W_k, \; P = \sum_{k=0}^{p} U_k (-A_0)^k \quad .$$

(1.14.28) Let $A(x)$ be a regular pencil such that $|A_1| \neq 0$. Prove that in (1.14.12) and (1.14.13) ${}_k = \psi_k = 0$, $k = 1,\ldots,n$. (Use the equality (1.11.20) for $A(x)$ and $A(x_0,x_1)$.

(1.14.29) Consider the following two pencils

$$A(x) = \begin{pmatrix} 2+x & 1+x & 3+2x \\ 3+x & 2+x & 5+2x \\ 3+x & 2+x & 6+3x \end{pmatrix}, \quad B(x) = \begin{pmatrix} 2+x & 1+x & 1+x \\ 1+x & 2+x & 1+x \\ 1+x & 1+x & 1+x \end{pmatrix}$$

over $R[x]$. Show that $A(x)$ and $B(x)$ are equivalent but not strictly equivalent.

(1.14.30) Let

$$A(x) = \sum_{k=0}^{p} A_k x^k \in M_{mn}(\mathbf{C}[x]) \quad .$$

Put

$$A(x_0,x_1) = \sum_{k=0}^{q} A_k x_0^{q-k} x_1^k$$

where $q = 0$ if $A(x) = 0$ and $A_q \neq 0$, $A_k = 0$, $q < k \leq p$ if $A(x) \neq 0$. Let $i_k(x_0,x_1)$, $k = 1,\ldots,r(A)$ be the invariant factors of $A(x_0,x_1)$. Prove that $i_k(x_0,x_1)$ is a homogeneous polynomial, $k = 1,\ldots,r(A)$. Show that $i_k(1,x)$, $k = 1,\ldots,r(A)$, are the invariant factors of $A(x)$.

(1.14.31) Let $A(x),B(x) \in M_{mn}(\mathbf{C}[x])$. $A(x)$ and $B(x)$ are called strictly equivalent ($A \underset{s}{\sim} B$) if $B(x) = PA(x)Q$, $P \in M_m(\mathbf{C})$, $Q \in M_n(\mathbf{C})$, $(|P|\ |Q| \neq 0)$. Prove that if

$$A \underset{s}{\sim} B \text{ then the } A(x_0,x_1) \text{ and } B(x_0,x_1)$$

have the same invariant factors.

(1.14.32) Prove that the pencils $A(x)$ and $B(x)$ are strictly equivalent if and only if $A^t(x) \underset{s}{\sim} B^t(x)$.

1.15 Similarity of matrices

Definition 1.15.1. Let $A, B \in M_m(D)$. The matrices A and B are called similar $(A \quad B)$ if

$$B = QAQ^{-1} , \tag{1.15.2}$$

for some $Q \in UM_m(D)$.

Clearly the similarity relation is an equivalence relation. So $M_m(D)$ is divided to equivalence classes which are called the similarity classes. It is a standard fact that each similarity class corresponds to all possible representations of some $T \in \mathrm{Hom}(M,M)$, where M is a D-module having a basis of n elements. Indeed, let $u^1,\ldots,u^m$ be a basis in M. Then T is represented by $A = (a_{ij}) \in M_m(D)$

$$Tu^i = \sum_{j=1}^{m} a_{ij}u^j, \quad i = 1,\ldots,m \ . \tag{1.15.3}$$

Let $\tilde{u}^1,\ldots,\tilde{u}^m$ be another basis in M. Assume that Q is a unimodular matrix which is given by (1.9.9). Then according to (1.15.3) and the arguments of Section 1.9, the representation of T in the basis $\tilde{u}^1,\ldots,\tilde{u}^m$ is given by the matrix B of the form (1.15.2).

The similarity notion of matrices is closely related to the strict equivalency of certain regular pencils.

Lemma 1.15.4. Let $A, B \in M_m(D)$ and associate with these matrices the following regular pencils

$$A(x) = -A+xI, \quad B(x) = -B+xI \ . \tag{1.15.5}$$

Then A and B are similar if and only if the pencils $A(x)$ and $B(x)$ are strictly equivalent.

Proof. Assume first that $A \approx B$. Then (1.15.2) implies (1.14.4) where $P = Q^{-1}$. Suppose now that $A(x) \underset{s}{\sim} B(x)$. So

$$B = QAP, \; QP = I \quad .$$

That is $P = Q^{-1}$ and $A \approx B$.

□

Clearly is $A(x) \underset{s}{\sim} B(x)$ then $A(x) \sim B(x)$. So we have

Corollary 1.15.6. Let $A,B \in M_n(D)$. Assume that D is a unique factorization domain. Assume that A and B are similar then the corresponding pencils $A(x)$ and $B(x)$ given by (1.15.5) have the same invariant polynomials.

In case that $D = F$ the above condition is also a sufficient condition in view of Lemma 1.15.4 and Corollary 1.14.18.

Theorem 1.15.7. Let $A,B \in M_n(F)$. Then A and B are similar if and only if the pencils $A(x)$ and $B(x)$ given by (1.15.5) have the same invariant polynomials.

It can be shown (see Problem 1.15.8) that even over Euclidean domains the condition that $A(x)$ and $B(x)$ have the same invariant polynomials does not imply in general that $A \approx B$.

Problems

(1.15.8) Let

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}, \; B = \begin{pmatrix} 1 & 1 \\ 0 & 5 \end{pmatrix} \in M_2(\mathbb{Z}) \quad .$$

Show that $A(x)$ and $B(x)$ given by (1.15.5) have the same invariant polynomials over $\mathbb{Z}[x]$. Prove that A and B are not similar over $\mathbb{Z}$.

(1.15.9) Let $A(x) \in M_n(\mathrm{UFD}[x])$ be given by (1.15.5). Let $i_1(x),\dots,i_n(x)$ be the invariant polynomials of $A(x)$. Using the equality (1.11.20) prove that each $i_k(x)$ can be assumed to be normalized polynomial and

$$\sum_{k=1}^{n} \deg i_k(x) = n \quad . \tag{1.15.10}$$

(1.15.11) Let $A \in M_n(F)$. Show that $A \approx A^t$.

1.16 The companion matrix

Theorem 1.15.7 shows that the invariant polynomials of $xI-A$ determine the similarity class of A. We now show that any set of normalized polynomials $i_1(x),\ldots,i_n(x) \in$ UFD[x] such that $i_j(x)|i_{j+1}(x)$, $j = 1,\ldots,n-1$ and which satisfy (1.15.10) are invariant polynomials of $xI-A$ for some $A \in M_n(\mathrm{UFD})$. To do so we introduce the notion of a companion matrix.

Definition 1.16.1. Let $p(x) \in D[x]$ be a normalized polynomial

$$p(x) = x^m + a_1x^{m-1}+\ldots+a_m \quad .$$

Then $C(p) = (c_{ij})_1^m \in M_m(D)$ is called the companion matrix of $p(x)$ if

$$(1.16.2)\quad c_{ij} = \delta_{(i+1)j}, \ i = 1,\ldots,m-1, \ j = 1,\ldots,m, \ c_{mj} = -a_{m-j+1}, \ j = 1,\ldots,m \quad .$$

Lemma 1.16.3. Let $p(x) \in$ UFD[x] be a normalized polynomial of degree m. Consider the pencil $C(x) = xI - C(p)$. Then the invariant polynomials of $C(x)$ are

$$(1.16.4)\qquad i_1(C) =\ldots= i_{m-1}(C) = 1, \ i_m(C) = p(x) \quad .$$

Proof. For $k < m$ consider a minor of $C(x)$ composed of the rows $1,\ldots,k$ and columns $2,\ldots,k+1$. Since this minor is the determinant of a lower triangular matrix with -1 on the main diagonal we deduce that its value is $(-1)^k$. So $\delta_k(C(x)) = 1$, $k = 1,\ldots,m-1$. This establishes the first equality in (1.16.4). Clearly, $\delta_m(C(x)) = |xI-C|$. Expand the determinant of $C(x)$ by the first row and use the induction hypothesis to prove $|xI-C| = p(x)$. This shows $i_m(C) = \delta_m(C)/\delta_{m-1}(C) = p(x)$.

Using the results of Problem 1.12.10 and Lemma 1.16.3 we get

Theorem 1.16.5. Let $p_j(x) \in$ UFD[x], $j = 1,\ldots,k$ be normalized polynomials of positive degrees such that $p_j(x)|p_{j+1}(x)$, $j = 1,\ldots,k-1$. Consider the matrix

$$C(p_1,\dots,p_k) = \sum_{j=1}^{k} \oplus\, C(p_j) \quad . \tag{1.16.6}$$

Then the non-trivial invariant polynomials of $xI - C(p_1,\dots,p_k)$ (i.e. those polynomials which are not the identity element) are $p_1(x),\dots,p_k(x)$.

Combining Theorems 1.15.7 and 1.16.5 we obtain a canonical representation for the similarity class in $M_n(\mathbf{F})$.

Theorem 1.16.7. Let $A \in M_n(\mathbf{F})$ and assume that $p_j(x) \in \mathbf{F}[x]$, $j = 1,\dots,k$ are the non-trivial normalized polynomials of $xI-A$. Then A is similar to $C(p_1,\dots,p_k)$.

Definition 1.16.8. For $A \in M_n(\mathbf{F})$ the matrix $C(p_1,\dots,p_k)$ is called the rational canonical form of A.

Let $\mathbf{D}$ be an integral domain and denote by $\mathbf{F}$ its quotient field. Let $A \in M_n(\mathbf{D})$. So $A \in M_n(\mathbf{F})$ and let $C(p_1,\dots,p_k)$ be the rational canonical form of A. We now examine the case when $A(p_1,\dots,p_k) \in M_n(\mathbf{D})$. Assume first that $\mathbf{D}$ is UFD. Let δ_k be the g.c.d. of $k \times k$ $xI-A$. So δ_k divides a minor of $p(x) = (xI-A)[\alpha|\alpha]$, $\alpha = \{1,\dots,k\}$. Clearly $p(x)$ is a normalized polynomial of degree k. Recall that $\mathbf{D}[x]$ is also UFD. (See Section 1.4.)

According to Theorem 1.4.12 the decomposition of $p(x)$ into irreducible factors in $\mathbf{D}[x]$ if of the form (1.4.9) where $a = 1$ and each $q_i(x)$ is a non-trivial normalized and irreducible polynomial in $\mathbf{D}[x]$. Since δ_k is a product of some irreducible factors of $p(x)$ then either $\delta_k = 1$ or δ_k is a non-trivial normalized polynomial in $\mathbf{D}[x]$. The same argument shows that $i_k = \delta_k/\delta_{k-1}$ is either identity or a non-trivial polynomial in $\mathbf{D}[x]$. Thus we demonstrated.

Theorem 1.16.9. Let $A \in M_n(\mathbf{D})$. Assume that $\mathbf{D}$ is a unique factorization domain. Then the rational canonical form $C(p_1,\dots,p_k)$ over the quotient field $\mathbf{F}$ belongs to $M_n(\mathbf{D})$.

In particular, we have

Corollary 1.16.10. Let $A \in M_n(\mathbf{C}[x_1,\dots,x_m])$. Then the rational canonical form A belongs to $M_n(\mathbf{C}[x_1,\dots,x_n])$.

Using the results of Theorem 1.4.13 we deduce that Theorem 1.16.9 applies to the ring of analytic functions in several variables although this ring is not UFD (see Section 1.3).

Theorem 1.16.11. *Let* $A \in M_n(H(\Omega))(\Omega \subseteq C^m)$. *Then the rational canonical form of* A *over the field of meromorphic functions belongs to* $M_n(H(\Omega))$.

Problems

(1.16.12) Let $p(x) \in UFD[x]$ be a normalized non-trivial polynomial. Assume $p(x) = p_1(x)p_2(x)$. where $p_i(x)$ is normalized non-trivial polynomial in $UFD[x]$ for $i = 1,2$. Using Problems 1.12.9 and 1.12.10 show that $xI - C(p_1,p_2)$ given by 1.16.6 has the same invariant polynomials as $xI - C(p)$ if and only if $(p_1,p_2) = 1$.

(1.16.13) Let $A \in M_n(UFD)$ and assume that $p_1(x),\ldots,p_k(x)$ are the non-trivial normalized invariant polynomials of $xI-A$. Let

$$p_j(x) = [\varphi_1(x)]^{m_{1j}}\cdots[\varphi_\ell(x)]^{m_{\ell j}},\ j = 1,\ldots,k \tag{1.16.14}$$

where $\varphi_1(x),\ldots,\varphi_\ell(x)$ are non-trivial normalized irreducible polynomials in $UFD[x]$ such that $(\varphi_i,\varphi_j) = 1$ for $i \neq j$. Prove that

$$\begin{gathered} m_{ik} \geq 1,\ i = 1,\ldots,\ell,\quad m_{ik} \geq m_{i(k-1)} \geq \cdots \geq m_{i1} \geq 0\ , \\ \sum_{i,j=1}^{\ell,k} m_{ij} = n\ . \end{gathered} \tag{1.16.15}$$

The polynomials $\varphi_i^{m_{ij}}$, for $m_{ij} > 0$ are called the elementary divisors of $xI-A$. Using the above problem show that $xI-A$ and $xI-E$ where

$$E = \sum_{m_{ij}>0} \oplus\, C(\varphi_i^{m_{ij}}) \tag{1.16.16}$$

have the same invariant polynomials. Thus over a field $\mathbb{F}$ $A \approx E$. Sometimes E is called the rational canonical form of A.

1.17 Splitting to invariant subspaces

Let V be a vector space of dimension m over a field F. Denote by $L(V)$ the vector space of all linear transformations $T : V \to V$. That is

$$L(V) = \mathrm{Hom}(V,V) \quad . \tag{1.17.1}$$

Let $T \in L(V)$. As we pointed out in Section 1.15 the set of all matrices $\mathcal{A} \subseteq M_m(F)$ which represent T in different bases is exactly an equivalence class of matrices with respect to the similarity relation. Theorem 1.15.7 shows that the class $\mathcal{A}$ is characterized by the invariant polynomials of $xI-A$ for some $A \in \mathcal{A}$. Since $xI-A$ and $xI-B$ have the same invariant polynomials if and only if $A \approx B$ we define.

Definition 1.17.2. Let $T \in L(V)$ and let $A \in M_m(F)$ be a representation matrix of T in a basis $u^1,\ldots,u^m$ given by the equality (1.15.3). Then the invariant polynomials $p_1(x),\ldots,p_m(x)$ of T are defined as the invariant polynomials of $xI-A$. The characteristic polynomials of T - is the polynomial $|xI-A|$.

The fact that the characteristic polynomial of T is independent of a representation matrix A follows from the identity (1.11.20)

$$|xI-A| = p_1(x)\cdots p_k(x) \tag{1.17.3}$$

where $p_1(x),\ldots,p_k(x)$ are non-trivial invariant polynomials of $xI-A$. In Section 1.16 we proved that the matrix $C(p_1,\ldots,p_k)$ is a representation matrix of T. In this section we shall consider another representation matrix A of T which is closely related to the matrix E (1.16.16). This form will be achieved by splitting V to a direct sum

$$V = U_1 \oplus \quad \oplus U_\ell \tag{1.17.4}$$

where each U_j is an invariant subspace of T.

Definition 1.17.5. A subspace $U \subseteq V$ is an invariant subspace of T if

(1.17.6) $$TU \subseteq U \quad .$$

U <u>is called trivial if</u> $U = [0]$. U <u>is called proper if</u> $U \subsetneq V$. U <u>is called irreducible if</u> U <u>cannot be expressed as a direct sum of two non-trivial invariant subspaces of</u> T.

Thus if V splits to a direct sum of non-trivial invariant subspaces of T then a direct sum of matrix representations of the restrictions of T to U_j gives a representation matrix of T. So, a simple representation of T can be achieved by splitting V to a direct sum of irreducible invariant subspaces. To do so we need to introduce the notion of the minimal polynomial of T. Consider the linear operators $I, T, T^2, \ldots, T^{m^2}$, where I is the identity operator $(Ix=x)$. Since the dimension of L(V) is m^2 these $m^2 + 1$ operators are linearly dependent. So there exists an integer q such that $I, T, \ldots, T^{q-1}$ are independent and $I, T, \ldots, T^q$ are linearly dependent.

<u>Definition</u> 1.17.7. <u>A polynomial</u> $\psi(x) \in F[x]$ <u>is called the minimal polynomial of</u> T <u>if</u> $\psi(x)$ <u>is a normalized polynomial of the smallest degree satisfying</u>

(1.17.8) $$\psi(T) = 0 \quad .$$

Here

$$\phi(T) = \sum_{i=0}^{\ell} c_i T^i, \; \phi(x) = \sum_{i=0}^{\ell} c_i x^i \in F[x]$$

and 0 is the zero operator $(0x=0)$. By the definition $\deg \psi > 1$. The minimal polynomial is characterized by the following property.

<u>Lemma</u> 1.17.9. <u>Assume that</u> T <u>annihilates</u> $\phi \in F[x]$. <u>That is</u> $\phi(T) = 0$. <u>Then</u> $\psi | \phi$.

<u>Proof.</u> Divide ϕ by ψ

$$\phi(x) = \chi(x)\psi(x) + \rho(x), \; \deg \rho < \deg \psi \quad .$$

Now (1.17.8) and the assumption of the lemma imply that $\rho(T) = 0$. As $\deg \rho < \deg \psi$ from the definition of the minimal polynomial we deduce that $\rho(x) = 0$. □

Since $F[x]$ is a unique factorization domain, let

$$(1.17.10) \qquad \psi(x) = \varphi_1(x)^{s_1} \cdots \varphi_\ell(x)^{s_\ell}, \quad (\varphi_i, \varphi_j) = 1, \quad \text{for } 1 \le i < j \le \ell, \quad \deg \varphi_i \geq 1, \; i = 1,\ldots,\ell$$

where each $\varphi_i(x)$ is a normalized irreducible polynomial in $F[x]$.

Theorem 1.17.11. Let $\psi(x)$ be the minimal polynomial of T. Assume that ψ splits to a product of co-prime factors as given in (1.17.10). Then the space V splits to a direct sume (1.17.4) where each U_j is a non-trivial invariant subspace of T. Moreover $\varphi_j^{s_j}(x)$ is the minimal polynomial of the restriction of T to U_j.

The proof of the theorem follows immediately from the lemma below.

Lemma 1.17.12. Let ψ be the minimal polynomial of T. Assume that ψ splits to a product of two co-prime factors

$$(1.17.13) \qquad \psi(x) = \psi_1(x)\psi_2(x), \; \deg \psi_i \geq 1, \; i = 1,2, \; (\psi_1,\psi_2) = 1 \quad ,$$

where each ψ_i is normalized. Then

$$(1.17.14) \qquad V = U_1 \oplus U_2 \quad ,$$

where each U_j is a non-trivial invariant subspace of T and ψ_j is the minimal polynomial of the restriction of T to U_j.

Proof. The assumption of the lemma imply the existence of polynomials $\theta_1(x)$ and $\theta_2(x)$ sucht that

$$(1.17.15) \qquad \theta_1(x)\psi_1(x) + \theta_2(x)\psi_2(x) = 1 \quad .$$

Define

(1.17.16) $$U_j = \{u \mid u \in V,\ \psi_j(T)u = 0\},\ j = 1,2 \quad .$$

Since any two polynomials in T commute, i.e.

$$\mu(T)\nu(T) = \nu(T)\mu(T)$$

we clearly have that each U_j is an invariant subspace of T. The equality (1.17.15) implies

$$I = \psi_1(T)\theta_1(T) + \psi_2(T)\theta_2(T) \quad .$$

That is, for any $u \in V$ we have

$$u = u_1 + u_2,\ u_1 = \psi_2(T)\theta_2(T)u \in U_1, u_2 = \psi_1(T)\theta_1(T)u \in U_2 \quad .$$

So

$$U_1 + U_2 = V \quad .$$

Suppose that $u \in U_1 \cap U_2$. Then

$$\psi_1(T)u = \psi_2(T)u = 0 \quad .$$

Thus

$$\theta_1(T)\psi_1(T)u = \theta_2(T)\psi_2(T)u = 0 \quad .$$

Finally

$$u = [\theta_1(T)\psi_1(T) + \theta_2(T)\psi_2(T)]u = 0$$

which proves that $\mathbf{U}_1 \cap \mathbf{U}_2 = \{0\}$. This establishes (1.17.14). Let T_j be the restriction of T of $\mathbf{U}_j$. By the definition of $\mathbf{U}_j$ (1.17.16) T_j annihilates ψ_j. Let $\bar{\psi}_j$ be the minimal polynomial of T_j. So $\bar{\psi}_j | \psi_j$, $j = 1,2$. Now

$$\bar{\psi}_1(T)\bar{\psi}_2(T)u = \bar{\psi}_1(T)\bar{\psi}_2(T)(u_1 + u_2) = \bar{\psi}_2(T)\bar{\psi}_1(T)u_1 + \bar{\psi}_1(T)\bar{\psi}_2(T)u_2 = 0 \quad .$$

Therefore (1.17.14) yields that T annihilates $\bar{\psi}_1\bar{\psi}_2$. Since $\psi(x)$ is the minimal polynomial of T we have $\psi_1\psi_2 | \bar{\psi}_1\bar{\psi}_2$. This finally implies $\psi_j = \bar{\psi}_j$, $j = 1,2$. Also as $\deg \psi_j \geq 1$ it follows that $\dim \mathbf{U}_j \geq 1$.

Problems

(1.17.17) Assume that (1.17.14) holds, where $T\mathbf{U}_j \subseteq \mathbf{U}_j$, $j = 1,2$. Let ψ_j be the minimal polynomial of the restriction of T to $\mathbf{U}_j$, $j = 1,2$. Prove that the minimal polynomial ψ of T is equal to $\psi_1\psi_2/(\psi_1,\psi_2)$.

(1.17.18) Let the assumptions of Problem (1.17.17) hold. Assume furthermore that $\psi = \phi^s$ where ϕ is irreducible over $\mathbf{F}[x]$. Then either $\psi = \psi_1$ or $\psi = \psi_2$.

(1.17.19) Let $C = C(p) \in M_m(\mathbf{D})$ be the companion matrix given by (1.16.2). Let $\varepsilon_i = (\delta_{i1},\ldots,\delta_{im})^t$, $i = 1,\ldots,m$, be a standard basis in $\mathbf{D}^m$. Show

$$(1.17.20) \qquad C\varepsilon_i = \varepsilon_{i-1} - a_{m-i+1}\varepsilon_m, \; i = 1,\ldots,m \; (\varepsilon_0 = 0) \quad .$$

Prove that $p(C) = 0$ and that any polynomial $0 \neq q(x) \in \mathbf{D}[x]$, $\deg q < m$, is not annihilated by C. (Consider $q(C)\varepsilon_i$ and use (1.17.20).) That is $p(x)$ is the minimal polynomial of $C(p)$.

(1.17.21) Let $A \in M_m(\mathbf{F})$. Using Theorem 1.16.7 and Problems 1.17.17 and 1.17.19 show that the minimal polynomial ψ of A is the last invariant polynomial $xI-A$. That is

$$\psi(x) = |xI-A|/\delta_{m-1}(x) \quad , \tag{1.17.22}$$

where $\delta_{m-1}(x)$ is g.c.d. of all $(m-1) \times (m-1)$ minor of $xI-A$.

(1.17.23) Show that the results of Problem 1.17.22 apply to $A \in M_m(\mathbf{UFD})$. In particular if A B then A and B have the same minimal polynomials.

(1.17.24) Deduce from Problem (1.17.21) the <u>Cayley-Hamilton</u> theorem which states that $T \in L(\mathbf{V})$ annihilates its characteristic polynomial.

(1.17.25) Let $A \in M_m(\mathbf{D})$. Prove that A annihilates its characteristic polynomial. (Prove this result by considering the quotient field of $\mathbf{D}$.)

(1.17.26) Use Problem (1.17.24) and Lemma 1.17.9 to show

$$\deg \psi \leq \dim \mathbf{V} \quad . \tag{1.17.27}$$

(1.17.28) Let $\psi = \varphi^s$ where φ is irreducible in $\mathbf{F}[x]$ and assume that $\deg \psi = \dim V$. Use Problem (1.17.18) and (1.17.26) to show that $\mathbf{V}$ is an irreducible invariant subspace of T.

(1.17.29) Let $p(x) \in \mathbf{F}[x]$ be a non-trivial normalized polynomial such that $p = \varphi^s$ where $\varphi(x)$ is irreducible in $\mathbf{F}[x]$. Let $T \in L(V)$ be represented by $C(p)$. Use Problem (1.17.28) to prove that $\mathbf{V}$ is an irreducible invariant subspace of T.

(1.17.30) Let $T \in L(\mathbf{V})$ and let E be the matrix given by (1.16.16) which is determined by the elementary divisors of T. Using Problem (1.17.29) prove that the representation E of T corresponds to a splitting of $\mathbf{V}$ to a direct sume of irreducible invariant subspaces of T.

(1.17.31) Deduce from Problems (1.17.28) and (1.17.30) that $\mathbf{V}$ is an irreducible invariant subspace if and only if the minimal polynomial ψ satisfies the assumptions of Problem (1.17.28).

1.18 An upper triangular form

Definition 1.18.1. Let M be a D-module and assume that $T \in \mathrm{Hom}(M,M)$. $\lambda \in D$ is called an eigenvalue if there exist $0 \neq u \in M$ such that

(1.18.2) $$Tu = \lambda u \quad .$$

The element (vector) u is called eigenelement (eigenvector) corresponding to λ. An element $0 \neq u$ is called generalized eigenelement (eigenvector) if

(1.18.3) $$(\lambda I-T)^k u = 0$$

for some positive integer k where λ is an eigenvalue of T. For $T \in M_m(D)$ λ is called eigenvalue if (1.18.2) holds for some $0 \neq u \in D^m$. The element u is called eigenelement (eigenvector) or generalized eigenvector if (1.18.2) or (1.18.3) holds respectively.

Lemma 1.18.4. Let $T \in M_m(D)$. Then λ is an eigenvalue of T if and only if λ is a root of the characteristic polynomial $|xI-T|$.

Proof. Let F be the quotient field of D. As (1.18.2) is equivalent to

$$(\lambda I-T)u = 0$$

by the definition $u \neq 0$, so the above system has a non-trivial solution. Therefore $|\lambda I-T| = 0$. That is λ is a root of the characteristic polynomial of T. Vice versa, if $|\lambda I-T| = 0$ then the above system has a non-trivial solution $u \in F^m$. Clearly au, $a \in D$ also satisfies the above equality. Choose $a \neq 0$ such that $au \in D^m$. Thus λ is an eigenvalue of T.

Theorem 1.18.5. Let $T \in M_m(D)$. Assume that the characteristic polynomial of T splits to linear factors over D

(1.18.6) $$|xI-T| = \prod_{i=1}^{m} (x-\lambda_i),\ \lambda_i \in D,\ i = 1,\dots,m\ .$$

Suppose that D is a Bezout domain. Then

(1.18.7) $$T = QAQ^{-1},\ Q \in UM_m(D)$$

where $A = (a_{ij})^m$, is an upper triangular matrix (i.e. $a_{ij} = 0$ for $j < i$) such that $a_{11},\dots,a_{mm}$ are the eigenvalues $\lambda_1,\dots,\lambda_m$ appearing in any specified order.

Proof. Let λ be an eigenvalue of T and consider the set of all $u \in D^m$ which satisfies (1.18.2). Clearly this set is a D-module M. According to Lemma 1.18.4 M contains non-zero vectors. Assume that D is BD. Then according to Theorem 1.11.12 M has a basis $u^1,\dots,u^k$ which can be completed to a basis $u^1,\dots,u^m$ in D^m. Let

(1.18.8) $$Tu^i = \sum_{j=1}^{m} b_{ji}u^j,\ i = 1,\dots,m,\ B = (b_{ji}) \in M_m(D)\ .$$

A straightforward computation shows that $T \approx B$.

As $Tu^i = \lambda u^i$, $i = 1,\dots,k$, we have that $b_{11} = \lambda$ and $b_{j1} = 0$ for $j > 1$. So

$$|xI-T| = |xI-B| = (x-\lambda)|xI-\tilde{B}|,\ \tilde{B} = (b_{ij}),\ i,j = 2,\dots,m\ ,$$

where the last equality follows by expanding $|xI-B|$ by the first column. So $|xI-\tilde{B}|$ splits over D. Using the induction assumption $\tilde{B} \approx A_1$ where A_1 is an $(m-1) \times (m-1)$ upper triangular matrix with the eigenvalues of $\tilde{B}$ on the main diagonal of A_1 appearing in any prescribed order. This establishes the theorem.

□

Clearly, the upper triangular form of A is not unique unless $A = aI$. See Problem 1.18.9. In what follows we shall use the definitions given below.

Definition 1.18.9. Let $T \in M_m(D)$ and assume that (1.18.6) holds. Then the spectrum of $T - \sigma(T)$ is defined to be the set

$$\sigma(T) = \{\lambda_1,\ldots,\lambda_m\} \quad . \tag{1.18.10}$$

By m_i denote the multiplicity of λ_i, that the number of times that λ_i is appearing in $\sigma(T)$. The eigenvalue λ_i is called algebraically simple if $m_i = 1$. Let $\lambda_1,\ldots,\lambda_\ell$ be all the distinct eigenvalues of T. That is

$$\sum_{i=1}^{\ell} m_i = m, \ \lambda_i \in \sigma(T), \ i = 1,\ldots,\ell \quad . \tag{1.18.11}$$

By $\sigma_d(T)$ we denote the distinct spectrum of T

$$\sigma_d(T) = \{\lambda_1,\ldots,\lambda_\ell\} \quad . \tag{1.18.12}$$

Problems

(1.18.13). Let Q correspond to the elementary row operation described in Definition 1.10.11 - (iii). Assume that A is an upper triangular matrix. Show that if $j < i$ then QAQ^{-1} is also an upper triangular matrix.

(1.18.14) Prove that if $T \in M_m(D)$ is similar to an upper triangular matrix $A \in M_m(D)$ then the characteristic polynomial of T splits to linear factors.

(1.18.15) Let $T \in M_m(D)$ and put

$$|xI-T| = x^m + \sum_{j=1}^{m} a_{m-j}x^j \quad . \tag{1.18.16}$$

Assume that the assumptions of Theorem 1.18.5 hold. Show that

$$(1.18.17) \qquad (-1)^k a_k = \sum_{\alpha \in Q_{k,m}} T[\alpha|\alpha] = s_k(\lambda_1,\ldots,\lambda_m) \quad ,$$

where $s_k(x_1,\ldots,x_m)$ is the k-th symmetric polynomial of $x_1,\ldots,x_m$. The coefficient $-a_1$ is called the trace of $A(tr(A))$. That is

$$(1.18.18) \qquad tr(A) = \sum_{i=1}^{m} a_{ii} = \sum_{i=1}^{m} \lambda_i \quad .$$

(1.18.19) Let $T \in M_m(D)$ and suppose that the assumptions of Theorem 1.18.5 hold. Assume that D is **UFD**. Using the results of Theorem 1.18.5 and Problem 1.17.23 prove that the minimal polynomial $\psi(x)$ of T is of the form

$$(1.18.20) \qquad \psi(x) = \prod_{i=1}^{\ell} (x-\lambda_i)^{s_i}, \ \lambda_i \neq \lambda_j \ \text{ for } \ i \neq j, \ 1 \leq s_i \leq m_i, \ i,j = 1,\ldots,\ell$$

where $\sigma_d(T) = \{\lambda_1,\ldots,\lambda_\ell\}$. (<u>Hint</u>: Consider the diagonal elements of $\psi(A)$.)

(1.18.21) Let $T \in M_m(\mathbf{UFD})$ and assume that the minimal polynomial of T is given by (1.18.20). Using Problem 1.17.21 and the equality (1.17.3) prove

$$(1.18.22) \qquad |xI-T| = \prod_{i=1}^{\ell} (x-\lambda_i)^{m_i}, \ \lambda_i \neq \lambda_j \ \text{ for } \ i \neq j, \ i,j = 1,\ldots,\ell \quad .$$

1.19 Jordan canonical form

Theorem 1.18.5 and Problem 1.18.14 shows that $T \in M_m(\mathbf{D})$ is similar to an upper triangular matrix A if and only if the characteristic polynomial of T splits to linear factors. Unfortunately the upper triangular form of T is not unique. In case that $\mathbf{D}$ is a field there exists an upper triangular matrix A which depends only on the eigenvvalues of T and this matrix is essentially unique. For convenience we state the theorem of an operator $T \in L(\mathbf{V})$.

Theorem 1.19.1. *Let* $T \in L(\mathbf{V})$. *Assume that the minimal polynomial* $\psi(x)$ *of* T *splits to a product of linear factors as given in* (1.18.20). *Then* $\mathbf{V}$ *splits to a direct sum of non-trivial irreducible invariant subspaces of* T

$$\mathbf{V} = \mathbf{W}_1 \oplus \cdots \oplus \mathbf{W}_q \quad . \tag{1.19.2}$$

In each subspace $\mathbf{W}$ *it is possible to choose a basis consisting of generalized eigenvectors* $x^1,\ldots,x^r$ *such that*

$$Tx^1 = \lambda_0 x^1 \quad , \tag{1.19.3}$$

$$Tx^{k+1} = \lambda_0 x^{k+1} + x^k, \; k = 1,\ldots,r-1 \tag{1.19.4}$$

where λ_0 *is equal to* λ_i *for some* i *and* $r \leq s_i$ (*in case that* $r = 1$ (1.19.4) *is void*). *Moreover for each* λ_i, $i = 1,\ldots,\ell$, *there exists an invariant subspace of* $\mathbf{W}$ *whose basis satisfies* (1.19.3) - (1.19.4) *with* $\lambda_0 = \lambda_i$ *and* $r = s_i$.

Proof. Assume first that the minimal polynomial of T is of the form

$$\psi(x) = x^s \quad . \tag{1.19.5}$$

So $T^s = 0$ and $T^{s-1} \neq 0$. Let $x^{11},\ldots,x^{1n_1}$ span the range of T^{s-1}

$$T^{s-1}V = [x^{11},\dots,x^{1n_1}] \quad .$$

In particular $x^{11},\dots,x^{1n_1}$ are linearly independent. Let $x^{s1},\dots,x^{sn_1}$ be the pre-images of $x^{11},\dots,x^{1n_1}$ for the map $T^{s-1}: V \to V$. So

$$T^{s-1}x^{sj} = x^{1j},\ j = 1,\dots,n_1 \quad .$$

Denote

$$x^{(s-k)j} = T^k x^{sj},\ k = 1,\dots,s-2,\ j = 1,\dots,n_1 \quad .$$

As $T^s = 0$ we have $Tx^{1j} = 0$, $j = 1,\dots,n_1$. The two equalities above are equivalent to

(1.19.6) $$Tx^{1j} = 0,\ Tx^{(k+1)j} = x^{kj},\ k = 1,\dots,s-1,\ j = 1,\dots,n_1 \quad .$$

We claim that $x^{11},\dots,x^{1n_1},\dots,x^{s1},\dots,x^{sn_1}$ are linearly independent. Indeed, suppose that

$$\sum_{i=1}^{s} \sum_{j=1}^{n_1} a_{ij}x^{ij} = 0 \quad .$$

Apply T^{s-1} to this equality. The (1.19.6) yields

$$\sum_{j=1}^{n_1} a_{sj}x^{1j} = 0 \quad .$$

As $x^{11},\dots,x^{1n_1}$ are linearly independent $a_{s1} = \dots = a_{sn_1} = 0$. Next apply T^{s-2} to obtain $a_{(s-1)1} = \dots = a_{(s-1)n_1} = 0$. Continuing in the same manner we deduce that

x^{ij}, $i = 1,\ldots,s$, $j = 1,\ldots,n_1$ are linearly independent. Assume now that we have found linear independent vectors

$$x^{ij},\ i = 1,\ldots,s-r+1,\ j = m_{r-1}+1,\ldots,m_r,\ r = 1,\ldots,p \ ,$$

(1.19.7)

$$m_0 = 0,\ m_r = n_1+\ldots+n_r,\ r = 1,\ldots,p,\ n_1 \geq 1,\ n_2 \geq 0,\ldots,n_p \geq 0 \ ,$$

for $1 \leq p \leq s$, such that (1.19.6) holds for $k = 1,\ldots,s-r$, $j = m_{r-1}+1,\ldots,m_r$, $r = 1,\ldots,p$ and

$$(1.19.8) \qquad T^{s-p} = \mathbf{V} = [x^{11},\ldots,x^{1m_1},\ldots,x^{p1},\ldots,x^{pm_1},\ x^{1(m_1+1)},\ldots,x^{1m_2},\ldots,$$

$$x^{(p-1)(m_1+1)},\ldots,x^{(p-1)m_2},\ldots,x^{1(m_{p-1}+1)},\ldots,x^{1m_p}] \ .$$

For $p = s$ we found the needed basis of $\mathbf{V}$. For $p < s$ consider the subspace $\mathbf{V}_{s-p-1} = T^{s-p-1}\mathbf{V}$. Clearly $x^{11},\ldots,x^{1m_1},\ldots,x^{(p+1)1},\ldots,x^{(p+1)m_1},\ldots,x^{1(m_{p-1}+1)},\ldots,x^{1m_p}$, $x^{2(m_{p-1}+1)},\ldots,x^{2m_p}$ belong to $\mathbf{V}_{s-p-1}$. Recall that these vectors were assumed to be linearly independent. If the above vectors span $\mathbf{V}_{s-p-1}$ put $r_{p+1} = 0$, $m_{p+1} = m_p$. Otherwise, let

$$\mathbf{V}_{s-p-1} = [x^{11},\ldots,x^{1m_1},\ldots,x^{(p+1)1},\ldots,x^{(p+1)m_1},\ldots,x^{1(m_{p-1}+1)},\ldots,x^{1m_p},$$

$$x^{2(m_{p-1}+1)},\ldots,x^{2m_p},\ x^{1(m_p+1)},\ldots,x^{1m_{p+1}}] \ .$$

Since $Tx^{1j} \in \mathbf{V}_{s-p}$, $j = m_p + 1,\ldots,m_{p+1}$ in view of (1.19.8) and (1.19.6) we can assume that $Tx^{1j} = 0$, $j = m_p + 1,\ldots,m_{p+1}$ (by adding to x^{1j} a linear combination of vectors appearing in (1.19.7)). Denote $x^{(s-p)j}$ the pre-images of x^{1j} for the map $T^{s-p-1}: \mathbf{V} \to \mathbf{V}$ for $j = m_p+1,\ldots,m_{p+1}$. Also let $x^{(s-p-k)j} = T^k x^{(s-p)j}$, $k = 1,\ldots,s-p-2$. Thus (1.19.6) holds for $k = 1,\ldots,s-r$, $j = m_{r-1}+1,\ldots,m_r$, $r = 1,\ldots,p+1$. We claim that the vectors appearing in (1.19.7) for $r = 1,\ldots,p+1$ linearly independent. This follows by applying $T^{s-p-1},\ldots,T^1$, $T^0 = I$ to the given vectors and using the identities (1.19.6) for all the involved vectors and taking in account that the vectors which span V_{s-p-1} are linearly independent. For each p such that $m_p > m_{p-1}$ let

$$W_j = [x^{1j}, x^{2j},\ldots,x^{(s-p+1)j}], \quad m_{p-1} + 1 \leq j \leq m_p \quad .$$

So (1.9.3) - (1.9.4) holds for $\lambda_0 = 0$ and $r = s-p+1$. As all x^{ij}, $i = 1,\ldots,s-r+1$, $j = m_{r-1} + 1,\ldots,m_r$, $r = 1,\ldots,s$, the equality (1.19.2) holds. Also $m_1 = n_1 > 0$ and $\dim W_1 = s$. It is left to show that W_j is an irreducible invariant subspace of T. From the equalities (1.19.6) we get that if $x^{(s-p+1)j} \in U$, where U is an invariant subspace of W_j then $x^{kj} \in U$ for $k = s-p+1,\ldots,1$. So $U = W_j$ and W_j is irreducible. This proves the theorem in case that the minimal polynomial of T is of the form (1.19.5). Assume that the minimal polynomial of T is of the form (1.18.20). According to Theorem 1.17.11 V splits to the direct sum of non-trivial invariant subspaces (1.17.4) such that the minimal polynomial of T_j - the restriction of T to U_j is $(x-\lambda_j)^{s_j}$. Call

$$T_j - \lambda_j I = Q_j : U_j \to U_j, \quad j = 1,\ldots,\ell \quad .$$

Clearly, the minimal polynomial of Q_j is x^{s_j}. So we can apply our results for Q_j, $j = 1,\ldots,\ell$. This immediately implies the theorem.

□

Let $H(n)$ be the following $n \times n$ matrix

(1.19.9) $$H(n) = (h_{ij})_1^n, \quad h_{ij} = \delta_{(i+1)j}, \quad i,j = 1,\ldots,n \quad .$$

That is $H(n)$ is 0-1 matrix of the form

$$H(n) = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix} .$$

We shall also denote $H(n)$ by H_n or simply H in case that the dimension of H is well defined. Let $V = [v^1,\ldots,v^r]$ when $v^i = x^{r-i+1}$, $i = 1,\ldots,r$ and $x^1,\ldots,x^r$ satisfy (1.19.3) - (1.19.4). Then in this basis T is represented by a <u>Jordan block</u> $\lambda_0 I + H$ of dimension r. Thus Theorem 1.19.1 implies:

<u>Theorem</u> 1.19.10. <u>Let</u> $A \in M_n(F)$. <u>Assume that the minimal polynomial</u> $\psi(x)$ <u>of</u> A <u>splits to linear factors as given in</u> (1.18.20). <u>The there exists a non-singular matrix</u> $P \in M_n(F)$ <u>such that</u>

$$(1.19.11) \quad \begin{aligned} &P^{-1}AP = J,\ J = \sum_{i=1}^{\ell} \sum_{j=1}^{q_i} \oplus (\lambda_i I(m_{ij}) + H(m_{ij})) \ , \\ &1 \le m_{iq_i} \le m_{iq_i-1} \le \ldots \le m_{i1} = s_i,\ i = 1,\ldots,\ell,\ \lambda_i \ne \lambda_j,\ \text{for } i \ne j \ . \end{aligned}$$

<u>Definition</u> 1.19.12. <u>The matrix</u> J <u>appearing in</u> (1.19.11) <u>is called the Jordan canonical form of</u> $A(T)$. <u>The polynomials</u>

$$(1.19.12) \qquad \varphi_{ij}(x) = (x-\lambda_i)^{m_{ij}},\ j = 1,\ldots,q_i,\ i = 1,\ldots,\ell \ ,$$

<u>are called the elementary divisors of</u> $A(T)$.

<u>Remark</u> 1.19.13. In case that the minimal polynomial of A does not split to linear factors in F we can find a finite extension field K such that ψ splits in K. Then (1.19.11) holds for $P \in M_n(K)$. We shall refer to J as the Jordan canonical form of A.

<u>Theorem</u> 1.19.14. <u>Let</u> $A \in M_n(F)$. <u>Assume that the minimal polynomial of</u> A <u>is of the form given by</u> (1.18.20). <u>The the elementary polynomials of</u> A <u>are the elementary divisors of</u> $xI-A$ <u>defined in Problem</u> 1.16.13. <u>That is, put</u>

(1.19.15) $$m_{iq_i+1} = \dots = m_{in} = 0, \; i = 1,\dots,\ell \; .$$

Then the invariant polynomials $i_1(x),\dots,i_n(x)$ of $xI-A$ are given by the equalities

(1.19.16) $$i_r(x) = \prod_{i=1}^{\ell} (x-\lambda_i)^{m_{i(n-r+1)}}, \; r = 1,\dots,n \; .$$

In particular if $p_1(x),\dots,p_k(x)$ are the non-trivial invariant polynomials of $xI-A$ then

(1.19.17) $$p_{k-j+1}(x) = \prod_{i=1}^{\ell} (x-\lambda_i)^{m_{ij}}, \; j = 1,\dots,k \; .$$

Proof. *Assume first that* $A = \lambda_0 I(m) + H(m)$. *Then for* $1 \le k \le m-1$ the minor of $xI-A$ composed of the rows $1,\dots,k$ and the columns $2,\dots,k+1$ is equal to $(-1)^k$. So the first $m-1$ determinant invariants are trivial. Also $\delta_m(xI-A) = |xI-A| = (x-\lambda_0)^m$. Hence

$$xI-A \sim xI-J \sim \sum_{i=1}^{\ell} \sum_{j=1}^{q_i} \oplus \, \mathrm{diag}(1,\dots,1, (x-\lambda_i)^{m_{ij}}) \; .$$

Applying the results of Problems (1.12.9) - (1.12.10) we deduce the equality (1.19.16). Clearly (1.19.16) is equivalent to (1.19.17).

□

This theorem shows that the Jordan canonical form of $A(T)$ is unique up to permutation of Jordan blocks.

Problems

(1.19.18) Show directly that to each eigenvalue λ_o of a companion matrix $C(p)$ corresponds one linear independent vector of the form $(1,\lambda_o,\dots,\lambda_o^{n-1})^t$.

(1.19.19) Let $A \in M_n(\mathbb{C})$. Assume that $\lambda \in \sigma(A)$. Let $U_1, U_2 \subseteq \mathbb{C}^n$ be the subspaces of all eigenvectors of A and A^t respectively corresponding to λ. Show that there exists bases $x^1, \ldots, x^m$ in U_1 and $y^1, \ldots, y^m$ in U_2 such that $(y^i)^t x^j = \delta_{ij}$, $i,j = 1, \ldots, m$.

(Hint Assume first that A is in the Jordan canonical form.)

(1.19.20) Let

$$Ax = \lambda x, \; A^t y = \mu y, \; A \in M_n(\mathbb{C}), \; 0 \neq x, y \in \mathbb{C}^n \; .$$

Show that if $\lambda \neq \mu$ then $y^t x = 0$.

(1.19.21) Verify directly that J annihilates its characteristic polynomial. Using the fact that any $A \in M_n(\mathbb{F})$ is similar to its Jordan canonical form over a finite extension field $\mathbb{K}$ deduce the Cayley-Hamilton theorem.

(1.19.22) Let $A, B \in M_n(\mathbb{F})$ show that $A \approx B$ if and only if A and B have the same Jordan canonical form.

1.20 Some applications of Jordan canonical form.

Definition 1.20.1. Let $A \in M_n(\mathbb{F})$ and assume that $|xI-A|$ splits in $\mathbb{F}$. Let λ_0 be an eigenvalue of A. Then the number of factors of the form $(x-\lambda_0)$ appearing in the minimal polynomial $\psi(x)$ of A is called the index of λ_0 and is denoted by index (λ_0). The number of linearly independent eigenvectors of A corresponding to λ_0 is called the geometric multiplicity of λ_0.

Using the results of the previous section we get

Lemma 1.20.2. Let the assumptions of Definition 1.20.1 hold. Then index (λ_0) is the size of the largest Jordan block corresponding to λ_0 (i.e. of the form $\lambda_0 I+H$) and the geometric multiplicity of λ_0 is the number of the Jordan blocks corresponding to λ_0 which appear in the Jordan canonical form of A.

Let $T \in L(V)$, $\lambda_0 \in \sigma(T)$ and consider invariant subspaces

$$X_r = \{x \mid x \in V,\ (\lambda_0 I - T)^r x = 0\},\quad r = 0,1,\dots \quad . \tag{1.20.3}$$

Using the decomposition (1.19.2) and the definition of Jordan form of T we obtain

Theorem 1.20.4. Let $T \in L(V)$ and assume that λ_0 is an eigenvalue of T. Let index $(\lambda_0) = m_1 \geq m_2 \geq \dots \geq m_\sigma \geq 1$ be the dimensions of all Jordan blocks corresponding to λ_0 which appear in Jordan canonical form of T. Then

$$\dim X_r = \sum_{i=1}^{\sigma} \min(r, m_i),\quad r = 0,1,\dots \quad . \tag{1.20.5}$$

In particular

$$\{0\} = X_0 \subsetneqq X_1 \subsetneqq X_2 \subsetneqq \dots \subsetneqq X_{m_1} = X_{m_1+1} = \dots,\ m_1 = \text{index}\ (\lambda_0) \quad . \tag{1.20.6}$$

Thus (1.20.6) gives yet another characterization of the index of λ_0. Also note in view of Definition 1.18.1 each X_k consists of generalized eigenvectors of T.

Definition 1.20.7. *An operator* $T \in L(V)$ *is said to have a simple structure if there exists a basis in* V *which consists entirely of eigenvectors of* T. *That is any representation matrix* A *of* T *is diagonable (i.e., similar to a diagonal matrix).*

For such T we must have $X_1 = X_{m_1}$. Theorem 1.19.1 yields.

Theorem 1.20.8. *Let* $T \in L(V)$. *Then* T *has a simple structure if and only if the minimal polynomial* ψ *of* T *splits to linear factors such that any two factors in* ψ *are relatively prime. That is the index of any eigenvalue of* T *equals to* 1.

Definition 1.20.9. *Let* $T \in \mathrm{Hom}(M,M)$ *where* M *is a* D-*module.* T *is called nilpotent if* $T^s = 0$ *for some positive integer* s.

We need in the sequel the following result.

Theorem 1.20.10. *Let* $T \in L(V)$ *be nilpotent. Assume that* U *is a non-trivial invariant subspace of* T *such that*

$$TV \subseteq U \quad . \tag{1.20.11}$$

Then it is possible to split V *to the direct sum of invariant subspaces* (1.19.2) *and to choose in each invariant subspace* W_i *basis* $y^{i1},\ldots,y^{ir_i}$ *satisfying* (1.19.3) - (1.19.4) *(with* $\lambda_0 = 0$*) such that the vectors* $y^{i1},\ldots,y^{ir_i'}$, $r_i - 1 \leq r_i' \leq r_i$, $i = 1,\ldots,q$, *form a basis in* U.

Proof. The proof of the theorem is a modification of the proof of Theorem 1.19.1 and we point out the changes one needs to make. Let x^{ij}, $i = 1,\ldots,s-r+1$, $j = m_{r-1}+1,\ldots,m_r$, $r = 1,\ldots,s$ be a basis of V satisfying (1.19.6). Then the condition (1.20.11) implies $x^{ij} \in U$, $i = 1,\ldots,s-r$, $j = m_{r-1}+1,\ldots,m_r$, $r = 1,\ldots,s-1$. We choose now the vectors x^{ij} in the following way. The vectors $x^{11},\ldots,x^{1n_1}$ are picked up to satisfy in addition

$$T^{s-1}U = [x^{11},\ldots,x^{1n_1'}],\ 0 \leq n_1' \leq n_1,\ (n_1' = 0 \text{ if } T^{s-1}U = [0]) \quad .$$

We claim that the vectors x^{ij} can be chosen in each stage to satisfy also

$$\begin{aligned}
T^{s-p}U = [&x^{11},\dots,x^{1m_1},\dots, x^{(p-1)1},\dots,x^{(p-1)m_1}, x^{p1},\dots,x^{pn'_1},\dots, \\
&x^{1(m_{p-2}+1)},\dots,x^{1m_{p-1}}, x^{2(m_{p-2}+1)},\dots,x^{2(m_{p-2}+n'_{p-1})},\dots, \\
&x^{1(m_{p-1}+n'_p)}] ,
\end{aligned} \tag{1.20.12}$$

where $0 \leq n'_p \leq n_p$. (If $n'_p = 0$ then $x^{1(m_{p-1}+1)} \notin T^{s-p}U$). Suppose we already proved the claim for $q = p-1$. Let

$$\begin{aligned}
T^{s-p}U = [&x^{11},\dots,x^{1m_1},\dots,x^{(p-1)1},\dots,x^{(p-1)m_1}, x^{p1},\dots,x^{pn'_1},\dots, \\
&x^{1(m_{p-2}+1)},\dots,z^{1m_{p-1}}, x^{2(m_{p-2}+1)},\dots,x^{2(m_{p-2}+n'_{p-1})}, y^1,\dots,y^t] .
\end{aligned}$$

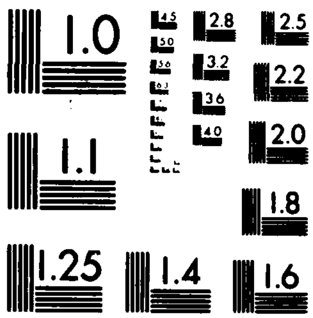
As in the proof of Theorem 1.19.1 we may assume that $Ty^j = 0$, $j = 1,\dots,t$. To finish the proof we have to show that the vectors x^{ij}, $i = 1,\dots,p-r+1$, $j = m_{r-1}+1,\dots,m_r$, $r = 1,\dots,p-1$, $y^1,\dots,y^t$ are linearly independent. Suppose that these vectors are linearly dependent. Applying T to the vector in question and using the fact that x^{ij} are linearly independent we deduce that $x^{11},\dots,x^{1m_1},x^{1(m_1+1)},\dots,x^{1m_2},\dots,x^{1m_{p-1}}$, $y^1,\dots,y^t$ have to be linearly independent. This is impossible since these vectors are part of the basis vectors for $T^{s-p}U$. So $t \leq n_p$ and we can choose vectors $x^{1(m_{p-1}+1},\dots,x^{1(m_{p-1}+n_p)}$ such that the first n'_p vectors coincide with $y^1,\dots,y^t$. The equality (1.20.12) establishes the theorem.

□

AD-A089 635   WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER   F/G 12/1
SPECTRAL THEORY OF MATRICES. I. GENERAL MATRICES.(U)
MAY 80   S FRIEDLAND   DAAG29-75-C-0024
UNCLASSIFIED   MRC-TSR-2082   NL

2 of 2

END
DATE
FILMED
10 80
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

1.21 <u>The equation</u> $AX - XB = 0$.

Let $A,B \in M_n(D)$. A possible way to determine whether $A \approx B$ over $M_n(D)$ is to consider the matrix equation

$$(1.21.1) \qquad AX - XB = 0 \quad .$$

Then $A \approx B$ if and only if there exists a solution X such that $|X|$ is an invertible element. If we consider X as a column vector $\hat{X}$ composed of n columns of X then the equation (1.21.1) has a simple form in the tensor notation (e.g. Marcus-Minc [1964], see also Problem 1.21.17)

$$(1.21.2) \qquad (I \otimes A - B^t \otimes I)\hat{X} = 0 \quad .$$

Thus is D is a Bezout domain then the set of all X $M_n(D)$ satisfying (1.21.1) form a D-module with a basis $X_1,\ldots,X_\nu$, (Theorem 1.11.12). So any matrix X which satisfies (1.21.1) is of the form

$$X = \sum_{i=1}^{\nu} x_i X_i, \; x_i \in D, \; i = 1,\ldots,\nu \quad .$$

Thus it is "left" to find whether a function

$$\delta(x_1,\ldots,x_r) = |\sum_{i=1}^{\nu} x_i X_i|$$

has an invertible value. In such a generality this is a difficult problem. A more modest task is to find the value ν and to determine if $\delta(x_1,\ldots,x_\nu)$ vanish identically. For that purpose it is enough to assume that D is actually a field F (for example the quotient field of D). Also, we may replace F by a finite extension K in which the characteristic polynomials of A and B split to linear factors. Finally, we are going to study slightly more general case

$$A \in M_m(K),\ B \in M_n(K),\ X \in M_{mn}(K) \quad .$$

Let $\psi(x)$ and $\varphi(x)$ and J and K be the minimal polynomials and Jordan canonical forms of A and B respectively.

$$\psi(x) = (x-\lambda_1)^{s_1} \cdots (x-\lambda_\ell)^{s_\ell},\ \lambda_i \neq \lambda_j \quad \text{for} \quad i \neq j \quad ,$$

$$\varphi(x) = (x-\mu_1)^{t_1} \cdots (x-\mu_k)^{t_k},\ \mu_i \neq \mu_j \quad \text{for} \quad i \neq j \quad ,$$

(1.21.3)

$$P^{-1}AP = J = \sum_{i=1}^{\ell} \oplus J_i,\ J_i = \sum_{j=1}^{q_i} \oplus (\lambda_i I(m_{ij}) + H(m_{ij})),\ 1 \leq m_{iq_i} \leq \cdots \leq m_{i1} = s_i \quad ,$$

$$Q^{-1}BQ = K = \sum_{i=1}^{k} \oplus K_i,\ K_i = \sum_{j=1}^{p_i} \oplus (\mu_i I(n_{ij}) + H(n_{ij})),\ 1 \leq n_{ip_i} \leq \cdots \leq n_{i1} = t_i \quad .$$

Let

$$Y = P^{-1}XQ \quad .$$

Then the system (1.21.1) is equivalent to

$$JY - YK = 0 \quad .$$

We partition Y conformally to the partitions of J and K as given in (1.21.3). So

$$Y = (Y_{ij}),\ Y_{ij} \in M_{m_i n_j}(K),\ m_i = \sum_{r=1}^{q_i} m_{ir},\ n_j = \sum_{r=1}^{p_j} n_{jr},\ i=1,\ldots,\ell\ ,\ j=1,\ldots,\ell \quad .$$

Thus, the matrix equation for Y reduces to ℓk matrix equations

(1.21.4) $$J_i Y_{ij} - Y_{ij} K_j = 0,\ i = 1,\ldots,\ell,\ j = 1,\ldots,k \quad .$$

The following two lemms analyze the equations (1.21.4).

<u>Lemma</u> 1.21.5. <u>Consider a matrix equation 1.21.4 for some choice of</u> $1 \leq i \leq \ell$ <u>and</u> $1 \leq j \leq k$. <u>If</u> $\lambda_i \neq \mu_j$ <u>then</u> $Y_{ij} = 0$.

<u>Proof.</u> Put

$$J_i = \lambda_i I(m_i) + \bar{J}_i,\ \bar{J}_i = \sum_{r=1}^{q_i} \oplus H(m_{ir}) \quad ,$$

$$K_j = \mu_j I(n_j) + \bar{K}_j \cdot \bar{K}_j = \sum_{r=1}^{p_i} \oplus H(n_{ir}) \quad .$$

Note that $\bar{J}_i^{\,u} = \bar{K}_j^{\,v} = 0$ for $u > m_i$ and $v > n_j$. Then (1.21.4) becomes

$$(\lambda_i - \mu_j)Y_{ij} = -\bar{J}_i Y_{ij} + Y_{ij}\bar{K}_j \quad .$$

Thus

$$\begin{aligned}(\lambda_i-\mu_j)^2\, Y_{ij} &= -\bar{J}_i(\lambda_i-\mu_j)Y_{ij} + (\lambda_i-\mu_j)Y_{ij}\bar{K}_j = \\ &= -\bar{J}_i(-\bar{J}_i Y_{ij} + Y_{ij}\bar{K}_j) + (-\bar{J}_i Y_{ij} + Y_{ij}\bar{K}_j)\bar{K}_u = \\ &= (-\bar{J}_i)^2 Y_{ij} + 2(-\bar{J}_i)Y_{ij}\bar{K}_j + Y_{ij}\bar{K}_j^{\,2} \quad .\end{aligned}$$

Continuing this procedure we get

$$(\lambda_i-\mu_j)^r Y_{ij} = \sum_{u=0}^{r} \binom{r}{u}(-\bar{J}_i)^u Y_{ij}\bar{K}_j^{\,r-u} \quad .$$

Whence, for $r = m_i+n_j$ either $\bar{J}_i^{\,u}$ or $\bar{K}_j^{\,r-u}$ is a zero matrix. Since $\lambda_i \neq \mu_j$ we deduce that $Y_{ij} = 0$.

□

Lemma 1.21.6. *Let* $Z = (z_{\alpha\beta}) \in M_{mn}(\mathbb{F})$ *satisfy the equation*

(1.21.7) $$H(m)Z = Z\,H(n) \quad .$$

Then the entries of Z *are of the form*

(1.21.8) $$\begin{aligned} z_{\alpha\beta} &= 0 && \text{for } \beta < \alpha+n-\min(m,n) \quad , \\ z_{\alpha\beta} &= z(\alpha+1)(\beta+1) && \text{for } \beta \geq \alpha+n-\min(m,n) \quad . \end{aligned}$$

In particular the subspace of all $m \times n$ *matrices* Z *satisfying* (1.21.7) *has the dimension* $\min(m,n)$.

Proof. As the last row of $H(m)$ and the first column of $H(n)$ are equal to zero row and column respectively, in view of the equality (1.21.7) the last row of $ZH(n)$ and the first column of $H(m)Z$ vanish. That is $z_{m\beta} = z_{\alpha 1} = 0$, $\beta = 1,\ldots,n-1$, $\alpha = 2,\ldots,m$. In all other cases by equating the (α,β) entries of $H(m)Z$ and $ZH(n)$ we get $z_{(\alpha+1)\beta} = z_{\alpha(\beta-1)}$, $\alpha = 1,\ldots,m-1$, $\beta = 2,\ldots,n$. The above set of equalities imply the condition (1.21.8).

Combine the two lemmas to obtain

Theorem 1.21.9. Consider the system of matrix equations (1.21.4). Then $Y_{ij} = 0$ if $\lambda_i \neq \mu_j$ (i.e., J_i and K_j do not have a common eigenvalue). Assume that $\lambda_i = \mu_j$. Partition Y_{ij} conformally with the partition of J_i and K_j as given in (1.21.3).

$$Y_{ij} = (Y_{ij}^{(uv)}),\ Y_{ij}^{(uv)} \in M_{m_{iu}n_{jv}}(K),\ u = 1,\ldots,\sigma_i,\ v = 1,\ldots,\rho_j \quad .$$

Then each $Y_{ij}^{(uv)}$ is of the form prescribed in Lemma 1.21.6 with $m = m_{iu}$ and $n = n_{jv}$. Assume that

$$(1.21.10) \qquad \lambda_1 = \mu_1,\ldots,\lambda_t = \mu_t,\ \lambda_i \neq \mu_j,\ i = t+1,\ldots,\ell,\ j = t+1,\ldots,k \quad .$$

Then the dimension of the linear subspace $\mathcal{V} \subset M_{mn}(K)$ of matrices $Y = (Y_{ij})$, $i = 1,\ldots,\ell$, $j = 1,\ldots,k$ satisfying (1.21.4) is given by the formula

$$(1.21.11) \qquad \dim \mathcal{V} = \sum_{i=1}^{t} \sum_{u,v=1}^{\sigma_i,\rho_i} \min(m_{iu},n_{iv}) \quad .$$

Let us consider a special case of (1.21.1)

$$(1.21.12) \qquad AX - XA = 0 \quad , \quad A \in M_n(D) \quad .$$

In that case, a D-module of all matrices $X \in M_n(D)$ satisfying (1.21.12) is in fact a ring (non-commutative in general) with an identity I. Denote

(1.21.13) $$C(A) = \{X \mid X \in M_n(D),\ AX = XA\} \ .$$

In case that D is a field F or more generally when $C(A)$ has a finite basis (according to Theorem 1.11.12 this assumption holds if D is a Bezout domain) then according to Theorem 1.21.9

$$\dim C(A) = \sum_{i=1}^{\ell} \sum_{u,v=1}^{q_i} \min(m_{iu}, m_{iv}) \ .$$

(Clearly the dimension of $C(A)$ is not changed if we let $X \in M_n(K)$). As $\{m_{iu}\}_1^{q_i}$ is a decreasing sequence we have

$$\sum_{v=1}^{q_i} \min(m_{iu}, m_{iv}) = u m_{iu} + \sum_{v=u+1}^{q_i} m_{iv} \ .$$

So

(1.21.14) $$\dim C(A) = \sum_{i=1}^{\ell} \sum_{u=1}^{q_i} (2u-1) m_{iu} \ .$$

Let $i_1(x), \ldots, i_n(x)$ be the invariant polynomials of $xI-A$. Use (1.19.15)-(1.19.16) to deduce

(1.21.15) $$\dim C(A) = \sum_{u=1}^{n} (2u-1) \deg i_{(n-u+1)}(x) \ .$$

With the help of the above formula we can determine when any commuting matrix with A is a polynomial in A. Clearly, the dimension of the subsapce spanned by the powers of A is equal to the degree of the minimal polynomial of A.

Corollary 1.21.16. *Let* $A \in M_n(F)$. *Then each commuting matrix with* A *can be expressed as a polynomial in* A *if and only if the minimal and the characterisitc polynomials of*

A are identical. That is, A is similar to a companion matrix $C(p)$, where $p(x) = |xI-A|$.

A matrix for which the minimal and the characteristic polynomial coincide is called nonderogatory, otherwise derogatory.

Problems

(1.21.17) Let $\mu : M_{mn}(D) \to M_{(mn)1}(D)$, such that

$$\mu(X) = \hat{X} = (\hat{x}_j),\ j = 1,\ldots,mn,\ \hat{x}_j = x_{ik} \quad \text{for } j = (k-1)m + i \ ,$$
$$i = 1,\ldots,m,\ k = 1,\ldots,n \ .$$

Prove

$$(1.21.18)\ \mu(AX) = (I(n) \otimes A)\mu(X),\ \mu(XB) = (B^t \otimes I(m)\mu(X),\ A \in M_m(D),\ B \in M_n(D) \ .$$

Here $A \otimes B$ is the Kronecker product

$$(1.21.19)\ A \otimes B = (a_{ij}B) \in M_{(mp)(nq)}(D),\ A = (a_{ij}) \in M_{mn}(D),\ B = (b_{k\ell}) \in M_{pq}(D) \ .$$

Prove

$$(1.21.20)\quad (A_1 \otimes A_2)(B_1 \otimes B_2) = (A_1B_1) \otimes (A_2B_2),\ A_i \in M_{m_i n_i}(D),\ B_i \in M_{n_i p_i}(D),$$
$$i = 1,2 \ .$$

(1.21.21) Let $P \in M_m(F)$, $Q \in M_n(F)$, $R \in M_{mn}(F)$. Put

$$A = \begin{pmatrix} P & R \\ 0 & Q \end{pmatrix},\quad B = \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix} \in M_{m+n}(F) \ .$$

Assume that the characteristic polynomials of P and Q are coprime. Show that there exists a matrix $X = \begin{pmatrix} I(m) & Y \\ 0 & I(n) \end{pmatrix}$ which satisfies (1.21.1). This in particular implies A B.

(1.21.22) Let $A = \sum_{i=1}^{\ell} \oplus A_i \in M_n(\mathbb{F})$. Prove that

(1.21.23) $$\dim C(A) \geq \sum_{i=1}^{\ell} \dim C(A_i)$$

and the equality sign holds if and only if

$$(|xI-A_i|, |xI-A_j|) = 1 \quad \text{for} \quad i \neq j, \; i,j = 1,\ldots,\ell \; .$$

(1.21.24) Let $A \in M_n(D)$. Show that the ring $C(A)$ is a commutative ring if and only if A satisfies the conditions of Corollary 1.21.16 where $\mathbb{F}$ is the quotient field of D.

(1.21.25) Let $A \in M_n(D)$. Let $B \in C(A)$. Then B is an invertible element in the ring $C(A)$ if and only if B is a unimodular matrix.

(1.21.26) Let

(1.21.27) $$C(A,B) = \{X | X \in M_{mn}(D), \; AX-XB = 0\} \; A \in M_m(D), \; B \in M_n(D) \; .$$

Show that $C(A,B)$ is a left (right) module of $C(A)$ $(C(B))$ under the matrix multiplications.

(1.21.28) Let $A,B \in M_n(D)$. Prove that $A \approx B$ if and only if (i) $C(A,B)$ is a $C(A)$-module with a basis containing one element U; (ii) any element basis U is a unimodular matrix.

1.22 A criterion for similarity of two matrices.

Definition 1.22.1. *Let* $A \in M_m(D)$ *and* $B \in M_n(D)$. *Denote by* $r(A,B)$ *and* $\nu(A,B)$ *the rank and the nullify of the matrix* $I(n) \otimes A - B^t \otimes I(m)$ *viewed as a matrix over* $M_{mn}(F)$ *where* F *is the quotient field of* D.

According to Theorem 1.21.9 we have

$$\begin{aligned} \nu(A,B) &= \sum_{i=1}^{t} \sum_{u,v=1}^{q_i,p_i} \min(m_{iu},n_{iv}) \ , \\ r(A,B) &= mn - \sum_{i=1}^{t} \sum_{u,v=1}^{q_i,p_i} \min(m_{iu},n_{iv}) \ . \end{aligned} \tag{1.22.2}$$

Theorem 1.22.3. *Let* $A \in M_m(D)$, $B \in M_n(D)$. *Then*

$$\nu(A,B) \le \frac{1}{2}[\nu(A,A) + \nu(B,B)] \ .$$

The equality sign holds if and only if $m = n$ *and* A *and* B *are similar over the quotient field* F.

Proof. Without loss of generality we may assume that $D = F$ and the characteristic polynomials of A and B splits in $F[x]$. For $x,y \in R$ consider the function $\min(x,y)$ (the minimum of the values x and y). Clearly, $\min(x,y)$ is a homogeneous concave function on R^2 $\min(ax,ay) = a\min(x,y)$, $a > 0$, $\min(\frac{x+u}{2}, \frac{y+v}{2}) \ge \frac{1}{2}[\min(x,y) + \min(u,v)]$. So

$$\min(a+b,c+d) \ge \frac{1}{2}[\min(a,c) + \min(b,d) + \min(a,d) + \min(b,c)] \ . \tag{1.22.4}$$

Moreover, a straightforward calculation shows that if $a = c$ and $b = d$ the equality sign holds if and only if $a = b$. Let

$$N = \max(m,n), \ m_{iu} = n_{jv} = 0, \ \text{ for } q_i < u \le N, \ p_i < v \le N, \ i = 1,\dots,\ell, \ j = 1,\dots,k \ .$$

Then

$$\nu(A,A) + \nu(B,B) = \sum_{i,u=1}^{\ell,N} (2u-1)m_{iu} + \sum_{j,u=1}^{k,N} (2u-1)n_{ju} \geq \sum_{i,u=1}^{t,N} (2u-1)(m_{iu} + n_{iu}) ,$$

and the equality sign holds if and only if $\ell = k = t$. Next consider the inequality

$$\sum_{i,u=1}^{t,N} (2u-1)(m_{iu} + n_{iu}) = \sum_{i=1}^{t} \sum_{u,v=1}^{N} \min(m_{iu} + n_{iu}, m_{iv} + n_{iv}) \geq$$

$$\frac{1}{2} \sum_{i=1}^{t} \sum_{u,v=1}^{N} \min(m_{iu},m_{iv}) + \min(n_{iu},n_{iv}) + \min(m_{iu},n_{iv}) + \min(n_{iu},m_{iv})$$

$$= \frac{1}{2} \sum_{i,u=1}^{t,N} (2u-1)(m_{iu} + n_{iu}) + \sum_{i=1}^{t} \sum_{u,v=1}^{q_i,p_i} \min(m_{iu},n_{iv}) .$$

By looking at the terms where $u = v$ from the equality case in (1.22.4) we deduce that the equality sign in the above inequality holds if and only if $m_{iu} = n_{iu}$, $u = 1,\dots,N$, $i = 1,\dots,t$. The above inequality is equivalent to

$$\frac{1}{2} \sum_{i,u=1}^{t,N} (2u-1)(m_{iu} + n_{iu}) \geq \sum_{i=1}^{t} \sum_{u,v=1}^{q_i,p_i} \min(m_{iu},n_{iv}) .$$

Combining all these results we obtain the inequality (1.22.3). The equality sign in (1.22.3) holds if and only if A and B have the same Jordan canonical forms. That is $m = n$ and A is similar to B over $\mathbb{F}$. □

Suppose that $A \approx B$. That is (1.15.2) holds. Then the rules for the tensor products (Problem 1.21.20) imply

$$I \otimes A - B^t \otimes I = [(Q^t)^{-1} \otimes I](I \otimes A - A^t \otimes I)[Q^t \otimes I] ,$$

(1.22.5)

$$I \otimes B - B^t \otimes I = [(Q^t \otimes Q](I \otimes A - A^t \otimes I)[Q^t \otimes Q^{-1}) .$$

That it is the three matrices

$$(1.22.6) \qquad I \otimes A - A^t \otimes I,\ I \otimes A - B^t \otimes I,\ I \otimes B - B^t \otimes I$$

are similar. In particular these matrices are equivalent. Over a field F the above matrices are equivalent if and only if they have the same nullity. Hence Theorem 1.22.3 yields.

Theorem 1.22.7. *Let $A,B \in M_n(F)$. Then A and B are similar if and only if the three matrices in (1.22.6) are equivalent.*

The obvious part of Theorem 1.22.7 extends trivially to any integral domain D.

Lemma 1.22.8. *Let $A,B \in M_n(D)$. If A and B are similar over D then the three matrices in (1.22.6) are equivalent over D.*

However, this condition is not sufficient for the similarity of A and B even in case $D = Z$. (See Problem 1.22.16.) The disadvantage of the similarity criterion stated in Theorem 1.22.7 is due to the appearance of the matrix $I \otimes A - B^t \otimes I$ which depends on A and B. It is interesting to note that the equivalence of just two matrices in (1.22.6) does not imply the similarity of A and B. Indeed $I \otimes A - A^t \otimes I = I \otimes (A + \lambda I) - (A + \lambda I)^t \otimes I$ for any λ, but A is not similar to $A + \lambda I$ for $\lambda \neq 0$. (Problem 1.22.17.) Also if $A = H(n)$ and $B = 0$ then $\nu(A,A) = \nu(A,B) = n$ (Problem 1.22.18). However, under certain assumptions the equality $\nu(A,A) = \nu(A,B)$ implies that $A \approx B$.

Theorem 1.22.9. *Let $A \in M_n(C)$. Then there exists a neighborhood of $A = (a_{ij})$*

$$(1.22.10) \qquad D(A,\rho) = \{B \mid B = (b_{ij}) \in M_n(C),\ \sum_{i,j=1}^{n} |b_{ij}-a_{ij}|^2 < \rho^2\},\ 0 < \rho \quad ,$$

where ρ depends on A such that if

$$(1.22.11) \qquad \nu(A,B) = \nu(A,A),\ B \in D(A,\rho) \quad ,$$

then B is similar to A.

<u>Proof.</u> Let r be the rank of $I \otimes A - A^t \otimes I$. So there exists indices

$$\alpha = ((\alpha_{11},\alpha_{21}),\dots,(\alpha_{1r},\alpha_{2r})),\ \beta = ((\beta_{11},\beta_{21}),\dots,(\beta_{1r},\beta_{2r})) \subseteq N \times N,$$
$$N = \{1,\dots,n\}$$

such that $|(I \otimes A - A^t \otimes I)[\alpha|\beta]| \neq 0$. Also $|(I \otimes A - A^t \otimes I)|[\gamma|\delta] = 0$ for $\gamma,\delta \in Q_{(r+1),n^2}$. Here we identify the sets $N \times N$ and $\{1,\dots,n^2\}$. First choose a positive ρ such that

(1.22.12) $$|(I \otimes A - B^t \otimes I)[\alpha|\beta]| \neq 0,\ B \in D(A,\rho) \quad .$$

Next consider the system of r equations in variables x_{ij}, $i,j,\dots,n$, out of n^2 equations of (1.21.11) $(X = (x_{ij}))$ which correspond to set α

(1.22.13) $$\sum_{k=1}^{n} (a_{ik}x_{kj} - x_{ik}b_{kj}) = 0,\ i = \alpha_{1\mu},\ j = \alpha_{2\mu},\ \mu = 1,\dots,r \quad .$$

Let

(1.22.14) $$x_{kj} = \delta_{kj} \quad \text{for} \quad (k,j) \neq (\beta_{1\mu},\beta_{2\mu}),\ \mu = 1,\dots,r \quad .$$

In view of (1.22.12) the system (1.22.13)-(1.22.14) has a unique solution for $B \in D(A,\rho)$. Also $X(A) = I$. Using the continuity argument we deduce the existence of a small positive ρ such that $|X(B)| \neq 0$ for $B \in D(A,\rho)$. We choose such ρ. Let V be the set of matrices B which satisfy

(1.22.15) $$|(I \otimes A - B^t \otimes I)[\gamma|\delta]| = 0,\ r,\delta \in Q_{(r+1),n^2} \quad .$$

Thus V is an algebraic variety. We claim that $V \cap D(A,\rho)$ is exactly the set of matrices of the form (1.22.11). Indeed, let $B \in V \cap D(A,\rho)$. Then according to (1.22.15) $\nu(A,B) \leq r$. On the other hand (1.22.12) implies that $r(A,B) \geq r$. These inequalities yield (1.22.11). Assume that B satisfies (1.22.11). So (1.22.15) holds. Whence $B \in V \cap D(A,\rho)$. Finally, in view of (1.22.15) we deduce that for $B \in V \cap D(A,\rho)$ the equalities (1.22.13) imply $AX(B) - X(B)B = 0$.

As $|X(B)| \neq 0$ we get that $A \approx B$.

□

<u>Problems</u>

(1.22.16) Show that for A and B given in Problem 1.15.8 the three matrices in (1.22.6) are equivalent over $\mathbf{Z}$ however A and B are not similar over $\mathbf{Z}$ (see Problem 1.15.8).

(1.22.17) Show that for $A \in M_n(\mathbf{F})$, $A \approx A + \lambda I$ if and only if $\lambda = 0$. (Compare the traces of A and $A + \lambda I$.)

(1.22.18) Show that if $A = H(n)$ and $B = 0$ then $\nu(A,A) = \nu(A,B) = n$.

(1.22.19) Let $A,B \in M_n(\mathbf{D})$. Assume that the three matrices in (1.22.6) are equivalent. Let I be a maximal ideal in $\mathbf{D}$. Let $\mathbf{F} = \mathbf{D}/I$ and we may view A,B as matrices over $\mathbf{F}$. Prove that A and B are similar over $\mathbf{F}$. (Note that the matrices in (1.22.6) are equivalent over $\mathbf{F}$.)

1.23 The equation $AX - XB = C$.

A related equation to (1.21.1) is a non-homogeneous equation

$$AX - XB = C, \; A \in M_m(\mathbb{F}), \; B \in M_n(\mathbb{F}), \; C \in M_{mn}(\mathbb{F}) \tag{1.23.1}$$

which is written in the tensor notation as

$$(I(n) \otimes A - B^t \otimes I(m))\hat{X} = \hat{C} \quad . \tag{1.23.2}$$

A necessary and sufficient condition for the solvability of (1.23.2) can be stated as follows. Consider a homogeneous system whose coefficient is the transposed coefficient matrix of (1.23.2) (see Problem 1.23.9)

$$(I(n) \otimes A^t - B \otimes I(m))\hat{Y} = 0 \quad .$$

Then (1.23.2) is solvable if and only if any solution $\hat{Y}$ is orthogonal to $\hat{C}$ (e.g. Problem 1.23.10). In matrix the above equality is equivalent to

$$A^tY - YB^t = 0 \quad .$$

The orthogonality of $\hat{Y}$ and $\hat{C}$ is written as $tr(Y^tC) = 0$ (see Problem 1.23.11).

Thus we proved

Theorem 1.23.3. Let $A \in M_m(\mathbb{F})$, $B \in M_n(\mathbb{F})$. Then (1.23.1) is solvable if and only if

$$tr(ZC) = 0 \tag{1.23.4}$$

for all $Z \in M_{nm}(\mathbb{F})$ satisfying

$$ZA - BZ = 0 \quad . \tag{1.23.5}$$

Using Theorem 1.23.3 we can obtain a stronger version of Problem 1.21.22.

<u>Theorem</u> 1.23.6. <u>Let</u> $G = (G_{ij})$, $G_{ij} \in M_{n_i n_j}(F)$, $G_{ij} = 0$ <u>for</u> $j < i$, $i,j = 1,\ldots,\ell$. <u>Then</u>

$$\dim C(G) \geq \sum_{i=1}^{\ell} \dim C(G_{ii}) \quad . \tag{1.23.7}$$

<u>Proof.</u> Consider first the case $\ell = 2$. Let $G = \begin{pmatrix} A & E \\ 0 & B \end{pmatrix}$. Consider a matrix $T = \begin{pmatrix} U & X \\ 0 & V \end{pmatrix}$ which commutes with G. So

$$AU = UA, \ BV = VB, \ AX - XB = UE - EV \quad . \tag{1.23.8}$$

According to Theorem 1.23.3 the matrices $U \in C(A)$ and $V \in C(B)$ satisfy the last equation of (1.23.8) if and only if $\mathrm{tr}[Z(UE-EV)]$ for all Z satisfying (1.23.5). Thus the dimension of the subspace of pairs (U,V) satisfying (1.23.8) is at least

$$\dim C(A) + \dim C(B) - \dim C(B,A) \quad .$$

On the other hand if $U = V = 0$ then the dimension of the subspace of matrices X satisfying (1.23.8) is $\dim C(A,B)$. The equality (1.21.11) implies $\dim C(A,B) = \dim C(B,A)$. Hence we established (1.23.7) for $\ell = 2$. The general case immediately follows by induction on ℓ.

We remark that contrary to the results given in Problem 1.21.22 the equality in (1.23.7) may occur even if $G_{ii} = G_{jj}$ for some $i \neq j$. See Problem (1.23.12).

<u>Problems</u>

(1.23.9) Let $A \otimes B$ be defined as in (1.21.19). Prove $(A \otimes B)^t = A^t \otimes B^t$.

(1.23.10) Consider a system

$$Ax = b, \quad A \in M_{mn}(\mathbb{F}), \; b \in M_{n1}(\mathbb{F}) \quad .$$

Prove that the above system is solvable if and only if $y^t b = 0$ where y is a solution of the system $A^t y = 0$. (Hint: Change varibles to obtain A in its diagonal form as in Section 1.12).

(1.23.11) Let $X, Y \in M_{mn}(\mathbb{D})$. Let $\mu(X), \mu(Y) \in M_{(mn)1}(\mathbb{D})$ be defined as in Problem 1.21.17. Prove that

$$\mu(Y)^t \mu(X) = \mathrm{tr}(Y^t X) \quad .$$

(1.23.12) Assume in Theorem 1.23.6 $\ell = 2$, $G_{11} = G_{22} = 0$, $G_{12} = I$. Show that in this case the equality sign holds in (1.23.7).

(1.23.13) Let $A_i \in M_{n_i}(\mathbb{F})$, $i = 1,2$ and suppose that A_1 and A_2 do not have a common eigenvalue. Assume that $A = A_1 \oplus A_2$. Let $C = (C_{ij})$, $X = (X_{ij})$, $C_{ij}, X_{ij} \in M_{n_i n_j}(\mathbb{F})$, $i,j = 1,2$. Using Problem 1.21.22 prove that the equation $AX - XA = C$ is solvable if and only if the equations $A_i X_{ii} - X_{ii} A_i = C_{ii}$, $i = 1,2$.

1.24 A case of two nilpotent matrices.

The following result is needed later.

Theorem 1.24.1. *Let* $A \in M_n(\mathbb{F})$ *be a nilpotent matrix. Put*

$$X_k = \{x \mid x \in \mathbb{F}^n,\ A^k x = 0\},\ k = 0,1,\ldots,\quad .$$

Assume that

$$0 = X_0 \subsetneqq X_1 \subsetneqq X_2 \subsetneqq \cdots \subsetneqq X_p = \mathbb{F}^n \quad .$$

Suppose that $B \in M_n(\mathbb{F})$ *satisfies*

$$BX_{i+1} \subset X_i,\ i = 0,\ldots,p-1 \quad . \tag{1.24.1}$$

Then

$$\nu(A,A) \leq \nu(B,B) \quad . \tag{1.24.3}$$

The equality sign holds if and only if B *is similar to* A.

Proof. We prove the lemma by induction on p. For $p = 1$ the theorem is trivial since $A = B = 0$. Suppose that the lemma holds for $p = q - 1$ and we prove it for $p = q$. A_1 and B_1 be the restrictions of A and B to X_{q-1}. Assume that A_1, A and B_1, B have the following Jordan canonical forms

$$P_1^{-1}A_1P_1 = \sum_{i=1}^{\mu'} \oplus H(m_i'),\ P^{-1}AP = \sum_{i=1}^{\mu} \oplus H(m_i) \quad ,$$

$$Q_1^{-1}B_1Q_1 = \sum_{i=1}^{\nu'} \oplus H(n_i'),\ Q^{-1}BQ = \sum_{i=1}^{\nu} \oplus H(n_i) \quad ,$$

(1.24.4)

$$m_1' \geq m_2' \geq \ldots \geq m_{\mu'}' \geq 1,\ n_1' \geq n_2' \geq \ldots \geq n_{\nu'}' \leq 1,\ \sum_{i=1}^{\mu'} m_i' = \sum_{i=1}^{\nu'} n_i' = \dim X_{q-1} \quad ,$$

$$m_1 \geq m_2 \geq \ldots \geq m_{\mu} \geq 1,\ n_1 \geq n_2 \geq \ldots \geq n_{\nu} \geq 1,\ \sum_{i=1}^{\mu} m_i = \sum_{i=1}^{\nu} n_i = \dim X_q = n \quad .$$

Clearly the Jordan canonical form of A_1 and A is determined completely by the dimensions of the subspaces $X_1,\ldots,X_{q-1}$. Indeed, put

$$\psi_i = \dim X_i,\ \theta_i = \psi_{i+1} - \psi_i,\ i = 0,1,\ldots \quad .$$

Then $\theta_{i-1} - \theta_i$ is the number of the Jordan blocks of order i in the Jordan canonical form of A and A_1. So

(1.24.5) $\mu = \mu',\ m_i' = m_i - 1,\ i = 1,\ldots,\theta_{q-1},\ m_i' = m_i$ for $i > \theta_{q-1}$,

$$\theta_{q-1} = n - \dim X_{q-1} \quad .$$

Formula (1.21.14) yields

(1.24.6) $$\nu(A,A) = \sum_{u=1}^{\mu} (2u-1)m_u = \sum_{u=1}^{\mu'} (2u-1)m_u' + \sum_{u=1}^{\theta_{q-1}} (2u-1) = \nu(A',A') + \theta_{q-1}^2 \quad .$$

According to Theorem 1.20.10 we have the following relations between the Jordan canonical form of B and B'

$\nu \geq \nu'$, $n_i' + 1 \geq n_i \geq n_i'$, $i = 1,\ldots,\nu'$, $n_i = 1$, $i = \nu-\nu'+1,\ldots,\nu$ (if $\nu > \nu'$).

So

$$(1.24.7)\ \nu(B,B) = \sum_{i=1}^{\nu}(2u-1)n_i \geq \sum_{i=1}^{\nu'}(2u-1)n_i + \sum_{u=1}^{\theta_{q-1}}(2u-1) = \nu(B',B') + \theta_{q-1}^2$$

and the equality sign holds if and only if $n_i = n_i' + 1$, $i = 1,\ldots,\theta_{q-1}$. Combine (1.24.6) with (1.24.7) to deduce (1.24.3) by induction. Suppose that the equality sign holds in (1.24.3). Since

$$\nu(A',A') = \nu(B',B'),\ n_i = n_i' + 1,\ i = 1,\ldots,\theta_{q-1}\ .$$

The induction assumptions imply that $A' \approx B'$. Use (1.24.5) and the above equalities to get $A \approx B$.

□

1.25 Components of a matrix and functions of matrices.

From this section and through the end of this chapter we shall assume that all the matrices are complex valued unless otherwise stated. Let $\varphi(x)$ be a polynomial ($\varphi(x) \in C[x]$). The following relations are easily established:

$$(1.25.1)\qquad \begin{aligned} &\varphi(B) = P\varphi(A)P^{-1},\ B = PAP^{-1},\ A,B,P \in M_n(C)\ , \\ &\varphi(A_1 \oplus A_2) = \varphi(A_1) \oplus \varphi(A_2),\ A_i \in M_{n_i}(C),\ i = 1,2\ . \end{aligned}$$

It often pays to know the explicit formula for $\varphi(A)$ in terms of the Jordan canonical form of A. In view of (1.25.1) it is enough to consider the case where J is composed of one Jordan block.

Lemma 1.25.2. Let $J = \lambda_0 I + H \in M_n(C)$. Then for any $\varphi \in C[x]$ we have

$$\varphi(J) = \sum_{k=0}^{n-1} \frac{\varphi^{(k)}(\lambda_0)}{k!} H^k\ .$$

Proof. For any φ we have the Taylor expansion $\varphi(x) = \sum_{k=0}^{N} \frac{\varphi^{(k)}(\lambda_0)}{k!}(x-\lambda_0)^k$, $N = \max(\deg \varphi, n)$. As $H^\ell = 0$ for $\ell > n$ from the above equality we deduce the lemma.

□

Using Jordan canonical form of A we obtain

Theorem 1.25.3. Let $A \in M_n(C)$. Assume that Jordan canonical form of A is given by (1.19.11). Then for $\varphi(x) \in C[x]$ we have

$$(1.25.4)\qquad \varphi(A) = P\{\sum_{i=1}^{\ell} \sum_{j=1}^{q_i} \oplus [\sum_{k=0}^{m_{ij}-1} \frac{\varphi^{(k)}(\lambda_i)}{k!} H^k(m_{ij})]\}P^{-1}\ .$$

Definition 1.25.5. Let the assumptions of Theorem 1.25.3 prevail. Then $Z_{ik} = Z_{ik}(A)$ is called the (i,k) component of A and is given by

$$(1.25.6)\qquad Z_{ik} = P\{0\oplus\dots 0\oplus[\sum_{j=1}^{q_i} \oplus H^k(m_{ij})]\oplus 0\dots\oplus 0\}P^{-1},\ k = 0,\dots,s_i-1,\ s_i = m_{i1},\quad i = 1,\dots,\ell\ .$$

Compare (1.25.4) with (1.25.6) to deduce

$$(1.25.7)\qquad \varphi(A) = \sum_{i=1}^{\ell} \sum_{j=0}^{s_i-1} \frac{\varphi^{(j)}(\lambda_i)}{j!} Z_{ij}\ .$$

Definition 1.25.8. Let $\Omega \subseteq C$ be an open set such that $\sigma(A) \subset \Omega$, where $A \in M_n(C)$. Then for $\varphi \in H(\Omega)$ define $\varphi(A)$ by the equality (1.25.7).

Using (1.25.6) it is easy to verify that the components of A satisfy

(1.25.9) $Z_{ij},\ i = 1,\dots,\ell,\ j = 0,\dots,s_i - 1,$ are linearly independent ,

$Z_{ij}Z_{\alpha\beta} = 0$ if/ $i = \alpha$, $Z_{ij}Z_{ik} = 0$ if $j + k > s_i$, $Z_{ij}Z_{ik} = Z_{i(k+j)}$, for $j+k \le s - 1$,

$$A = P(\sum_{i=1}^{\ell} \lambda_i Z_{i0} + Z_{i1})P^{-1}\ .$$

Consider the component $Z_{i(s_i-1)}$ The relations (1.25.9) imply

$$(1.25.10)\qquad AZ_{i(s_i-1)} = Z_{i(s_i-1)}A = \lambda_i Z_{i(s_i-1)}.$$

Thus the columns of $Z_{i(s_i-1)}$ and $Z^t{}_{i(s_i-1)}$ are the eigenvectors of A and A^t respectively. Clearly, $Z_{i(s_i-1)} \neq 0$. More precisely we have

Lemma 1.25.11. *Let* $A \in M_n(\mathbf{C})$. *Suppose that* λ_i *is an eigenvalue of* A. *Let* X_i *be the generalized eigenspace of* A *corresponding to* λ_i

$$(1.25.12) \qquad X_i = \{x | x \in \mathbf{C}^n, (\lambda_i I - A)^{s_i} x = 0\} \quad .$$

Then

$$(1.25.13) \qquad r(Z_{i(s_i-1)}) = \dim[(\lambda_i I - A)^{s_i-1} X_i] \quad .$$

Proof. It is enough to assume that A is in its Jordan form. Then X_i is the subspace of all $x = (x_1,\ldots,x_n)^t$ where the first $\sum_{\alpha=1}^{i-1} \sum_{j=1}^{q_\alpha} m_{\alpha j}$ and the last $\sum_{\alpha=i+1}^{\ell} \sum_{j=1}^{q_\alpha} m_{\alpha j}$ vanish. So $(\lambda_i I - A)^{s_i-1} X$ contains only those eigenvectors which correspond to Jordan blocks of the length s_i. Obviously, the rank $Z_{i(s_i-1)}$ is exactly the number of such blocks.

□

Definition 1.15.14. *Let* $A \in M_n(\mathbf{C})$. *Then the spectral radius of* A - $\rho(A)$ - *is defined*

$$(1.25.15) \qquad \rho(A) = \max_{\lambda \in \sigma(A)} |\lambda| \quad .$$

The peripherial spectrum of A - $\sigma_p(A)$ - *is the set of all eigenvalues* $\{\lambda_1,\ldots,\lambda_m\}$ *(each one appearing according to its multiplicity) which are on the spectral circle* $|x| = \rho(A)$. *The district peripherial spectrum of* A - $\sigma_{dp}(A)$ - *is given*

$$(1.25.16) \qquad \sigma_{dp}(A) = \sigma_d(A) \cap \sigma_p(A) \quad .$$

The index A - index (A) *is defined by*

(1.25.17) $$\operatorname{index}(A) = \max_{\lambda \varepsilon \sigma_p(A)} \operatorname{index}(\lambda) \quad .$$

Problems

(1.25.18) Let $A \varepsilon M_n(\mathbb{C})$, $\psi(x) \varepsilon \mathbb{C}[x]$ be the minimal polynomial of A. Assume that Ω is an open set in $\mathbb{C}$ such that $\sigma(A) \subset \Omega$. Let $\varphi \varepsilon H(\Omega)$. Then the values

(1.25.19) $$\varphi^{(k)}(\lambda),\ k = 0,1,\ldots,\ \operatorname{index}(\lambda) - 1,\ \lambda \varepsilon \sigma_d(A)$$

are called the <u>values</u> of φ on the <u>spectrum</u> of A $(\sigma(A))$. Two functions, φ,θ are said to be <u>coinciding</u> on $\sigma(A)$ if they have the same values on $\sigma(A)$. Assume that $\varphi \varepsilon \mathbb{C}[x]$ and let

$$\varphi(x) = \omega(x)\psi(x) + \theta(x),\ \deg\theta < \deg\psi \quad .$$

Prove that $\theta(x)$ coincides with $\varphi(x)$ on $\sigma(A)$. Let

$$\frac{\varphi(x)}{\psi(x)} = \omega(x) + \frac{\theta(x)}{\psi(x)} = \omega(x) + \sum_{i=1}^{\ell} \sum_{j=1}^{s_i} \frac{\alpha_{ij}}{(x-\lambda_i)^j} \quad , s_i = \operatorname{index}(\lambda_i),\ i = 1,\ldots,\ell \quad ,$$

where $\psi(x)$ is given by (1.18.20). Show that α_{ij}, $j = s_i, s_i-1,\ldots,s_i-p$, are determined recursively by $\varphi^{(j)}(\lambda_i)$, $j = 0,\ldots,p$. (Multiply the above equality by $\psi(x)$ and evaluate this identity and its derivatives at λ_i.) Thus for any $\varphi \varepsilon H(\Omega)$ define $\theta(x)$ by the equality

(1.25.20) $$\theta(x) = \psi(x) \sum_{i=1}^{\ell} \sum_{j=1}^{s_i} \frac{\alpha_{ij}}{(x-\lambda_i)^j} \quad .$$

The polynomial $\theta(x)$ is called the <u>Lagrange-Sylvester</u> (L-S) <u>interpolation</u> polynomial of φ (corresponding to ψ). Prove that

(1.25.21) $$\varphi(A) = \theta(A) \quad .$$

Let $\theta_j(x)$ be L-S polynomials of $\varphi_j \varepsilon H(\Omega)$, $j = 1,2$. Show that $\theta_1(x)\theta_2(x)$ coincides with L-S polynomial of $\varphi_1(x)\varphi_2(x)$ on $\sigma(A)$. Use this fact to prove the identity

(1.25.22) $$\varphi_1(A)\varphi_2(A) = \varphi(A), \quad \varphi(x) = \varphi_1(x)\varphi_2(x) \quad .$$

(1.25.23) Prove (1.25.22) by using the definition (1.25.7) and the relations (1.25.9).

(1.25.24) Let the assumptions f Problem 1.25.18 prevail. Assume that a sequence $\varphi_m \varepsilon H(\Omega)$, converges to $\varphi \varepsilon H(\Omega)$. That is $\varphi_m(x)$, $m = 1,2,\ldots,$ converge uniformly to $\varphi(x)$ on any compact set of Ω. This in particular implies

$$\lim_{m\to\infty} \varphi_m^{(j)}(\lambda) = \varphi^{(j)}(\lambda), \; j = 0,1,\ldots, \; \lambda \varepsilon \Omega \quad .$$

Use the definition (1.25.7) to show

(1.25.25) $$\lim_{m\to\infty} \varphi_m(A) = \varphi(A) \quad .$$

Apply this result to prove

(1.25.26) $$e^A = \sum_{m=0}^{\infty} \frac{A^m}{m!} \; (= \lim_{N\to\infty} \sum_{m=0}^{N} \frac{A^m}{m!})$$

(1.25.27) $$(\lambda I - A)^{-1} = \sum_{m=0}^{\infty} \frac{A^m}{\lambda^{m+1}} \quad \text{for} \quad |\lambda| > \rho(A) \quad .$$

1.26 Cesaro convergence of matrices

Let

(1.26.1) $$A_k = (a_{ij}^{(k)}) \in M_{mn}(\mathbb{C}),\ k = 0,1,2,\ldots$$

be a sequence of matrices. The p-th Cesaro sequence is defined as

(1.26.2) $$A_{k,0} = A_k,\ A_{k,p} = (a_{ij}^{(k,p)}) = \sum_{j=0}^{k} A_{j,p-1}/(k+1), k = 0,1,\ldots, \quad .$$

Definition 1.26.3. *A sequence* $\{A_k\}$ *is said to be convergent to* $A = (a_{ij})$ *if*

$$\lim_{k\to\infty} a_{ij}^{(k)} = a_{ij},\ i = 1,\ldots,m,\ j = 1,\ldots,n, \iff \lim_{k\to\infty} A_k = A \quad .$$

A sequence $\{A_k\}$ *is said to be* $p(\geq 0)$ - *Cesaro convergent to* A

$$\lim_{k\to\infty} A_{k,p} = A$$

and exactly $p(\geq 1)$ - *Cesaro convergent if in addition to the above equality the sequence* $A_{k,p-1}$, $k = 0,1\ldots,$ *is not convergent.*

It is a standard fact (e.g. Hardy [1949]) that if $\{A_k\}$ is p-Cesaro convergent then $\{A_k\}$ is $(p+1)$ - Cesaro convergent. A standard example of exactly 1 - Cesaro convergence sequence is $\{\lambda^k\}$, $|\lambda| = 1$, $\lambda \neq 1$. More precisely we have (e.g. Hardy [1949] or Problem 1.26.11).

Lemma 1.26.4. *Let* $|\lambda| = 1$, $\lambda \neq 1$. *Then the sequence* $\binom{k}{p-1}\lambda^k$, $k = 0,1,\ldots,$ *is exactly p-Cesaro convergent to zero for* $p \geq 1$.

We now show how to recover the component $Z_{\alpha(s_\alpha-1)}$ for $0 \neq \lambda_\alpha \in \sigma_p(A)$ by using the notion of Cesaro convergence.

Theorem 1.26.5. Let $A \in M_n(\mathbf{C})$. Assume that $\rho(A) > 0$ and $\lambda_\alpha \in \sigma_p(A)$. Put

$$(1.26.6) \qquad A_k = (s_\alpha - 1)!\, [\bar{\lambda}_\alpha A/|\lambda_\alpha|^2]^k / k^{s_\alpha - 1}, \quad s_\alpha = \text{index}\,(\lambda_\alpha) \ .$$

Then

$$(1.26.7) \qquad \lim_{k \to \infty} A_{k,p} = Z_{\alpha(s_\alpha - 1)}, \quad p = \text{index}\,(A) - \text{index}\,(\lambda_\alpha) + 1 \ .$$

The sequence A_k is exactly p-Cesaro convergent unless $\sigma_{dp}(A) = \{\lambda_\alpha\}$ or

$$\text{index}\,(\lambda) < \text{index}\,(\lambda_\alpha) \quad \text{for any} \quad \lambda \in \sigma_p(A),\ \lambda \neq \lambda_2 \ .$$

In that case the sequence $\{A_k\}$ converges to $Z_{\alpha(s_\alpha - 1)}$.

Proof. It is enought to consider the case where $\lambda_\alpha = \rho(A) = 1$. By letting $\varphi(x) = x^k$ in (1.25.7) we get

$$(1.26.8) \qquad A^k = \sum_{i=1}^{\ell} \sum_{j=0}^{s_i - 1} \binom{k}{j} \lambda_i^{k-j} Z_{ij} \ .$$

So

$$A_k = \sum_{i=1}^{\ell} \sum_{j=0}^{s_i - 1} \frac{(s_\alpha - 1)!}{k^{s_\alpha - 1}} \frac{k(k-1)\ldots(k-j+1)}{j!} \lambda_i^{k-j} Z_{ij} \ .$$

Since the components Z_{ij}, $i = 1,\ldots,\ell$, $j = 1,\ldots,s_i - 1$, are linearly independent it is enough to analyze the sequence $\frac{(s_\alpha - 1)!}{k^{s_\alpha - 1}} \binom{k}{j} \lambda_i^{k-j}$, $k = j, j+1, \ldots$. Clearly for $|\lambda| < 1$

and any j or for $|\lambda| = 1$ and $j < s_\alpha - 1$ this sequence converges to zero. For $\lambda_i = 1$ and $j = s_\alpha - 1$ the above sequence converges to 1. For $|\lambda_i| = 1$, $\lambda_i \neq 1$ and $j \geq s_\alpha - 1$ the given sequence is exactly $j - s_\alpha + 2$ Cesaro convergent to zero in view of Lemma 1.26.4. These arguments establish the theorem. □

Corollary 1.26.9. Let the assumptions of Theorem 1.26.5 hold. Then

$$(1.26.10) \qquad \lim_{N\to\infty} \left[\sum_{k=0}^{N} (s-1)!\, (\rho^{-1}(A)A)^k / k^{s-1} \right] / (N+1) = Z, \quad s = \text{index}\,(A)$$

where $Z = 0$ unless $\lambda_1 = \rho(A) \in \sigma(A)$ and index (λ_1) = index (A) in which case $Z = Z_{1(s-1)}$.

Problems

(1.26.11) Let $|\lambda| = 1$, $\lambda \neq 1$ be fixed. Using the formula

$$\sum_{j=0}^{k-1} \lambda^j = \frac{\lambda^k - 1}{\lambda - 1}$$

prove by differentiating the above equality r times

$$\sum_{j=0}^{k-1} \binom{j}{r} \lambda^j = k \left[\sum_{\ell=0}^{r-1} \alpha(\lambda,r,\ell) \binom{k-1}{\ell} \lambda^{k-1} \right] + (-1)^r \frac{\lambda^k - 1}{(\lambda-1)^{r+1}}, \quad k \geq 1$$

where $\alpha(\lambda,r,\ell)$ are some fixed non-zero functions. Use the induction on r to prove Lemma 1.26.4.

(1.26.12) Let $\varphi(x)$ be a normalized polynomial of degree $p-1$. Prove that the sequence $\{\varphi(k)\lambda^k\}$, $|\lambda| = 1$, $\lambda \neq 1$, is exactly p-Cesaro convergent.

(1.26.13) Let $A \in M_n(\mathbf{C})$. Put

(1.26.14) $$Z_{ij}(A) = (z_{\mu\nu}^{(ij)}),\ \mu,\nu = 1,\dots,n,\ \lambda_i \in \sigma_d(A),\ j = 0,\dots,\text{index}\ (\lambda_i) - 1\ .$$

Denote

(1.26.15) $\text{index}_{\mu\nu}(\lambda_i) = 1 + \max\{j : z_{\mu\nu}^{(ij)} \neq 0,\ j = 0,\dots,\text{index}\ (\lambda_i) - 1\}$, where $\text{index}_{\mu\nu}(\lambda_i) = 0$ if $z_{\mu\nu}^{(ij)} = 0,\ j = 0,\dots,\text{index}\ (\lambda_i) - 1$.

(1.26.16) $\rho_{\mu\nu}(A) = \max\{|\lambda_i| : \text{index}_{\mu\nu}(\lambda_i) > 0\}$ where $\rho_{\mu\nu}(A) = -\infty$ if $\text{index}_{\mu\nu}(\lambda_i) = 0$ for all $\lambda_i \in \sigma(A)$.

(1.26.17)

$$\text{index}_{\mu\nu}(A) = \max\{\text{index}_{\mu\nu}(\lambda_i) : \text{index}_{\mu\nu}(\lambda_i) > 0,\ |\lambda_i| = \rho_{\mu\nu}(A)\}\ .$$

Here $\text{index}_{\mu\nu}(A) = 0$ if $\rho_{\mu\nu}(A) = -\infty$. The quantities $\text{index}_{\mu\nu}(\lambda_i)$, $\text{index}_{\mu\nu}(A)$, $\rho_{\mu\nu}(A)$ are called the (μ,ν) index of λ_i, the (μ,ν) index of A and the (μ,ν) spectral radius of A respectively. Or shortly the local indices of λ_i and A and the local spectral radius of A. Show that Theorem 1.26.5 and Corollary 1.16.9 could be stated in a local form. That is for $1 \leq \mu,\nu \leq n$ assume that $|\lambda_\alpha| = \rho_{\mu\nu}(A)$, $s_\alpha = \text{index}_{\mu\nu}(\lambda_\alpha)$, $A^k = (a_{\mu\nu}^{(k)})$, $A_k = (a_{\mu\nu,k})$, $A_{k,p} = (a_{\mu\nu,kp})$ where A_k is given by (1.26.6) and $A_{k,p}$ by (1.26.2). Prove

(1.26.7)' $$\lim_{k\to\infty} a_{\mu\nu,kp} = z_{\mu\nu}^{(\alpha(s_\alpha-1))},\ p = \text{index}_{\mu\nu}(A) - \text{index}_{\mu\nu}(\lambda_\alpha) + 1\ ,$$

(1.26.10)' $$\lim_{N\to\infty}\{\sum_{k=0}^{N} (s-1)!\ [\rho_{\mu\nu}^{-1}(A)]^k a_{\mu\nu,k}/k^{s-1}\}/(N+1) = z_{\mu\nu},$$
$$s = \text{index}_{\mu\nu}(A),\ \rho_{\mu\nu}(A) > 0\ ,$$

where $z_{\mu\nu} = 0$ unless $\lambda_1 = \rho_{\mu\nu}(A) \in \sigma(A)$ and $\text{index}_{\mu\nu}(\lambda_1) = \text{index}_{\mu\nu}(A)$ in which case $z_{\mu\nu} = z_{\mu\nu}^{1(s-1)}$. Finally A is called <u>irreducible</u> if $\rho_{\mu\nu}(A) = \rho(A)$ for $\mu,\nu = 1,\dots,n$. Thus for irreducible A the local and the global versions of Theorem 1.26.5 and Corollary 1.26.9 coincide.

1.27 An iteration process

Consider an iteration given by

(1.27.1) $$x^{i+1} = Ax^i + b,\ i = 0,1,\ldots$$

where $A \in M_n(C)$, $x^i, b \in M_{n1}(C)$. Such an iteration can be used to solve a system

(1.27.2) $$x = Ax + b \quad .$$

Also, the iteration (1.27.1) appears naturally in certain physical instances where a given physical system evolves discretely in time according to (1.27.1) (e.g. Berman-Plemmons [1979]. Assume that x is the only solution of (1.27.2) and put $y^i = x^i - x$. Then

(1.27.3) $$y^{i+1} = Ay^i,\ i = 0,1,2,\ldots \quad .$$

We would like to know under what conditions $\lim_{i\to\infty} y^i = 0$, regardless of the initial condition y^0. This is equivalent to the statement that $x^{(i)}$ converges to x for any initial condition $x^{(0)}$. In some other instances the evolusion of a certain physical system is given by (1.27.3). In that case it is important to know whether the system would not blow up. That is there exists a constant $M(y^0)$ (depending on y^0) such that

(1.27.4) $$\|y^i\| \leq M(y^0),\ i = 0,1,\ldots \quad .$$

Here for $B \in M_{mn}(C)$ we define the norm of B - $\|B\|$ - as

(1.27.5) $$\|B\| = \max_{1\leq i\leq m} \sum_{j=1}^{n} |b_{ij}|,\ B = (b_{ij}) \in M_{mn}(C) \quad .$$

Definition 1.27.6. The system (1.27.3) is called stable if the sequence y^i, $i = 0,1,\ldots,$ converges to zero for all choices of y^0. The system (1.27.3) is called bounded if the sequence y^i, $i = 0,1,\ldots,$ is bounded for all choices of y^0, i.e. (1.27.4) holds.

Clearly, the solution to (1.27.3) is

$$y^i = A^i y^0,\ i = 0,1,\ldots, \quad ,$$

so (1.27.3) is stable if and only if

(1.27.7) $$\lim A^i = 0$$

and (1.27.3) is bounded if and only if

(1.27.8) $$\|A^i\| \leq M,\ i = 0,1,\ldots,$$

for some positive number M.

Theorem 1.27.9 Let $A \in M_n(C)$. Then the powers A^i, $i = 0,1,\ldots,$ converge to zero matrix if and only if

(1.27.10) $\rho(A) < 1$.

<u>These powers are bounded if and only if</u>

(1.27.11) $\rho(A) \leq 1$, <u>and</u> index $(A) = 1$ <u>if</u> $\rho(A) = 1$.

<u>Proof.</u> Consider the formula (1.26.8). Since all the components of A are linearly independent (1.27.7) is equivalent to

$$\lim_{k\to\infty} \binom{k}{j}\lambda_i^{k-j} = 0,\ \lambda_i \in \sigma_d(A),\ j = 0,1,\dots,\ \text{index}\ (\lambda_i) - 1 \ .$$

Of course, the above conditions are equivalent to (1.27.10). The condition (1.27.8) is equivalent to the statement that the sequence $\binom{k}{j}\lambda_i^{k-j}$, $k = 0,1,\dots,$ is bounded. This immediately yields that $\rho(A) \leq 1$. If $\rho(A) < 1$ then (1.27.7) holds which implies (1.27.8). Suppose that $\rho(A) = 1$ and let $\lambda_i \in \sigma_p(A)$, i.e. $|\lambda_i| = 1$. Then the sequence $\binom{k}{j}\lambda_i^{k-j}$ is bounded if and only if $j = 0$. That is we must have index $(\lambda_i) = 1$. This establishes (1.27.11).

□

<u>Problems</u>

(1.27.12) Let $A \in M_n(\mathbb{C})$ and $\psi(x)$ the minimal polynomial of A is given by (1.18.20). Verify

(1.27.13) $$e^{At} = \sum_{i=1}^{\ell} \sum_{j=0}^{s_i-1} \frac{t^j e^{\lambda_i t}}{j!} Z_{ij} \ .$$

Use (1.25.9) or (1.25.26) to prove

(1.27.14) $$\frac{d}{dt}(e^{At}) = Ae^{At} = e^{At}A \ .$$

Prove that the system

(1.27.15) $$\frac{dx}{dt} = Ax,\ x(t) \in M_{n1}(\mathbb{C})$$

has a solution

$$x(t) = e^{A(t-t_0)} x(t_0) \quad . \tag{1.27.16}$$

The system (1.27.15) is called stable if $\lim_{t\to\infty} x(t) = 0$ for any solution (1.27.16). The system (1.27.15) is called bounded if any solution (1.27.16) satisfies

$$\|x(t)\| < M, \; t > t_0, \; M = M(x(t_0)) \quad .$$

Prove that (1.27.15) is stable if and only if

$$\mathrm{Re}\{\lambda\} < 0 \quad \text{for} \quad \lambda \in \sigma(A) \quad , \tag{1.27.17}$$

and (1.27.15) is bounded if and only if

$$\mathrm{Re}\{\lambda\} \le 0, \; \text{index}\,(\lambda) = 1 \quad \text{if} \quad \mathrm{Re}\{\lambda\} = 0 \quad . \tag{1.27.18}$$

1.28 <u>The Cauchy formula for functions of matrices.</u>

Let $A \in M_n(\mathbf{C})$, $\varphi \in H(\Omega)$, where Ω is an open domain in $\mathbf{C}$. Here we do not assume that Ω is connected. If $\sigma(A) \subset \Omega$ then it is possible to define $\varphi(A)$ by (1.25.7). It is possible to give an integral formula for (A) by using the Cauchy integration formula for $\varphi(\lambda)$. The resulting expression is simply looking and very useful in theoretical studies of $\varphi(A)$. Moreover, this formula remains valid for bounded operators in Banach spaces (e.g. Kato [1970]). To do so we consider the function $\varphi(x,\lambda) = (\lambda-x)^{-1}$. The domain of the analycity of $\varphi(x,\lambda)$ is the whole complex plane $\mathbf{C}$ punctured at λ. Thus if $\lambda \notin \sigma(A)$ according to (1.25.7)

$$(1.28.1) \qquad (\lambda I-A)^{-1} = \sum_{i=1}^{\ell} \sum_{j=0}^{s_i-1} (\lambda-\lambda_i)^{-(j+1)} Z_{ij} \quad .$$

<u>Definition</u> 1.28.2. <u>The function</u> $(\lambda I-A)^{-1}$ <u>is called the resolvent of</u> A <u>and is denoted by</u>

$$(1.28.3) \qquad R(\lambda,A) = (\lambda I-A)^{-1} \quad .$$

Let $\Gamma = \{\Gamma_1,\ldots,\Gamma_k\}$ be a set of disjoint simply connected rectifiable curves such that Γ forms a boundary ∂D of an open domain where

$$(1.28.4) \qquad D \cup \Gamma \subset \Omega, \ \Gamma = \partial D \quad .$$

For $\varphi \in H(\Omega)$ the classical Cauchy integration formula states (e.g. Rudin [1974])

$$(1.28.5) \qquad \varphi(\lambda) = \frac{1}{2\pi i} \int_{\Gamma} (x-\lambda)^{-1} \varphi(x)dx, \ \lambda \in D \quad .$$

By differentiating the above equality j times we get

$$(1.28.6) \qquad \frac{\varphi^{(j)}(\lambda)}{j!} = \frac{1}{2\pi i} \int_{\Gamma} (x-\lambda)^{-(j+1)} \varphi(x)dx, \ \lambda \in D \quad .$$

We now are ready to state the Cauchy integration formula for $\varphi(A)$.

<u>Theorem</u> 1.28.7. <u>Let</u> Ω <u>be an open domain in the complex plane. Assume that</u> $\Gamma = \{\Gamma_1,\ldots,\Gamma_k\}$ <u>be a set of disjoint simple connected rectifiable curves such that</u> Γ

is a boundary of an open domain D, and $\Gamma \cup D \subset \Omega$. Let $A \in M_n(\mathbf{C})$ and assume that $\sigma(A) \subset D$. Then for any $\varphi(x)$ analytic in Ω we have

$$\varphi(A) = \frac{1}{2\pi i} \int_\Gamma R(x,A)\varphi(x)dx \quad . \tag{1.28.8}$$

Proof. Insert the expression (1.28.1) into the above integral to get

$$\frac{1}{2\pi i} \int_\Gamma R(x,A)\varphi(x)dx = \sum_{k=1}^{\ell} \sum_{j=0}^{s_k-1} \left[\frac{1}{2\pi i} \int_\Gamma (x-\lambda_k)^{-(j+1)}\varphi(x)dx\right] Z_{kj} \quad .$$

Now use the identity (1.28.6) to deduce

$$\frac{1}{2\pi i} \int_\Gamma R(x,A)\varphi(x)dx = \sum_{k=1}^{\ell} \sum_{j=0}^{s_k-1} \frac{\varphi^{(j)}(\lambda_k)}{j!} Z_{kj} \quad .$$

The definition (1.25.7) of $\varphi(A)$ yields the equality (1.28.8).

□

We illustrate the usefulness of Cauchy integral formula (1.28.8) by two examples.

Theorem 1.28.9. Let $A \in M_n(\mathbf{C})$ and assume that $\lambda_p \in \sigma(A)$. Let

$$\lambda_p \in D, \ \lambda \notin D \cup \Gamma \ \text{ for } \ \lambda \in \sigma(A), \ \lambda \neq \lambda_p \quad , \tag{1.28.10}$$

where D is an open domain in $\mathbf{C}$ and Γ satisfies the assumptions of Theorem 1.28.7. Then the p,q component of A is given by

$$Z_{pq}(A) = \frac{1}{2\pi i} \int_\Gamma R(x,A)(x-\lambda_p)^{q-1}dx \quad . \tag{1.28.11}$$

__Proof.__ As in the proof of Theorem 1.28.7

$$\frac{1}{2\pi i}\int_\Gamma R(x,A)(x-\lambda_p)^{q-1}\,dx = \sum_{k=1}^{\ell}\sum_{j=0}^{s_k-1}\left[\frac{1}{2\pi i}\int_\Gamma \frac{(x-\lambda_p)^{q-1}}{(x-\lambda_k)^j}\,dx\right]Z_{kj} \quad .$$

Suppose that $\lambda_k \neq \lambda_p$. Then $\frac{(x-\lambda_p)^{q-1}}{(x-\lambda_k)^j}$ is analytic in $\Gamma \cup D \subset \Omega$. Put $\varphi(x) = \frac{(x-\lambda_p)^q}{(x-\lambda_k)^k}$.

and use (1.28.5) to deduce that the corresponding

integral which appears in the above equality vanishes. The same result applies for $\lambda_k = \lambda_p$ and $j < q-1$. For $j > q+1$, put $\varphi(x) = 1$ then

$\frac{(x-\lambda_p)^{q-1}}{(x-\lambda_k)^j} = \frac{\varphi(x)}{(x-\lambda_p)}$ j-q+1. Apply (1.28.6) to deduce also that the corresponding integral with this term is equal to zero. Hence

$$\frac{1}{2\pi i}\int_\Gamma R(x,A)(x-\lambda_p)^{-q+1}dx = \left[\frac{1}{2\pi i}\int_\Gamma \frac{dx}{x-\lambda_p}\right]Z_{pq} = Z_{pq}.$$

□

Our next example generalize the first part of Theorem 1.27.9 to a compact set of matrices.

__Definition__ 1.28.12. __A set__ $\mathcal{A} \subseteq M_n(\mathbb{C})$ __is called stable if__

(1.28.13) $$\lim_{k\to\infty} \sup_{A\in\mathcal{A}} \|A^k\| = 0 \quad .$$

__Theorem__ 1.28.14. __Let__ $\mathcal{A} \subseteq M_n(\mathbb{C})$ __be a compact set. Then__ $\mathcal{A}$ __is stable if and only if__ $\rho(A) < 1$ __for any__ $A \in \mathcal{A}$.

To prove the theorem we need a well known result on the roots of normalized polynomials in $\mathbb{C}[x]$ (e.g. Ostrowski [1966]).

Lemma 1.28.15. Consider a normalized polynomial

$p(x) = x^m + \sum_{i=1}^{m} a_i x^{m-i} \in C[x]$. Then the zeros $\xi_1,\ldots,\xi_m$ of $p(x)$ are continuous functions of its m coefficients. That is for a given $\varepsilon > 0$ there exists $\delta(\varepsilon)$ depending on $a_1,\ldots,a_m$, such that if $|b_i - a_i| < \delta(\varepsilon)$, $i = 1,\ldots,m$, it is possible to enumerate the zeros of $q(x) = x^m + \sum_{i=1}^{m} b_i x^{m-i}$ by $\eta_1,\ldots,\eta_m$, where $|\eta_i - \xi_i| < \varepsilon$, $i = 1,\ldots,m$. In particular the function

$$\rho(p) = \max_{1 \le i \le m} |\xi_i| \tag{1.28.16}$$

is a continuous function of $a_1,\ldots,a_m$.

Corollary 1.28.17. For $A \in M_n(C)$ the function $\rho(A)$ is a continuous function of A. That is, for a given $\varepsilon > 0$ there exists $\delta(\varepsilon, A) > 0$ such that

$$|\rho(B) - \rho(A)| < \varepsilon \quad \text{for} \quad \|B-A\| < \delta(\varepsilon,A) \quad .$$

Proof of Theorem 1.28.14. Suppose that (1.28.13) holds. Then by Theorem 1.27.9 $\rho(A) < 1$ for all $A \in$. Assume now that $\mathcal{A}$ is compact and $\rho(A) < 1$ for $A \in \mathcal{A}$. According to Corollary 1.28.17

$$\rho = \max_{A \in \mathcal{A}} \rho(A) = \rho(A^*) < 1, \; A^* \in \mathcal{A} \quad .$$

Recall that $(xI-A)^{-1} = (p_{ij}(x)/|xI-A|)$ where $p_{ij}(x)$ are the (i,j) cofactors of the matrix $(xI-A)$. So

$$\big||xI-A|\big| = |\Pi_{\lambda_i \in \sigma(A)}(x-\lambda_i)| \ge [|x| - \rho(A)]^n \ge (|x| - \rho)^n, \quad \text{for} \quad |x| > \rho \quad .$$

Let $\rho < \rho' < 1$. Then the above inequality and the expression of $(xI-A)^{-1}$ in terms of its cofactors imply that $\|(xI-A)^{-1}\| \le K$, $|x| = \rho'$, since $\mathcal{A}$ is a bounded set. Apply (1.28.8) to get

$$A^p = \frac{1}{2\pi i} \int_{|x|=\rho'} (xI-A)^{-1} x^p dx \quad . \tag{1.28.18}$$

Combine this equality and the estimate $\|(xI-A)^{-1})\| \leq K$ to get $\|A^p\| \leq K(\rho')^{p+1}$.

As $\rho' < 1$ we obtain (1.28.13). □

A generalization of the second part of Theorem 1.27.9 to compact sets $\mathcal{A} \subset M_n(\mathbf{C})$ is a far more complicated result and will be stated in the next chapter.

Problems

(1.28.19) Let $A \in M_n(\mathbf{C})$. Using (1.28.1) deduce that

(1.28.20) $$Z_{i(s_i-1)} = \lim_{x\to\lambda_i} (x-\lambda_i)^{s_i}(xI-A)^{-1}, \quad i = 1,\ldots,\ell \quad .$$

Put $R(x,A) = (r_{\mu\nu}(x))$. Using the definitions of Problem (1.26.13) show

(1.28.21) $$z_{\mu\nu}^{i(s-1)} = \lim_{x\to\lambda_i} (x-\lambda_i)^s r_{\mu\nu}(x), \quad s = \mathrm{index}_{\mu\nu}(\lambda_i) > 0 \quad .$$

(1.28.22) A set $\mathcal{A} \subseteq M_n(\mathbf{C})$ is called exponentially stable if

(1.28.23) $$\lim_{T\to\infty} \sup_{t\geq T, A\in\mathcal{A}} \|e^{At}\| = 0 \quad .$$

Prove that a compact $\mathcal{A}$ is exponentially stable if and only if $\mathrm{Re}\{\lambda\} < 0$, $\lambda \in \sigma(A)$ for any A.

(1.28.24) A matrix $B \in M_n(\mathbf{C})$ is called projection (idempotent) if $B^2 = B$. Let Γ be a set of simply connected rectifiable curves such that Γ form a boundry ∂D of an open domain D. Let $A \in M_n(\mathbf{C})$ and assume that $\Gamma \cap \sigma(A) = \emptyset$. Put

(1.28.25) $$P_D(A) = \frac{1}{2\pi i}\int_\Gamma R(x,A)dx, \quad A(D) = \frac{1}{2\pi i}\int_\Gamma R(x,A)x\,dx, \quad \sigma_D(A) = \sigma(A) \cap D \quad .$$

Show that $P_D(A)$ is a projection. $P_D(A)$ is called the projection of A on D and $A(D)$ is called the restriction of A to D. Prove

(1.28.26) $$P_D(A) = \sum_{\lambda_i \epsilon \sigma_D(A)} Z_{i0}, \quad A(D) = \sum_{\lambda_i \epsilon \sigma_D(A)} (\lambda_i Z_{i0} + Z_{i1}) \ .$$

Show that the rank of $P_D(A)$ is equal to the number of eigenvalues in D counted with their multiplicities. Prove that there exists a neighborhood of A such that $P_D(B)$ and $B(D)$ are analytic functions in B in this neighborhood. In particular, if D satisfies the assumptions of Theorem 1.28.9 then $P_D(A)$ is called the projection of A on λp. According to

(1.28.26) $P_D(A) = Z_{p0}$.

(1.28.7) Let $B = QAQ^{-1} \epsilon M_n(\mathbf{C})$. Assume that D satisfies the assumptions of Problem 1.28.24. Prove that $P_D(B) = QP_D(A)Q^{-1}$.

(1.28.28) Let $A \epsilon M_n(\mathbf{C})$ and assume that the minimal polynomial $\psi(x)$ of A is given by (1.18.20). Let

$$\mathbf{C}^n = U_1 \oplus \ldots \oplus U_\ell$$

where each U_p is an invariant subspace of A, i.e. $AU_p \subseteq U_p$, such that the minimal polynomial of the restriction of A to U_p is $(x-\lambda_p)^{s_p}$. Prove that

(1.28.29) $$U_p = Z_{p0}\mathbf{C}^n \ .$$

(It is enough to consider the case when A is in Jordan canonical form.)

(1.28.30) Let D_i, $i=1,\ldots,k$ satisfy the assumptions of Problem (1.28.4). Assume that $D_i \quad D_j = \phi$, for $i \neq j$, $i,j = 1,\ldots,k$. Prove that $P_{D_i}(A)\mathbf{C}^n \cap P_{D_j}(A)\mathbf{C}^n = [0]$ for $i \neq j$. Assume furthermore that $D_i \cap \sigma(A) \neq \phi$, $i=1,\ldots,k$, $\sigma(A) \subset \bigcup_{i=1}^{k} D_i$. Let

$P_{D_i}(A)\mathbf{C}^n = [y_1^{(i)},\ldots,y_{n_i}^{(i)}]$, $i=1,\ldots,k$, $X = (y_1^{(1)},\ldots,y_{n_1}^{(1)},\ldots,y_{n_k}^{(k)}) \epsilon M_n(\mathbf{C})$. Show that

(1.28.31) $$X^{-1}AX = \sum_{i=1}^{k} \oplus B_i, \quad \sigma(B_i) = D_i \cap \sigma(A), \ i = 1,\ldots,k \ .$$

(1.28.32) Let $A \epsilon M_n(\mathbf{C})$ and $\lambda_p \epsilon \sigma(A)$. Prove that

$$Z_{p0} = \prod_{\lambda_j \varepsilon \sigma_d(A), \lambda_j \neq \lambda_p} (A-\lambda_j I)^{s_j} / (\lambda_p - \lambda_j)^{s_j}, \quad s_j = \text{index } (\lambda_j) \quad ,$$

if index $(\lambda_p) = 1$. (Use the Jordan canonical form of A.)

1.29 A canonical form over H_A

Consider the space $M_n(C)$. Clearly $M_n(C)$ can be identified with C^{n^2}. As in Example 1.1.13 denote by H_A the set of analytic functions $f(B)$ where B ranges over the neighborhood $D(A,\rho)$ of the form (1.22.10) $(\rho = \rho(f) > 0)$. Thus the $B = (b_{ij})$ is an element in $M_n(H_A)$. Let $C \in M_n(H_A)$ and assume that $C = C(B)$ is similar to B over H_A. Then

$$(1.29.1) \qquad C(B) = X^{-1}(B)BX(B) \quad ,$$

where $X(B) \in M_n(H_A)$ and $|X(A)| \neq 0$. Our problem is to find a "simple" form for $C(B)$ (simpler than B itself!). Let M_A denote the quotient field of H_A - i.e., the set of meromorphic functions in the neighborhood of A. Thus if we let $X \in M_n(M_A)$ then we may take $C(B)$ to be $R(B)$ - the rational canonical form of B (1.16.6). According to Theorem 1.16.11 $R(B) \in M_n(H_A)$. However B and $R(B)$ are not similar over H_A in general and we shall give the necessary and sufficient conditions for $B \approx R(B)$. Thus, if $C(B) = (c_{ij}(B))$, we may ask how many independent variables are amont the functions $c_{ij}(B)$, $i,j = 1,\ldots,n$. For $X(B) = I$ the number of independent variables in $C(B) = B$ is obviously n^2. Therefore it is reasonable to define $C(B)$ to be simpler then B if $C(B)$ contains less independent variables than B. Given $C(B)$ we can view $C(B)$ as a map

$$(1.29.2) \qquad C(B) : D(A,\rho) \to M_n(C) \quad ,$$

where $D(A,\rho)$ is given by (1.22.10), for some $\rho > 0$. It is a well known result (e.g. (Gunning-Rossi [1965]) that the number of independent variables in $C(B)$ is equal to the rank of the Jacobian of $JC(B)$ of $C(B)$

(1.29.3) $$JC(B) = \frac{\partial}{\partial b_{ij}} \mu[C(B)] \quad , \quad i,j = 1,\dots,n \in M_{n^2}(H_A) \quad ,$$

where μ is the map given by (1.21.17).

Definition 1.29.4. Let $r(JC)$ be the rank of the Jacobian $JC(B)$, i.e. the number of linearly independent matrices in the set $\frac{\partial C(B)}{\partial b_{ij}}$, $i,j = 1,\dots,n$, over the field M_A. Let $r(JC(A))$ be the rank of the Jacobian $JC(A)$, i.e. the number of linearly independent matrices in the set $\frac{\partial C(A)}{\partial b_{ij}}$, $i,j,\dots,n$, over $\mathbf{C}$.

Lemma 1.29.5. Let $C(B)$ be a similar to B over H_A. Then

(1.29.6) $$r(JC(A)) > \nu(A,A) \quad .$$

Proof. Differentiating the relation $X^{-1}(B)X(B) = I$ with respect to b_{ij} we get

$$\frac{\partial X^{-1}(B)}{\partial b_{ij}} = -X^{-1} \frac{\partial X}{\partial b_{ij}} X^{-1} \quad .$$

So

(1.29.7) $$\frac{\partial C(B)}{b_{ij}} = X^{-1}\left[-\left(\frac{\partial X}{\partial b_{ij}} X^{-1}\right) B + B\left(\frac{\partial X}{\partial b_{ij}} X^{-1}\right) + E_{ij}\right] X \quad ,$$

where

(1.29.8) $$E_{ij} = (\delta_{i\alpha}\delta_{j\beta}) \in M_{mn}(\mathbf{C}), \quad i = 1,\dots,m, \ j = 1,\dots,n \quad ,$$

and $m = n$. So

$$X(A) \frac{\partial C(A)}{\partial b_{ij}} X^{-1}(A) = AP_{ij} - P_{ij}A + E_{ij}, P_{ij} = \frac{\partial X}{\partial b_{ij}}(A) \quad X^{-1}(A) \quad .$$

Clearly, $AP_{ij} - P_{ij}A$ is in range $(\hat{A})$ where

(1.29.9) $$\hat{A} = (I \otimes A - xA^t \otimes I) : M_n(C) \to M_n(C) \quad .$$

According to Definition 1.22.1

$$\dim \text{range } (\hat{A}) = r(A,A) \quad .$$

Let

(1.29.10) $$M_n(\mathbf{C}) = \text{range } (\hat{A}) \oplus [\Gamma_1, \ldots, \Gamma_{\nu(A,A)}] \quad .$$

As E_{ij}, $i,j,\ldots,n$ is basis in $M_n(\mathbf{C})$

$$\Gamma_p = \sum_{i,j=1}^{n} \alpha_{ij}^{(p)} E_{ij}, \; p = 1,\ldots,\nu(A,A) \quad .$$

Therefore

$$T_p = \sum_{i,j=1}^{n} \alpha_{ij}^{(p)} \frac{\partial C(A)}{\partial b_{ij}} = X^{-1}(A)[Q_p + \Gamma_p], \; Q_p \in \text{range } (\hat{A}) \quad .$$

According to (1.29.10) $T_1,\ldots,T_{\nu(A,A)}$ are linearly independent. That is the inequality (1.29.6) holds.

Clearly the rank of the Jacobian $JC(B)$ is at least the rank of $JC(A)$ so according to Lemma (1.29.6) $C(B)$ has to contain at least $\nu(A,A)$ independent variables. In fact this number can be achieved.

<u>Theorem</u> 1.29.11. <u>Let</u> $A \in M_n(\mathbf{C})$ <u>and assume that</u> $\Gamma_1,\ldots,\Gamma_{\nu(A,A)}$ <u>be any</u> $\nu(A,A)$ <u>matrices satisfying</u> (1.29.10). <u>Then for any non-singular matrix</u> $P \in M_n(\mathbf{C})$ <u>it is possible to find</u> $X(B) \in M_n(H_A)$, $|X(0)| \neq 0$, <u>such that</u>

$$(1.29.12) \qquad X^{-1}(B)BX(B) = P^{-1}AP + \sum_{i=1}^{\nu(A,A)} f_i(B)P\Gamma_i P^{-1},$$

$$f_i(B) \in H_A,\ f_i(A) = 0,\ i = 1,\ldots,\nu(A,A) \quad .$$

Proof. Let $R_1,\ldots,R_{r(A,A)}$ be a basis in range $(\hat{A})$. So there exist T_i such that

$$AT_j - T_jA = R_j,\ j = 1,\ldots,r(A,A) \quad .$$

Let us assume that $X(B)$ is of the form

$$(1.29.13) \qquad X(B)P^{-1} = I + \sum_{j=1}^{r(A,A)} g_j(B)T_j,\ g_j(A) = 0,\ j = 1,\ldots,r(A,A) \quad .$$

Thus the theorem will follow if we can prove that the system

$$(1.29.14) \qquad B[I + \sum_{j=1}^{r(A,A,)} g_j(B)T_j] = [I + \sum_{j=1}^{r(A,A)} g_j(B)T_j][A + \sum_{i=1}^{r(A,A)} f_i(B)\Gamma_i]$$

is solvable in some disc $D(A,\rho)$. This is a system of n^2 scalar equalities in n^2 unknowns $f_1,\ldots,f_{\nu(A,A)}, g_1,\ldots,g_{r(A,A)}$. Since $f_i(A) = g_j(A) = 0$, $i = 1,\ldots,\nu(A,A)$, $j = 1,\ldots,\nu(A,A)$ the above system is trivially satisfied for $B = A$. According to the implicit function theorem the above system is uniquely solvable in some neighborhood of A if the Jacobian of these n^2 equation is non-sinoular (e.g. Gunning-Rossi [1965]). Let $B = A + F$. Since all f_i, g_j are analytic in H_A we can expand these function in power series in the entries of $F = (f_{ij})$. Let $\alpha_i(F)$ and $\beta_j(F)$ be the linear terms in the expansions of f_i and g_j respectively. Then the Jacobian of the system (1.29.14) B non-singular if and only if the first terms α_i and β_j are uniquely determined by F. The linear part of (1.29.14) reduces to

$$F + \sum_{j=1}^{r(A,A)} \beta_j AT_j = \sum_{j=1}^{r(A,A)} \beta_j T_j A + \sum_{i=1}^{\nu(A,A)} \alpha_i \Gamma_i \quad .$$

That is

$$F = \sum_{j=1}^{r(A,A)} \beta_j R_j + \sum_{i=1}^{\nu(A,A)} \alpha_i \Gamma_i \quad .$$

In view of (1.29.10) $\alpha_1,\ldots,\alpha_{\nu(A,A)}$ and $\beta_1,\ldots,\beta_{r(A,A)}$ are determined uniquely. So (1.29.14) is solvable in $D(A,\rho)$ for some positive ρ.

This establishes the theorem.

Note that if $A = aI$ the form (1.29.12) is not simpler than B. Also by mapping $T \to P^{-1}TP$ we get

$$(1.29.15) \qquad M_n(C) = \text{range } (\widehat{P^{-1}AP}) \oplus [P^{-1}\Gamma_1 P,\ldots,P^{-1}\Gamma_{\nu(A,A)}P] \quad .$$

Next we consider the rational form $R(B)$ of B.

Lemma 1.29.16. *Let* $B \in M_n(H_A)$. *The rational canonical form of* B *over* $\mathcal{M}_A$ *is a companion matrix* $C(p) \in M_n(H_A)$, *where* $p(x) = |xI-B|$.

Proof. The rational canonical form of B is $C(p_1,\ldots,p_k)$ given by (1.16.6). We claim that $k = 1$. Otherwise $p(x)$ and $p'(x)$ have a common factor over $\mathcal{M}_A$ which in view of Theorem 1.4.13 implies that $p(x)$ and $p'(x)$ have a common factor over H_A. That is for any $B \in D(A,\rho)$ $|xI-B|$ has at least one multiple eigenvalue. Evidently, this cannot be true. Indeed consider $C = P^{-1}BP$ where $P \in M_n(C)$ and $J = P^{-1}AP$ is the Jordan canonical form of A. So $C \in D(J,\rho')$. Choose C to be an upper diagonal (this is possible since J is an upper diagonal matrix). So the eigenvalues of C are $c_{11},\ldots,c_{nn}$, and we can choose them to be distinct. Thus $p(x)$ and $p'(x)$ are co-prime over $\mathcal{M}_A$ and therefore $k = 1$. So $p_1(x) = |xI-C(p_1)| = |xI-B|$.

Theorem 1.29.17. *Let* $A \in M_n(C)$. *Then* $B \in M_n(H_A)$ *is similar to the companion matrix* $C(p)$, $p(x) = |xI-B|$, *over* H_A *if and only if* $\nu(A,A) = n$, *that is the minimal and the characteristic polynomials of* A *coincide*, i.e. A *is nonderogatory.*

Proof. Assume first that $C(B)$ in (1.29.1) can be chosen to be $C(p)$. Then for $B = A$ we get that A is similar to a companion matrix. According to Corollary 1.21.16 $\nu(A,A) = n$. Assume now that $\nu(A,A) = n$. According to (1.21.15) we must have that $i_1(x),\ldots,i_{n-1}(x)$ are equal to 1. That is, the minimal and the characteristic polynomials of A coincide. So A is similar to a companion matrix. From the form of (1.29.12) we can assume A is a companion matrix. Choose $\Gamma_i = F_{ni}$, $i = 1,\ldots,n$ when F_{ij} are defined in (1.29.8).

It is left to show that range $(\hat{A}) \cap [F_{n1},\ldots,F_{nn}] = [0]$. Suppose $\Gamma = \sum_{i=1}^{n} \alpha_i F_{ni} \in$ range $(\hat{A})$. According to Theorem 1.23.3 and Corollary 1.21.16 this assumption is equivalent to $tr(\Gamma A^k) = 0$, $k = 0,1,\ldots,n-1$. Let $\alpha = (\alpha_1,\ldots,\alpha_n)$. Since the first $n-1$ rows of Γ are zero we have

$$0 = tr(\Gamma A^k) = \alpha A^k \varepsilon_n, \quad \varepsilon_n = (\delta_{n1},\ldots,\delta_{nn})^t.$$

For $k = 0$ the above equality implies $\alpha_n = 0$. Suppose we already prved that these equalities for $k = 0,\ldots,\ell$ imply that $\alpha_n = \ldots = \alpha_{n-\ell} = 0$. Consider the equality $tr(\Gamma A^{\ell+1}) = 0$. Use Problem (1.17.20) to deduce

$$A^{\ell+1}\varepsilon_n = \varepsilon_{n-\ell-1} + \sum_{i=0}^{\ell} f_{(\ell+1),i}\varepsilon_{n-i} .$$

So $tr(\Gamma A^{\ell+1}) = \alpha_{n-\ell-1}$ as $\alpha_n = \ldots = \alpha_{n-\ell} = 0$. Thus $\alpha_{n-\ell-1} = 0$ and we proved that range $(\hat{A}) \cap [F_{n1},\ldots,F_{nn}] = [0]$.

According to Theorem (1.29.11)

$$C(B) = X^{-1}(B)BX(B) = A + \sum_{i=1}^{n} f_i(B)F_{ni} .$$

So $C(B)$ is a companion matrix. As $|xI-C(B)| = |xI-B|$, $C(B) = C(p)$. □

For the next theorem we need the following lemma which is an immediate consequence of Problem 1.23.13.

Lemma 1.29.18. *Let* $A_i \in M_{n_i}(\mathbb{C})$, $i = 1,2$ *and assume*

$$M_{n_i}(\mathbb{C}) = \text{range}\ (\hat{A_i}) \oplus \sum_{j=1}^{\nu(A_i,A_i)} [\Gamma_j^{(i)}],\ i = 1,2.$$

Suppose that A_1 *and* A_2 *do not have a common eigenvalue. Then*

$$M_{n_1}(\mathbb{C}) = \text{range}\ (\widehat{A_1 \oplus A_2}) \oplus \sum_{j=1}^{\nu(A_1,A_1)} \oplus [\Gamma_j^{(1)} \oplus 0] \oplus \sum_{j=1}^{\nu(A_2,A_2)} \oplus [0 \oplus \Gamma_j^{(2)}] \ .$$

Theorem 1.29.19. *Let* $A \in M_n(\mathbb{C})$. *Assume that* $\sigma_d(A)$ *consists of* ℓ *distinct eigenvalues* $\lambda_1,\ldots,\lambda_\ell$, *where* n_i *is the multiplicity of* λ_i, $i = 1,\ldots,\ell$. *Then* B *is similar over* H_A *to the following matrix*

$$C(B) = \sum_{i=1}^{\ell} \oplus C_i(B),\ C_i(B) \in M_{n_i}(H_A),\ [\lambda_i I(n_i) - C_i(A)]^{n_i} = 0\ ,\quad i = 1,\ldots,\ell\ . \tag{1.29.20}$$

□

Moreover $C(A)$ *is the Jordan canonical form of* A.

Proof. Choose P in the equality (1.29.12) such that $P^{-1}AP$ is the Jordan canonical form $\sum_{i=1}^{\ell} \oplus \Gamma_j^{(i)}$. Then the equality (1.29.12) yields the theorem.

Problems

(1.29.21) Let $A = \sum_{i=1}^{k} \oplus H(n_i)$, $n = \sum_{i=1}^{k} n_i$. Partition any $B \in M_n(\mathbb{C})$ conformally with A, $B = (B_{ij})$, $B_{ij} \in M_{n_i n_j}(\mathbb{C})$, $i,j = 1,\ldots,k$. Using the results of Theorems 1.21.9 and 1.23.3 prove that the matrices

$$\Gamma_{(\alpha,\beta,\gamma)} = (\Gamma_{ij}^{(\alpha,\beta,\gamma)}),\ \Gamma_{ij}^{(\alpha,\beta,\gamma)} = 0 \text{ if } (\alpha,\beta) \neq (i,j),\ \Gamma_{\alpha\beta}^{(\alpha,\beta,\gamma)} = F_{n_\alpha}^{\gamma},$$

$$\gamma = 1,\ldots,\min(n_\alpha,n_\beta),\ \alpha,\beta = 1,\ldots,k$$

satisfy (1.29.10).

(1.29.22) Let A be a matrix given by (1.21.3). Use Theorem 1.29.19 and Problem 1.29.21 to find a set of matrices $\Gamma_1,\ldots,\Gamma_{\nu(A,A)}$ which satisfy (1.29.10).

(1.29.23) Let $A \in M_n(\mathbb{C})$ and assume that λ_i is a simple root of the characteristic polynomial of A. Use Theorem 1.21.9 to prove the existence of $\lambda(B) \in H_A$ such that $\lambda(B)$ is an eigenvalue of B and $\lambda(A) = \lambda_i$.

(1.29.24.) Let A satisfy the assumptions of Theorem 1.29.19. Denote by D_p a domain satisfying the assumptions of Theorem 1.28.9, $p = 1,\ldots,\ell$. Let $P_k(B)$ be a projection of $B \in M_n(H_A)$ on D_k, $k = 1,\ldots,\ell$. According to Problem 1.28.24 $P_k(B) \in M_n(H_A)$, $k = 1,\ldots,\ell$. Let $P_k(A)\mathbb{C} = [x^{k1},\ldots,x^{kn_k}]$, $k = 1,\ldots,\ell$, $B \in D(A,\rho)$ where ρ is some positive number. Let $X(B)$ be formed by the columns $P_k(B)x^{k1},\ldots,P_k(B)x^{kn_k}$, $k = 1,\ldots,\ell$. Prove that $C(B)$ given by (1.29.1) satisfies (1.29.20). This gives another proof to Theorem 1.29.19.

1.30. Analytic, pointwise and rational similarity.

Definition 1.30.1. Let $\Omega \subseteq \mathbb{C}^m$ and $A,B \in M_n(H(\Omega))$. Then

(i) A and B are called analytically similar $(A \underset{a}{\approx} B)$ if A and B are similar over $H(\Omega)$,

(ii) A and B are called pointwise similar if $A(x)$ and $B(x)$ are similar over $\mathbb{C}$ for all $x \in \Omega_0$, where Ω_0 is some open domain $\Omega_0 \supseteq \Omega$,

(iii) A and B are called rationally similar $(A \underset{r}{\approx} B)$ if A and B are similar over the field of meromorphic functions $M(\Omega)$.

Theorem 1.30.2. Let $\Omega \subseteq \mathbb{C}^m$ and assume that $A,B \in M_n(H(\Omega))$. Then the following applications hold. $A \underset{a}{\approx} B \Rightarrow A \underset{p}{\approx} B \Rightarrow A \underset{r}{\approx} B$.

Proof. Suppose that

$$B(x) = P^{-1}(x)A(x)P(x) \quad , \tag{1.30.3}$$

where $P,P^{-1} \in M_n(H(\Omega))$. Let $x_0 \in \Omega$. Then (1.30.3) is holding in some neighborhood of x_0. So $A \underset{p}{\approx} B$. Assume now that $A \underset{p}{\approx} B$. Let $C(p_1,\ldots,p_k)$ and $C(q_1,\ldots,q_e)$ be the rational canonical forms of A and B respectively over the field of meromorphic functions $M(\Omega)$. $C(p_1,\ldots,p_k) = S(x)^{-1}A(x)S(x)$, $C(q_1,\ldots,q_e) = T(x)^{-1}B(x)T(x)$, $S(x),T(x) \in M_n(H(\Omega))$, $|S(x)| \not\equiv 0$, $|T(x)| \not\equiv 0$. According to Theorem 1.16.11 $C(p_1,\ldots,p_k)$, $C(q_1,\ldots,q_\ell) \in M_n(H(\Omega))$. Let $\Omega \subseteq \Omega_0$ be an open connected domain such that $A,B,S,T \in M_n(H(\Omega))$ and $A(x)$ and $B(x)$ are similar over $\mathbb{C}$ for any $x \in \Omega_0$. Let $x_0 \in \Omega_0$ such that $|S(x_0)T(x_0)| \neq 0$. Then for all $x \in D(x_0,\rho)$

$$C(p_1,\ldots,p_k)(x) = C(q_1,\ldots,q_\ell)(x) \quad .$$

Now the analyticity of $C(p_1,\ldots,p_k)$ and $C(q_1,\ldots,q_\ell)$ imply that these matrices are identical in $H(\Omega)$. That is $A \underset{r}{\approx} B$.

□

Assume that $A \underset{a}{\approx} B$. Then according to Lemma 1.22.8 the three matrices

$$(1.30.4) \qquad I \otimes A(x) - A^t(x) \otimes I, \ I \otimes A(x) - B^t(x) \otimes I, \ I \otimes B(x) - B^t(x) \otimes I$$

are equivalent over $H(\Omega)$. Theorem 1.22.7 yields

Theorem 1.30.5. Let $A,B \in M_n(H(\Omega))$. Assume that the three matrices in (1.30.4) are equivalent over $H(\Omega)$. Then $A \underset{p}{\approx} B$.

In case that $\Omega \subseteq \mathbf{C}$, $H(\Omega)$ is EDD (see Section 1.5). So we can determine when these three matrices are equivalent.

The problem of finding a canonical form of $A \in M_n(H(\Omega))$ under the analytic similarity is a very hard problem which will be discussed in the next few sections for the ring of local analytic functions in one variable. In what follows we determine when A is analytically similar to its rational canonical form over H_ζ -the ring of local analytic functions in the neighborhood of $\zeta \in \mathbf{C}^m$.

For $A,B \in M_n(H(\Omega))$ denote by $r(A,B)$ and $\nu(A,B)$ the rank and the nullity of the matrix $C = I \otimes A - B^t \otimes I$ over the field $M(\Omega)$. By $r(A(x),B(x))$ and $\nu(A(x),B(x))$ denote the rank and the nullity of $C(x)$ over $\mathbf{C}$. As the rank of $C(x)$ is determined by the largest size of a non-vanishing minor, we clearly have

$$(1.30.6) \qquad r(A(\zeta),B(\zeta)) \le r(A(x),B(x)) \le r(A,B); \ \nu(A,B) \le \nu(A(x),B(x)) \le \nu(A(\zeta),B(\zeta)),$$
$$x \in D(\zeta,\rho)$$

for some positive ρ. Moreover for any $\rho > 0$ there exist at least one $x_0 \in D(\zeta,\rho)$ such that

$$(1.30.7) \qquad r(A(x_0),B(x_0)) = r(A,B), \ \nu(A(x_0),B(x_0)) = \nu(A,B) \ .$$

Theorem 1.30.8. *Let* $A \in M_n(H_\zeta)$. *Assume that* $C(p_1,\ldots,p_k)$ *is the rational canonical form of* A *over* M_ζ *and* $C(\sigma_1,\ldots,\sigma_\ell)$ *is the rational canonical form of* $A(\zeta)$ *over* $\mathbb{C}$. *That is* $p_i = p_i(\lambda,x)$ *and* $\sigma_j = \sigma_j(\lambda)$ *are normalized polynomials in* λ *belonging to* $H_\zeta[\lambda]$ *and* $\mathbb{C}[\lambda]$ *respectively for* $i = 1,\ldots,k$, $j = 1,\ldots,\ell$. *Then* (i) $\ell \geq k$ *and*

(ii) $\prod_{i=0}^{m} \sigma_{\ell-i}(\lambda) \mid \prod_{i=0}^{m} p_{k-i}(\lambda,\zeta)$ *for* $m = 0,1,\ldots,k-1$. *Moreover* $\ell = k$ *and* $p_i(\lambda,\zeta) = \sigma_i(\zeta)$ *for* $i = 1,\ldots,k$ *if and only if*

$$r(A(\zeta),A(\zeta)) = r(A,A),\ \nu(A(\zeta),A(\zeta)) = \nu(A,A) \quad , \tag{1.30.9}$$

which is equivalent to the condition

$$r(A(\zeta),A(\zeta)) = r(A(x),A(x)),\ \nu(A(\zeta),A(\zeta)) = \nu(A(x),A(x)),\ x \in D(\zeta,\rho) \tag{1.30.10}$$

for some positive ρ.

Proof. Let

$$u_{n-k+i}(\lambda,x) = \prod_{\alpha=1}^{i} p_\alpha(\lambda,x),\ v_{n-\ell+j}(\lambda) = \prod_{\beta=1}^{j} \sigma_\beta(\lambda),\ i = 1,\ldots,k,\ j = 1,\ldots,\ell,$$
$$u_\alpha(\lambda,x) = v_\beta(\lambda) = 1, \quad \text{for} \quad \alpha \leq n-k,\ \beta \leq n-\ell \quad .$$

So $u_i(\lambda,x)$ and $v_i(\lambda)$ are the g.c.d. of all minors of order i of the matrices $\lambda I - A$ and $\lambda I - A(\zeta)$ over the rings $M_\zeta[\lambda]$ and $\mathbb{C}[\lambda]$ respectively. (See Section 1.16.) As $u_i(\lambda,x) \in H_\zeta[\lambda]$ it is clear that $u_i(\lambda,\zeta)$ divides all minors of $\lambda I - A(\zeta)$ of order i. So $u_i(\lambda,\zeta) \mid v_i(\lambda)$, $i = 1,\ldots,n$. Since $v_{n-\ell}(\lambda) = 1$ we must have that $u_{n-\ell}(\lambda,x) = 1$. That is $k \leq \ell$. Also

$$u_n(\lambda,x) = |\lambda I - A(x)|,\ v_n(\lambda) = |\lambda I - A(\zeta)| \quad .$$

Whence $u_n(\lambda,\zeta) = v_n(\lambda)$. Therefore $\frac{v_n(\lambda)}{v_i(\lambda)} \Big| \frac{u_n(\lambda,\zeta)}{u(\lambda,\zeta)}$. This establishes claim (ii) of the theorem. Clearly if $C(q_1,\ldots,q_\ell) = C(p_1,\ldots,p_k)(\zeta)$ then $k = \ell$ and $p_i(\lambda,\zeta) = q_i(\lambda)$, $i = 1,\ldots,\ell$. Assume now that (1.30.9) holds. According to (1.21.15)

$$\nu(A,A) = \sum_{i=1}^{k} (2i-1)\deg p_{k-i+1}(\lambda,x),\ \nu(A(\zeta),A(\zeta)) = \sum_{j=1}^{\ell} (2j-1)\deg q_{\ell-j+1}(\lambda) \quad .$$

Note that the degrees of the invariant polynomials of $(\lambda I - A)$ and $(\lambda I - A(\zeta))$ are satisfying the assumptions of Problem 1.30.13. From the results of Problem 1.30.13 it follows that the second equality in (1.30.9) holds if and only if $k = \ell$ and $\deg p_i(\lambda,x) = \deg q_i(\lambda)$, $i = 1,\ldots,k$. Since $p_i(\lambda,x)$ and $q_i(x)$ are normalized polynomials in λ the conditions (ii) yield $p_i(\lambda,\zeta) = q_i(\lambda)$, $i = 1,\ldots,k$. Finally (1.30.6)-(1.30.7) imply the equivalence of the conditions of (1.30.9) and (1.30.10).

$\square$

<u>Corollary</u> 1.30.11. <u>Let</u> $A \in M_n(H_\zeta)$. <u>Assume that</u> (1.30.10) <u>holds. Then</u> $A \underset{a}{\approx} B$ <u>if and only if</u> $A \underset{p}{\approx} B$.

<u>Proof</u>. According to Theorem 1.30.2 it is enough to show that $A \underset{p}{\approx} B$ implies $A \underset{a}{\approx} B$. Since A satisfies (1.30.10) then the assumption that $A \underset{p}{\approx} B$ implies that B also satisfies (1.30.10). According to Theorem 1.30.8 A and B are analytically similar to their rational canonical form. From Theorem 1.30.2 it follows that A and B have the same rational canonical form.

$\square$

<u>Problems</u>

(1.30.12) Let

$$A(x) = \begin{pmatrix} 0 & x \\ 0 & 0 \end{pmatrix}, \quad B(x) = \begin{pmatrix} 0 & x^2 \\ 0 & 0 \end{pmatrix} \quad .$$

Show that $A(x)$ and $B(x)$ are rationally similar over the ring of rational functions to $H(2)$. Prove that $A \underset{p}{\not\approx} H(2)$, $B \underset{p}{\not\approx} H(2)$. Show that $A \underset{p}{\approx} B$ over $C[x]$. Prove that the matrices given in (1.30.4) are not equivalent over $C[x]$. That is $A \underset{a}{\not\approx} B(x)$.

(1.30.13) Let n be a positive integer and assume that $\{m_i\}^n$ and $\{\ell_i\}^n$ are two sequences of non-negative integers satisfying $0 \leq m_n \leq m_{n-1} \leq \ldots \leq m_1$,

$0 \leq \ell_n \leq \ell_{n-1} \leq \ldots \leq \ell_1$, $\sum_{i=1}^{k} \ell_i \leq \sum_{i=1}^{k} m_i$, $k = 1,\ldots,n-1$, $\sum_{i=1}^{n} \ell_i = \sum_{i=1}^{n} m_i = n$. Prove (by induction) that

$$\sum_{i=1}^{n} (2i-1)m_i \leq \sum_{i=1}^{n} (2i-1)\ell_i$$

and the equality holds if and only if $m_i = \ell_i$, $i = 1,\ldots,n$.

(1.30.14) Let $\zeta_n \in C$, $n = 1,2,\ldots,$. Assume that $\lim_{n\to\infty} \zeta_n = \zeta$. Suppose that $\Omega \subseteq C$, such that $\zeta_n \in C$, $n = 1,2,\ldots,$, $\zeta \in \Omega$. Using the fact that if $\in H(\Omega)$ satisfies $(\zeta_n) = 0$, $n = 1,\ldots,$ implies $\equiv 0$ prove that for $A,B \in M_n(\Omega)$ the assumption $A(\zeta_n) \approx B(\zeta_n)$, $n = 1,2,\ldots,$ implies that $A \underset{r}{\approx} B$.

1.31. A global splitting

From this section through the end of the chapter we shall restrict ourselves to the matrices whose entries are analytic in domains $\Omega \subseteq \mathbf{C}$. In what follows we give a global version of Theorem 1.29.19.

Theorem 1.31.1. *Let* $A(x) \in M_n(H(\Omega))$, *where* Ω *is a connected set in* $\mathbf{C}$. *Suppose that*

$$(1.31.2) \qquad |\lambda I - A(x)| = \varphi_1(\lambda,x)\varphi_2(\lambda,x) \quad ,$$

where $\varphi_1(\lambda,x)$ *and* $\varphi_2(\lambda,x)$ *are two nontrivial normalized polynomials in* $H(\Omega)[x]$ *of the degrees* n_1 *and* n_2 *respectively such that* $(\varphi_1(\lambda,x_0),\varphi_2(\lambda,x_0)) = 1$ *for any* $x_0 \in \Omega$. *Then there exists a unimodular matrix* $X(x)$ *such that*

$$(1.31.3) \qquad \begin{aligned} &X^{-1}(x)A(x)X(x) = C_1(x) \oplus C_2(x),\ X(x),X^{-1}(x) \in M_n(H(\Omega)) \quad , \\ &C_i(x) \in M_{n_i}(H(\Omega)),\ |\lambda I - C_i(x)| = \varphi_i(\lambda,x),\ i = 1,2 \quad . \end{aligned}$$

Proof. Let $P_i(x)$ be the projection of $A(x)$ on the eigenvalues of $A(x)$ satisfying $\varphi_i(\lambda,x) = 0$. Since $(\varphi_1(\lambda,x_0), \varphi_2(\lambda,x_0)) = 1$ clearly $P_i(x) \in M_n(H(\Omega))$ for $i = 1,2$. (See Problem 1.28.24.) Also for any x_0 the rank of $P_i(x_0)$ is n_i. Since $H(\Omega)$ is EDD each $P_i(x)$ can be brought to the Smith normal form

$$P_i(x) = U_i(x)\ \mathrm{diag}(\varepsilon_1^{(i)}(x),\dots,\varepsilon_{n_i}^{(i)}(x),0,\dots,0)V_i(x),$$
$$U_i,V_i,U_i^{-1},V_i^{-1} \in M_n(H(\Omega)) \quad .$$

As $r(P_i(x_0)) = n_i$ for any $x_0 \in \Omega$ we immediately deduce that $\varepsilon_j^{(i)}(x) = 1,\ j = 1,\dots,n_i$, $i = 1,2$. Let $u_1^{(i)}(x),\dots,u_n^{(i)}(x)$ be the columns of $U_1(x)$, $i = 1,2$. Since $V_i^{-1}(x) \in M_n(H(\Omega))$ we get that

$$P_i(x)\mathbf{C}^n = [u_1^{(i)}(x),\ldots,u_{n_i}^{(i)}(x)] \quad . \tag{1.31.4}$$

for any $x \in \Omega$. Put

$$X(x) = (u_1^{(1)}(x),\ldots,u_{n_1}^{(1)}(x),\ u_1^{(2)}(x),\ldots,u_{n_2}^{(2)}(x)) \in M_n(H(\Omega)) \quad .$$

According to Problem (1.28.30) $|X(x_0)| \neq 0$ for any $x_0 \in \Omega$. So $X^{-1}(x) \in M_n(H(\Omega))$. Now (1.31.3) follows from (1.28.31).

□

1.32. *First variation of a geometrically simple eigenvalue.*

Theorem 1.32.1. *Let* $A(x)$ *be a continuous family of* $n \times n$ *complex values matrices for* $|x - x_0| < \delta$, *where the parameter* x *is either real or complex. Suppose that*

$$A(x) = A_0 + (x - x_0)A_1 + |x - x_0|o(1) \quad . \tag{1.32.2}$$

Assume furthermore that λ *is a geometrically simple eigenvalue of multiplicity* m. *Let* $x^1,\ldots,x^m$ *and* $y^1,\ldots,y^m$ *be the eigenvectors of* A_0 *and* A_0^t *corresponding to* λ *which form a biorthonormal system* $(y^i)^t x^j = \delta_{ij}$, $i,j = 1,\ldots,m$. *Then it is possible to enumerate the eigenvalues of* $A(x)$ *by* $\lambda_i(x)$, $i = 1,\ldots,n$, *such that*

$$\lambda_i(x) = \lambda + (x - x_0)\mu_i + |x - x_0|o(1), \; i = 1,\ldots,m \quad , \tag{1.32.3}$$

where $\mu_1,\ldots,\mu_m$ *are the eigenvalues of the matrix*

$$S = (s_{ij}) \in M_m(\mathbf{C}), \; s_{ij} = (y^i)^t A_1 x^j, \; i,j = 1,\ldots,m \quad . \tag{1.32.4}$$

Proof. By considering the matrix $P^{-1}A(x)P$, where $P \in M_n(\mathbf{C})$ we can assume that A_0 is in the Jordan canonical form such that the first m diagonal entries of A_0 are λ. From the proofs of Theorems 1.29.11 and 1.29.19 we have the existence of

$$X(B) = I + Z(B), \; Z \in M_n(H_0), \; Z(0) = 0$$

such that

$$X^{-1}(B)(A_0 + B)X(B) = \sum_{i=1}^{\ell} \oplus C_i(B), \; C_1(0) = \lambda I(m) \quad . \tag{1.32.5}$$

Substituting

$$B(x) = A(x) - A_0 = (x-x_0)A_1 + |x-x_0|o(1),$$
$$X(x) = X(B(x)) = I + (x-x_0)X_1 + |x-x_0|o(1)$$

we get

$$C(x) = X^{-1}(x)A(x)X(x) = A_0 + (x-x_0)(A_1 + A_0X_1 - X_1A_0) + |x-x_0|o(1) \quad .$$

According to (1.32.5) $\lambda_1(x),\dots,\lambda_m(x)$ are the eigenvalues of $C_1(B(x))$. As $C_1(B(x_0)) = \lambda I(m)$ by considering the matrix $[C_1(B(x)) - \lambda I(m)]/(x - x_0)$ we deduce that $(\lambda_i(x) - \lambda)/(x - x_0)$ are continuous functions at $x = x_0$. Also

$$[C_1(B(x)) - \lambda I(m)]/(x - x_0) = ((\eta^i)^t(A_1 + A_0X_1 - X_1A_0)\xi^j) + o(1) \quad ,$$

where $\xi^i = y^i = (\delta_{i1},\dots,\delta_{in}), i = 1,\dots,m$. Now note that since ξ^i and η^i are the eigenvectors of A_0 and A_o^t respectively corresponding to λ for $i = 1,\dots,m$ $(\eta^i)^t(A_0X_1 - X_1A_0)\xi^j = 0$ for $1 \le i,j \le m$. This establishes the result for a particular choice of eigenvectors $\xi^1,\dots,\xi^m$ and $\eta^1,\dots,\eta^m$. It is left to note that any other choice of the eigenvectors $x^1,\dots,x^m$ and $y^1,\dots,y^m$ which form a biorthogonal system amounts to a new matrix S^1 which is similar to S. In particular S and S^1 have the same eigenvalues.

□

<u>Problems</u>

(1.32.6) Let

$$A(x) = \begin{pmatrix} 0 & 1 \\ x & 0 \end{pmatrix} \quad .$$

Find the eigenvalues and the eigenvectors of $A(x)$ in terms of $\sqrt{x}$. Show that (1.32.3) does not apply for $x_0 = 0$ in this case. Let $B(x) = A(x^2)$. Verify that (1.32.3) holds for $x_0 = 0$ even though $\lambda = 0$ is not geometrically simple eigenvalue for $B(0)$.

1.33. Analytic similarity over H_0.

Let $A,B \in M_n(H_0)$. That is

$$A(x) = \sum_{k=0}^{\infty} A_k x^k, \quad |x| < r(A)$$

(1.33.1)

$$B(x) = \sum_{k=0}^{\infty} B_k x^k, \quad |x| < r(B) \quad .$$

Definition 1.33.2. Let $A,B \in M_n(H_0)$. Denote by $\eta(A,B)$ the index and $\kappa_p(A,B)$ the number of local invariant polynomials of degree p of the matrix $I \otimes A(x) - B^t(x) \otimes I$.

Theorem 1.33.3. Let $A,B \in M_n(H_0)$. Then A and B are analytically similar over H_0 if and only if A and B are rationally similar and there exist $\eta(A,A)+1$ matrices $T_0,\ldots,T_\eta \in M_n(\mathbb{C})$, $(\eta = \eta(A,A))$ such that $|T_0| \neq 0$ and

$$\sum_{i=0}^{k} (A_i T_{k-i} - T_{k-i} B_i) = 0, \; k = 0,1,\ldots,\eta(A,A) \quad . \tag{1.33.4}$$

Proof. The necessary part of the theorem is obvious. Assume now that $A(x) \underset{r}{\approx} B(x)$ and the matrices $T_0,\ldots,T_\eta$ satisfy (1.33.4) and T_0 is non-singular. Put

$$C(x) = T(x)B(x)T^{-1}(x), \; T(x) = \sum_{k=0}^{\eta} T_k x^k \quad .$$

As $|T_0| \neq 0$, $B(x) \underset{a}{\approx} C(x)$, so $A(x) \underset{r}{\approx} C(x)$. This in particular means that $r(A,A) = r(A,C)$. Also (1.33.4) is equivalent to $A(x) - C(x) = x^{\eta+1}O(1)$. Hence

$$(I \otimes A(x) - A^t(x) \otimes I) - (I \otimes A(x) - C^t(x) \otimes I) = x^{\eta+1}O(1) \quad .$$

In view of Lemma 1.13.4 the matrices $I \otimes A(x) - A^t(x) \otimes I$, $I \otimes A(x) - C^t(x) \otimes I$ are equivalent over H_0. In particular $\eta(A,A) = \eta(A,C)$. Also $I,0,\ldots,0$ satisfy the system (1.33.4) where $B_i = C_i$, $i = 0,1,\ldots,\eta$. According to Theorem 1.13.14 there exists $P(x) \in M_n(H_0)$ such that

$$A(x)P(x) - P(x)C(x) = 0, \quad P(0) = I \quad .$$

This shows that $A(x) \underset{a}{\approx} C(x)$. By the constriction $C(x) \underset{a}{\approx} B(x)$, so $A(x) \underset{a}{\approx} B(x)$.

Note that if $\eta(A,A) = 0$ then the assumptions of Theorem 1.33.3 are equivalent $A(x) \underset{p}{\approx} B(x)$. Then the implication that $A(x) \underset{a}{\approx} B(x)$ follows from Corollary 1.30.11.

Suppose that the characteristic polynomials of $A(x)$ splits over H_0. That is

$$(1.33.5) \qquad |\lambda I - A(x)| = \prod_{i=1}^{n} (\lambda - \lambda_i(x)), \ \lambda_i(x) \in H_0, \ i = 1,\ldots,n \quad .$$

As H_0 is **ED** according to Theorem 1.18.5 $A(x)$ is similar to an upper triangular matrix. Using Theorem 1.29.19 and Theorem 1.18.5 we get that $A(x)$ is analytically similar to

$$(1.33.6) \qquad \begin{aligned} &C(x) = \sum_{i=1}^{\ell} \oplus C_i(x), \ C_i(x) \in M_{n_i}(H_0), \ (\alpha_i I(n_i) - C_i(0))^{n_i} = 0 \quad , \\ &\qquad \alpha_i = \lambda_i(0), \quad \alpha_i \neq \alpha_j \quad \text{for} \quad i \neq j, \ i,j = 1,\ldots,\ell \end{aligned}$$

and each $C_i(x)$ is an upper triangular matrix. In this case we can be more specific above the form of the upper triangular matrix.

Theorem 1.33.7. Let $A(x) \in M_n(H_0)$. Assume that the characteristic polynomial of $A(x)$ splits in H_0. Then $A(x)$ is analytically similar to a block diagonal matrix $C(x)$ of the form (1.33.6) such that each $C_i(x)$ is an upper triangular matrix whose off-diagonal entries are polynomials in x. Moreover, the degree of each polynomial entry above the diagonal in the matrix $C_i(x)$ do not exceed $\eta(C_i,C_i)$, $i = 1,\ldots,\ell$.

Proof. Clearly, in view of Theorem 1.29.19 we may assure that $\ell = 1$. That is $A(0)$ has one eigenvalue α_0. Furthermore by considering $A(x) - \alpha_0 I$ we may assume that $A(0)$ is nilpotent. Also by Theorem 1.18.15 we may assume that $A(x)$ is already in the upper triangular form. Suppose now that in addition to all the above assumptions $A(x)$ is nilpotent. Define

$$X_k = \{y \mid A^k y = 0,\ y \in M_{n1}(H_0)\},\ k = 0,1,\ldots, \quad .$$

So

$$[0] = X_0 \subsetneqq X_1 \subsetneqq X_2 \subsetneqq \cdots \subsetneqq X_p = M_{n1}(H_0) \quad .$$

According to Theorem 1.11.12 it is easy to show the existence of a basis $y^1(x),\ldots,y^n(x)$ in $M_n(H_0)$ such that $y^1(x),\ldots,y^{\psi_k}(x)$ is a basis in X_k. As $A(x)X_{k+1} \subseteq X_k$ we have

$$A(x)y^j = \sum_{i=1}^{\psi_k} g_{ij}(x)y^i(x),\ \psi_k < j \le \psi_{k+1} \quad .$$

Define $g_{ij}(x) = 0$ for $i > \psi_k$ and $\psi_k < j \le \psi_{k+1}$. Put

$$G(x) = (g_{ij}(x))_1^n,\ T(x) = (y^1(x),\ldots,y^n(x)) \in M_n(H_0) \quad .$$

Since $y^1(x),\ldots,y^n(x)$ is a basis $T^{-1}(x) \in M_n(H_0)$. So

$$G(x) = T^{-1}(x)A(x)T(x),\ s = \eta(A,A) = \eta(G,G) \quad .$$

Put

$$G(x) = \sum_{j=0}^{\infty} G_j x^j,\ G^{(k)} = \sum_{j=0}^{k} G_j x^j,\ k = 0,1,\ldots, \quad .$$

We claim that $G^{(s)}(x) \underset{a}{\approx} G(x)$. First note that

$$(I \otimes G(x) - G^t(x) \otimes I) - \{I \otimes G^{(s)}(x) - [G^{(s)}(x)]^t \otimes I\} = x^{(s+1)}O(1) \quad .$$

From Lemma 1.13.4 it follows that the matrices $I \otimes G(x) - G^t(x) \otimes I$, $I \otimes G^{(s)}(x) - [G^{(s)}(x)]^t \otimes I$ have the same local invariant polynomials up to the degrees. So

$$r(G,G) \leq r(G^{(s)},G^{(s)})$$

which is equivalent to the inequality

$$(1.33.8) \qquad \nu(G^{(s)},G^{(s)}) \leq \nu(G,G) \quad .$$

Let

$$Y_k = \{y \mid y = (y_1,\ldots,y_n)^t,\ y_j = 0 \ \text{ for } \ j > \psi_k\} \quad .$$

Clearly if $g_{ij}(x) = 0$ then the (i,j) entry of $G^{(s)}(x)$ is also equal to zero. By the construction $g_{ij}(x) = 0$ for $i > \psi_k$ and $\psi_k < j \leq \psi_{k+1}$. So $G^{(s)}(x)Y_{k+1} \subseteq Y_k$, $k = 0,\ldots,p-1$. By Theorem 1.24.1

$$(1.33.9) \qquad \nu(G(x_0),G(x_0)) \leq \nu(G^{(s)}(x_0),G^{(s)}(x_0))$$

for all x_0 in the neighborhood of the origin. So

$$\nu(G,G) \leq \nu(G^{(s)},G^{(s)}).$$

This proves that the equality sign holds in (1.33.8) which in return implies the equality sign in (1.33.9) for $0 \leq |x_0| < \rho$. By Theorem 1.24.1 $G(x_0) \approx G^{(s)}(x_0)$ for $0 \leq |x_0| < \rho$. Using Theorem 1.30.2 we deduce that $G \underset{r}{\approx} G^{(s)}$. As

$$G(x)I - IG^{(s)}(x) = x^{s+1}O(1)$$

Theorem 1.33.3 implies that $G(x) \underset{a}{\approx} G^{(s)}(x)$. This establishes the theorem in case that $A(x)$ is a nilpotent matrix. Next consider the case where $A(x)$ is an upper triangular matrix whose diagonal entries (the eigenvalues of $A(x)$) are arranged in the following order

$$\begin{aligned} & A(x) = (A_{ij}(x))^{\ell}, \; A_{ij}(x) \in M_{n_i n_j}(H_0) \;, \\ (1.33.10) \qquad & A_{ij}(x) = 0 \;\; \text{for} \;\; j < i, \; (A_{ii}(x) - \lambda_i(x) I(n_i))^{n_i} = 0 \;, \\ & \lambda_i(x) \not\equiv \lambda_j(x), \;\; \text{for} \;\; i \neq j, \; i,j = 1,\ldots,\ell \;. \end{aligned}$$

We already showed

$$A_{ii}(x) = T_i^{-1}(x) F_{ii}(x) T_i(x), \; T_i, T_i^{-1} \in M_n(H_0) \;,$$

and each $F_{ii}(x) - \lambda_i(x)I(n_i)$ is a nilpotent upper triangular matrix with polynomial entries of the form described above. Put

$$G(x) = (G_{ij}(x))_1^{\ell} = T^{-1}(x)A(x)T(x), \; T(x) = \sum_{i=1}^{\ell} \oplus T_i(x) \;.$$

As $\lambda_i(x) \not\equiv \lambda_j(x)$ for $i \neq j$ Problem 1.21.21 implies

$$\nu(G,G) = \sum_{i=1}^{\ell} \nu(G_{ii}, G_{ii}) \;.$$

Let $G^{(k)}(x)$ be defined as above. According to Theorem 1.23.6

$$\nu(G^{(k)},G^{(k)}) \geq \sum_{i=1}^{\ell} (G_{ii}^{(k)},G_{ii}^{(k)}) \quad .$$

Using Theorem 1.24.1 as before we get

$$\nu(G_{ii},G_{ii}) \leq \nu(G_{ii}^{(k)},G_{ii}^{(k)}) \quad .$$

Combine the above inequalities to deduce that

$$\nu(G,G) \leq \nu(G^{(s)},G^{(s)}) \quad .$$

Compare the above inequality with (1.33.8) to obtain the equality sign in (1.33.8). So

$$\nu(G_{ii}^{(s)},G_{ii}^{(s)}) = \nu(G_{ii},G_{ii}),\ i = 1,\dots,\ell \quad . \tag{1.33.11}$$

Let

$$D_i(x) = \lambda_i(x)I(n_i) = \sum_{j=0}^{\infty} D_{ij}x^j,\ D_i^{(k)}(x) = \sum_{j=0}^{k} D_{ij}x^j,\ i = 1,\dots,\ell \quad ,$$

(1.33.12)

$$D(x) = \sum_{i=1}^{\ell} \oplus D_i(x),\ D^{(k)}(x) = \sum_{i=1}^{\ell} \oplus D_i^{(k)}(x) \quad .$$

So

$$\nu(G_{ii}^{(s)} - D_i^{(s)},\ G_{ii}^{(s)} - D_i^{(s)}) = \nu(G_{ii} - D_i,\ G_{ii} - D_i) \quad .$$

As before using Theorem 1.24.1 we deduce that $G_{ii}^{(s)} - D_i^{(S)} \underset{r}{\approx} G_{ii} - D_i$, i.e., $G_{ii}^{(s)} - D_i^{(s)} + D_i \underset{r}{\approx} G_{ii}$. Since $\lambda_i(x) \not\equiv \lambda_j(x)$ for $i \neq j$ we finally deduce that $G \underset{r}{\approx} G^{(s)} - D^{(s)} + D$. Also $GI - I(G^{(s)} - D^{(s)} + D) = x^{s+1}O(1)$. Therefore according to Theorem 1.33.3 $G \underset{a}{\approx} G^{(s)} - D^{(s)} + D$. *The proof of the theorem is completed.* □

<u>Theorem</u> 1.33.13. <u>Let</u> $P(x)$ <u>and</u> $Q(x)$ <u>be matrices of the form</u> (1.33.6)

$$P(x) = \sum_{i=1}^{p} \oplus P_i(x),\ P_i(x) \in M_{m_i}(H_0),\ (\alpha_i I(m_i) - P_i(0))^{m_i} = 0 ,$$
$$\alpha_i \neq \alpha_j \quad \text{for} \quad i \neq j,\ i,j = 1,\ldots,p$$

(1.33.14)

$$Q(x) = \sum_{j=1}^{q} \oplus Q_j(x),\ Q_j \in M_{n_j}(H_0),\ (\beta_j I(n_j) - Q_j(0))^{n_j} = 0 ,$$
$$\beta_i \neq \beta_j \quad \text{for} \quad i \neq j,\ i,j = 1,\ldots,q .$$

<u>Assume furthermore that</u>

(1.33.15) $\alpha_i = \beta_i,\ i = 1,\ldots,t,\ \alpha_i \neq \beta_j,\ i = t+1,\ldots,p, j = t+1,\ldots,q,$
$$0 \leq t \leq \min(p,q) .$$

<u>Then the non-constant local invariant polynomials of</u> $I \otimes P(x) - Q^t(x) \otimes I$ <u>consist of the non-constant local invariant polynomials of</u> $I \otimes P_i(x) - Q_i^t(x) \otimes I,\ i = 1,\ldots,t.$ <u>That is</u>

(1.33.16) $$\kappa_p(P,Q) = \sum_{i=1}^{t} \kappa_p(P_i,Q_i),\ p = 1,2,\ldots, .$$

<u>In particular if</u> $C(x)$ <u>is of the form</u> (1.33.6) <u>then</u>

(1.33.17) $$\eta(C,C) = \max_{1\leq i\leq \ell} \eta(C_i,C_i) .$$

Proof. According to Theorem 1.13.14

$$\kappa_p(P,Q) = \dim W_{p-1} - \dim W_p \quad ,$$

where $W_p \subseteq M_n(\mathbf{C})$ is the subspace of $n \times n$ matrices X_0 such that

$$(1.33.18) \qquad \sum_{j=0}^{k} (P_{k-j}X_j - X_jQ_{k-j}) = 0, \; k = 0,\dots,p \quad .$$

Here

$$P(x) = \sum_{j=0}^{\infty} P_j x^j, \; P_i(x) = \sum_{j=1}^{\infty} P_j^{(i)} x^j, \; P_j = \sum_{i=1}^{p} \oplus P_j^{(i)} \quad ,$$

$$Q(x) = \sum_{j=0}^{\infty} Q_j x^j, \; Q_i(x) = \sum_{j=0}^{\infty} Q_j^{(i)} x^j, \; Q_j \; = \sum_{i=1}^{q} \oplus Q_j^{(i)} \quad .$$

Partition X_j to $(X_{\alpha\beta}^{(j)})$, $X_{\alpha\beta}^{(j)} \in M_{m_\alpha n_\beta}(C)$, $\alpha = 1,\dots,p$, $\beta = 1,\dots,q$. We claim that $X_{\alpha\beta}^{(j)} = 0$ for if either $\alpha > t+1$ or $\beta > t+1$ or $\alpha \neq \beta$. This statement follows easily by induction since in view of Lemma 1.21.5 the equation

$$P_0^{(\alpha)} Y - YQ_0^{(\beta)} = 0$$

has only the trivial solution for those α and β. Thus (1.33.18) splits to the systems

$$\sum_{j=0}^{k} (P_{k-j}^{(i)} X_{ii}^{(j)} - X_{ii}^{(j)} Q_{k-j}^{(i)}) = 0, \; i = 1,\dots,t \quad .$$

Apply the characterizations of $\kappa_p(P,Q)$ and $\kappa_p(P_i,Q_i)$, $i = 1,\ldots,t$ to deduce (1.33.16). Clearly (1.33.16) implies (1.33.17).

$\square$

We close this section by remarking that the main assumption of Theorem 1.33.7 that the characteristic polynomial of $A(x)$ splits in H_0 is not a heavy restriction in view of the Weierstrass preparation theorem, Theorem 1.6.5. That is the eigenvalues of $A(y^m)$ split in H_0. If we choose $m = n!$ then this statement holds for all $A(x) \in M_n(H_0)$. According to Problem 1.33.19 $A(x) \underset{a}{\approx} B(x)$ if and only if $A(y^m) \underset{a}{\approx} B(y^m)$. Therefore the classification problem of analytic similarity classes reduces by Theorem 1.33.7 to determine the structure of the polynomial entries which are lying above the main diagonal. Thus given the rational canonical form of $A(x)$ and the index $\eta(A,A)$ the set of all possible analytic similarity classes which may correspond to A is a certain finite dimensional variety.

The case $n = 2$ is classified completely (Problem 1.33.20). In this case to a given rational canonical form there are at most countable number of analytic similarity classes.

For $n = 3$ we already have examples in which to a given rational canonical form there may correspond a finite dimensional variety of distinct similarity classes (Problem 1.33.21).

Problems

(1.33.19) Let $A(x), B(x) \in M_n(H_0)$. Let m be a positive integer. Assume that $A(y^m)T(y) = T(y)B(y^m)$, $T(y) \in M_n(H_0)$. Show that

$$A(x)Q(x) = Q(x)B(x), \quad Q(y^m) = \frac{1}{m}\sum_{k=1}^{m} T\left(ye^{\frac{2\pi i}{m}k}\right), \quad Q(x) \in M_n(H_0) .$$

Prove that $A(x) \underset{a}{\approx} B(x)$ if and only if $A(y^m) \underset{a}{\approx} B(y^m)$.

(1.33.20) Let $A(x) \in M_2(H_0)$ and assume that

$$|\lambda I - A(x)| = (\lambda-\lambda_1(x))\,(\lambda-\lambda_2(x)),\ \lambda_1(x),\lambda_2(x) \in H_0).$$

$$\lambda_i(x) = \sum_{j=0}^{\infty} \lambda_j^{(i)}x^j,\ i = 1,2,\quad \lambda_j^{(1)} = \lambda_j^{(2)},\ j = 0,\dots,p,\quad \lambda_{p+1}^{(1)} \neq \lambda_{p+1}^{(2)}\ ,$$

$$-1 \leq p \leq \infty (p = \infty \text{ means } \lambda_1(x) = \lambda_2(x))\ .$$

Prove that either $A(x)$ is similar to a diagonal matrix or $A(x)$ is similar to

$$B(x) = \begin{pmatrix} \lambda_1(x) & x^k \\ 0 & \lambda_2(x) \end{pmatrix}\ ,\quad k = 0,1,\dots,p\quad (p > -1)\ .$$

In the second case $\eta(A,A) = k$. (<u>Hint</u>: Use a similarity transformation of the form DAD^{-1} where D is a diagonal matrix.)

(1.33.21) Let $A \in M_3(H_0)$. Assume that $A(x) \underset{r}{\approx} C(p)$, $p(\lambda,x) = \lambda(\lambda-x^{2m})(\lambda-x^{4m})$, $m \geq 1$. Prove that $A(x)$ is analytically similar to a matrix

$$B(x,a) = \begin{pmatrix} 0 & x^{k_1} & a(x) \\ 0 & x^{2m} & x^{k_2} \\ 0 & 0 & x^{4m} \end{pmatrix}\ ,\quad 0 \leq k_1,\ k_2 \leq \infty\ (x^{\infty} = 0)$$

where $a(x)$ is a polynomial such that $\deg a < 4m$. (Use the previous problem.) Assume that $k_1 = k_2 = m$. Show that $B(x,a) \underset{a}{\approx} B(x,b)$ if and only if

(i) $b-a$ is divided by x^m in case that $a(0) \neq 1$.

(ii) $b-a$ is divided by x^{m+k} in case that $a(0) = 1$ and $a^{(i)}(0) = 0$, $i = 1,\dots,k-1$ $a^{(k)}(0) \neq 0$ for $1 \leq k < m$.

(iii) $b-a$ is divided by x^{2m} if $a(0) = 1$, $a^{(i)}(0)$, $i = 1,\dots,m-1$.

That is for $k_1 = k_2 = m$ the set of all analytic similarity classes of matrices $B(x,a)$ can be regarded as a union of $m + 1$ copies of $\mathbb{C}^m - \{0\}$, which are m nonfixed coefficients of $b(x)$.

(1.33.22) Let P and Q satisfy the assumptions of Theorem 1.33.13. Show that P and Q are analytically similar if and only if $p = q = t$, $m_i = n_i$, $i = 1,\ldots,t$ and $P_i(x) \underset{a}{\approx} Q_i$ for $i = 1,\ldots,t$.

1.34. Similarity to diagonal matrices.

Theorem 1.34.1. Let $A(x) \in M_n(H_0)$ and assume that the characteristic polynomial splits in H_0 as given in (1.33.5). Let

$$B(x) = \mathrm{diag}(\lambda_1(x),\ldots,\lambda_n(x)) \ . \tag{1.34.2}$$

Then $A(x)$ and $B(x)$ are not analytically similar if and only if there exists a non-negative integer p such that

$$\begin{aligned} &\kappa_p(A,A) + \kappa_p(B,B) < 2\kappa_p(A,B) \ , \\ &\kappa_j(A,A) + \kappa_j(B,B) = 2\kappa_j(A,B), \ j = 0,\ldots,p-1, \ \text{if} \ p \geq 1 \ . \end{aligned} \tag{1.34.3}$$

In particular $A(x)$ and $B(x)$ are analytically similar if and only if the three matrices given (1.22.6) are equivalent over H_0.

Proof. Suppose first that (1.34.3) holds. Then the three matrices in (1.22.6) are not equivalent. Hence $A(x) \not\approx_a B(x)$. Assume now that $A(x) \not\approx_a B(x)$. Without a restriction in the generality we may assume that $A(x) = C(x)$ where $C(x)$ is given in (1.33.6). Let

$$B(x) = \sum_{j=1}^{\ell} \oplus B_j(x), \ B_j(0) = \alpha_j I(n_j), \ j = 1,\ldots,\ell, \ m_j = n_0+\ldots+n_j \ (n_0=0) \ ,$$
$$j = 1,\ldots,\ell \ .$$

We prove (1.34.3) by induction on the dimension n. For $n = 1$ the theorem is obvious. Assume that the theoren holds for $n \leq N-1$. Let $n = N$. If $A(0) \not\approx B(0)$ then Theorem 1.22.3 implies the inequality (1.34.3) for $p = 0$ and the theorem is proved. Suppose now that $A(0) \approx B(0)$. That is $A_j(0) = B_j(0) = \alpha_j I(n_j)$, $j = 1,\ldots,\ell$. Suppose that $\ell > 1$. By Theorem 1.33.13

$$\kappa_p(A,A) = \sum_{j=1}^{\ell} \kappa_p(A_j,A_j),\ \kappa_p(A,B) = \sum_{j=1}^{\ell} \kappa_p(A_j,A_j) \quad .$$

$$\kappa_p(B,B) = \sum_{j=1}^{\ell} \kappa_p(B_j,B_j),\ p \geq 1 \quad .$$

Since $A(x) \underset{a}{\not\approx} B(x)$ if and only if $A_j(x) \underset{a}{\not\approx} B_j(x)$ for some j (Problem 1.33.22), using the induction hypothesis we deduce (1.34.3). It is left to consider the case

$$A(0) = B(0) = \alpha_0 I,\ \kappa_0(A,A) = \kappa_0(A,B) = \kappa_0(B,B) = 0 \quad .$$

Put

$$A^{(1)}(x) = (A(x) - \alpha_0 I)/x,\ B^{(1)}(x) = (B(x) - \alpha_0 I)/x \quad .$$

Clearly

$$\kappa_p(A,A) = \kappa_{p-1}(A^{(1)},A^{(1)}),\ \kappa_p(A,B) = \kappa_{p-1}(A^{(1)},B^{(1)}),\ \kappa_p(B,B) = \kappa_{p-1}(B^{(1)},B^{(1)}),\ p = 1,2,\ldots \quad .$$

Also $A(x) \underset{a}{\approx} B(x)$ if and only if $A^{(1)}(x) \underset{a}{\approx} B^{(1)}(x)$. We now continue the process as above. If at stage k either $A^{(k)}(0) \not\approx B^{(k)}(0)$ or $A^{(k)}(0)$ has more than one eigenvalue we conclude the theorem as before. The only possibility which is left is

$$A(x) = B(x) = \lambda(x)I,\ \lambda(x) \in H_0 \quad .$$

However this case violates the assumption $A(x) \underset{a}{\not\approx} B(x)$. The proof of the theorem is completed.

□

1.35. Strict similarity to diagonal matrices.

Let $A(x) \in M_n(H_0)$. According to the Weierstrass preparation theorem (Theorem 1.6.5) the eigenvalues of $A(y^m)$ are analytic in y for some $1 \le m \le n$. That is the eigenvalues $\lambda_1(x),\ldots,\lambda_n(x)$ of $A(x)$ are multivalued analytic functions in x which have the expansion

$$\lambda_j(x) = \sum_{k=0}^{\infty} \lambda_{jk} x^{k/m}, \ j = 1,\ldots,n \ .$$

In particular each $\lambda_j(x)$ has at most m branches. For more properties of the eigenvalues $\lambda_j(x)$, $j = 1,\ldots,n$ see Kato, [1976, Chapter 2]. Let $A(x) \in M_n(\mathbf{C}[x])$. So

$$A(x) = \sum_{k=0}^{m} A_k x^k, \ A_k \in M_n(\mathbf{C}), \ k = 0,1,\ldots,m \ . \tag{1.35.1}$$

The eigenvalues of $A(x)$ satisfy the equation

$$|\lambda I - A(x)| = \lambda^n + \sum_{j=0}^{n} a_j(x)\lambda^{n-j} = 0, \ a_j(x) \in C[x], \ j = 1,\ldots,n \ . \tag{1.35.2}$$

Thus the eigenvalues $\lambda_1(x),\ldots,\lambda_n(x)$ are algebraic functions of x (see for example Gunning-Rossi [1965]). The equation (1.35.2) describes one or several compact Riemann manifolds according to the polynomial (1.35.2) is irreducible or reducible in $C[x,\lambda]$. According to the Weierstrass preparation theorem $\lambda_j(x)$ has the following expansion around $x = \zeta$

$$\lambda_j(x) = \sum_{k=0}^{\infty} \lambda_{jk}(\zeta)(x-\zeta)^{k/m}, \ j = 1,\ldots,n \ . \tag{1.35.3}$$

The number m *can be chosen* to be independent of ζ. For example $m = n!$ will always be correct. These expansions are called the Puiseaux expansion of $\lambda_j(x)$. Since $A(x)$ is a polynomial matrix each $\lambda_j(x)$ can be expanded in the neighborhood of ∞. To do so we note that

$$A(x) = x^m B(\tfrac{1}{x}), \quad B(y) = \sum_{k=0}^{m} A_k y^{m-k} .$$

Expand now the eigenvalues of $B(y)$ at $y = 0$ to get

$$\lambda_j(x) = x^m \sum_{k=0}^{\infty} \lambda_{jk}(\infty) x^{-k/m}, \quad j = 1,\ldots,n . \tag{1.35.4}$$

Definition 1.35.5. Let $A(x), B(x) \in M_n(C[x])$. Then $A(x)$ and $B(x)$ are called strictly similar $(A \underset{s}{\approx} B)$ if there exists $P \in M_n(C)$, $|P| \neq 0$ such that $B(x) = PA(x)P^{-1}$.

From Lemma 1.22.8 it follows that if $A(x) \underset{s}{\approx} B(x)$ then the three matrices in (1.30.4) are equivalent over $C[x]$. To take in account the point $x = \infty$ we need to homogenize as in Section 1.14.

Definition 1.35.6. Let $A(x)$ be given by (1.35.1). Denote by $A(x_0, x_1)$ the corresponding homogeneous matrix

$$A(x_0,x_1) = \sum_{k=0}^{m'} A_k x_0^{m'-k} x_1^k \in M_n(C[x_0,x_1]) , \tag{1.35.7}$$

where $m' = 0$ if $A(x) = 0$ and $A_{m'} \neq 0$, $A_j = 0$ for $m' < j \leq m$ if $A(x) \neq 0$.

Clearly if $B(x) = PA(x)P^{-1}$ then $B(x_0,x_1) = PA(x_0,x_1)P^{-1}$. According to Lemma 1.22.8 the matrices

$$\begin{aligned} &I \otimes A(x_0,x_1) - A^t(x_0,x_1) \otimes I, \; I \otimes A(x_0,x_1) - B^t(x_0,x_1) \otimes I , \\ &\qquad I \otimes B(x_0,x_1) - B^t(x_0,x_1) \otimes I . \end{aligned} \tag{1.35.8}$$

are equivalent over $C[x_0,x_1]$. Thus from Lemma 1.10.3 it follows

Lemma 1.35.9. Let $A(x), B(x) \in M_n(C[x])$. Assume that $A(x) \underset{s}{\approx} B(x)$. Then the three matrices in (1.35.8) have the same invariant polynomials over $C[x_0,x_1]$.

<u>Definition</u> 1.35.10. <u>Let</u> $A(x), B(x) \in M_n(C[x])$. <u>And assume that</u> $A(x_0,x_1)$ <u>and</u> $B(x_0,x_1)$ <u>are the homogeneous matrices corresponding to</u> $A(x)$ <u>and</u> $B(x)$ <u>respectively. Denote by</u> $i_k(x_0,x_1)$, $k = 1,\ldots,r(A,B)$ <u>the invariant factors of the matrix</u> $I \otimes A(x_0,x_1) - B^t(x_0,x_1) \otimes I$.

As in the proof of Lemma 1.14.8 we have that $i_k(x_0,x_1)$ is a homogeneous polynomial for $k = 1,\ldots,r(A,B)$. Moreover $i_k(1,x)$, $k = 1,\ldots,r(A,B)$, are the invariant factors of $I \otimes A(x) - B^t(x) \otimes I$. (See Problems 1.14.30 - 1.14.31.)

We now answer the problem when $A(x)$ is strictly similar to a diagonal matrix $B(x) \in M_n(\mathbf{C}[x])$ of the form (1.34.2).

<u>Theorem</u> 1.35.11. <u>Let</u> $A(x) \in M_n(\mathbf{C}[x])$. <u>Assume that the characteristic polynomial of</u> $A(x)$ <u>splits to linear factors over</u> $\mathbf{C}[x]$. <u>Then</u> $A(x)$ <u>is strictly similar to the diagonal matrix given by</u> (1.34.2) <u>if and only if the three matrices in</u> (1.34.8) <u>have the same invariant factors over</u> $C[x_0,x_1]$.

<u>Proof</u>. Without loss of generality we may assume that $B(x)$ is of the form

$$B(x) = \sum_{i=1}^{\ell} \lambda_i(x) I(n_i), \quad \lambda_i(x) \neq \lambda_j(x) \quad \text{for } i \neq j, \; i,j = 1,\ldots,\ell \;. \tag{1.35.12}$$

Thus for all ζ except a finite number of points we have

$$\lambda_i(\zeta) \neq \lambda_j(\zeta), \quad i,j = 1,\ldots,\ell \;. \tag{1.35.13}$$

Let $P_j(x)$ be the projection of $A(x)$ on $\lambda_j(x)$, $j \equiv 1,\ldots,\ell$. Suppose that (1.35.13) is satisfied at x_0. According to Problem 1.28.24 each $P_j(x)$ is analytic in the neighborhood of ζ. Assume that (1.35.13) does not hold for $\zeta \in \mathbf{C}$. The assumption that the three matrices in (1.35.8) have the same invariant polynomials imply that the matrices in (1.30.4) are equivalent over H_ζ. Now use Theorem 1.34.1 to get that $A(x)$ and $B(x)$ are analytically similar over H_ζ. Clearly $P_j(B)$ the projection of $B(x)$ on $\lambda_j(x)$ is equal to $0 \oplus I(n_j) \oplus 0$. From Problem (1.28.27) we deduce that $P_j(x)$ is also analytic in the neighborhood of ζ. The same arguments apply for $\zeta = \infty$. This follows by considering the matrices $A(x_0,1)$ and $B(x_0,1)$. Thus $P_j(x)$ is analytic on the Riemann sphere. In particular $P_j(x)$ is bounded. The Liouville theorem (e.g. Rudin [1974]) implies that P_j is a constant matrix. Let

$$P_j C^n = [x^{j1},\dots,x^{jn_j}],\ u = 1,\dots,\ell,\ X = (x^{11},\dots,x^{1n_1},\dots,x^{\ell 1},\dots,x^{\ell n_\ell}) \in M_n(\mathbf{C}) .$$

According to Problem 1.28.28 $X^{-1}A(x)X = B(x)$ for ζ which satisfy (1.35.13). Finally the analyticity of $A(x)$ and $B(x)$ implies the validity of the above equality for all x.

□

Let $A(x) \in M_n(\mathbf{C})$. Suppose that $A(x)$ is strictly similar to a diagonal matrix $B(x)$. Consider the corresponding homogeous matrix $A(x_0,x_1)$. Then we obviously have that for any $\zeta_0,\zeta_1 \in C$, $A(\zeta_0,\zeta_1)$ is similar to a diagonal matrix, i.e., $A(\zeta_0,\zeta_1)$ is diagonable. However if $A(\zeta_0,\zeta_1)$ is diagonable this does not imply that $A(x)$ is strictly equivalent to some diagonal matrix. For example

$$(1.35.14) \qquad A(x) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} x^2 .$$

See Problem 1.35.24. We now give a sufficient condition on $A(x)$, such that $A(\zeta_0,\zeta_1)$ is diagonable for any ζ_0,ζ_1, which implies that $A(x)$ is strictly similar to a diagonal matrix.

Definition 1.35.15. Let $A(x) \in M_n(\mathbf{C}[x])$ be of the form (1.35.1) normalized by the condition $A_m \neq 0$ if $m \geq 1$. Let $\lambda_p(x)$ and $\lambda_q(x)$ be two eigenvalues of $A(x)$. The eigenvalues $\lambda_p(x)$ and $\lambda_q(x)$ are said to touch at ζ if the Puiseaux series of $\lambda_p(x)$ and $\lambda_q(x)$ at $x = \zeta$ satisfy the following relations

$$(1.35.16) \qquad \lambda_{pk}(\zeta) = \lambda_{qk}(\zeta), \ k = 0,\ldots,s$$

for a finite or infinite ζ.

Theorem 1.35.17. Let $A(x) \in M_n(C[x])$ be of the form (1.35.1) normalized by the condition $A_m \neq 0$ if $m > 1$. Assume that the corresponding homogeneous matrix $A(x_0,x_1)$ is diagonable for any $\zeta_0,\zeta_1 \in C$. Suppose furthermore that no two distinct eigenvalues of $A(x)$ touch at any finite or infinite point of the Riemann sphere. Then there exists a constant matrix $X \in M_n(C)$, $|X| \neq 0$ such that

$$(1.35.18) \qquad X^{-1}A(x)X = \sum_{k=0}^{m} D_k x^k ,$$

where $D_0,\ldots,D_m$ are diagonal matrices.

Proof. Clearly we can view $A(x)$ as a matrix in $M_n(\mathcal{M})$, where $\mathcal{M}$ is the field of rational functions. Let $\mathbb{K}$ be a finite extension of all such that $|\lambda I - A(x)|$ splits to linear factors over $\mathbb{K}$. Thus $A(x)$ ℓ distinct eigenvalues $\lambda_1(x),\ldots,\lambda_\ell(x)$, such that $\lambda_k(x)$ has multiplicity n_k, $k = 1,\ldots,\ell$. Thus for all except a finite number of points (1.35.13) holds. Assume that ζ satisfies (1.35.13). Denote by $P_j(\zeta)$ the projection of $A(x)$ on $\lambda_j(\zeta)$, $j = 1,\ldots,\ell$. According to Problem 1.28.24 $P_j(x)$ is analytic in the neighborhood of ζ. Also in view of Problem 1.28.32 $P_j(\zeta)$ is given by the formula

$$(1.35.19) \qquad P_j(\zeta) = \prod_{k=1,k\neq j}^{\ell} [A(\zeta) - \lambda_k(\zeta)I]/[\lambda_j(\zeta) - \lambda_k(\zeta)] .$$

We claim that in the neighborhood of any point ζ, $P_j(y^s+\zeta)$ is analytic in y. (In case that $\zeta = \infty$ $P_j(y^{-s})$ is analytic in y in the neighborhood of the origin.) Clearly it is enough to consider the points ζ at which (1.35.13) is violated. For simplicity of notation we consider $P_1(x)$ in the neighbornood of ζ. Assume first that ζ is finite. Suppose that

$$\eta = \lambda_1(\zeta)=\ldots=\lambda_u(\zeta),\ \lambda_k(\zeta) \neq \lambda_1(\zeta),\ k = u+1,\ldots,\ell \quad .$$

According to Theorem 1.29.19 there exists $Q(x) \in M_n(H_\zeta)$, $|Q(\zeta)| \neq 0$ such that

$$Q^{-1}(x)A(x)Q(x) = C_1(x) \oplus C_2(x),\ C_1(x) \in M_{m_1}(H_\zeta) \quad ,$$

$$C_2(x) \in M_{m_2}(H_\zeta),\ m_1 = \sum_{i=1}^{u} n_i,\ m_2 = n-m_1 \quad .$$

Moreover the eigenvalues of $C_1(x)$ and $C_2(x)$ are $\lambda_1(x),\ldots,\lambda_u(x)$ and $\lambda_{u+1}(x),\ldots,\lambda_\ell(x)$ with the multiplicities $n_1,\ldots,n_u$ and $n_{u+1},\ldots,n_\ell$ respectively. Since $A(x)$ is diagonable for each x, $C_1(x)$ and $C_2(x)$ are also diagonable in the neighborhood of ζ. So

$$\prod_{k=u+1}^{\ell} (C_2(x) - \lambda_k(x)I) = 0 \quad ,$$

which yields

$$\prod_{k=2}^{\ell} (A(x) - \lambda_k(x)I)/[\lambda_1(x) - \lambda_k(x)] =$$

$$Q^{-1}(x)\{\prod_{k=2}^{\ell} [C_1(x) - \lambda_k(x)I]/[\lambda_1(x) - \lambda_k(x)] \oplus 0\}Q(x) \quad .$$

The assumption that $\eta = \lambda_1(\zeta)=\ldots=\lambda_u(\zeta)$ and the diagonability of $C_1(x)$ imply

$$C_1(x) = \eta I + (x-\zeta)C_3(x) \quad .$$

Therefore the Puiseaux series of $\lambda_1(x),\ldots,\lambda_u(x)$ satisfy $\lambda_{jk}(\zeta) = 0$, $k = 1,\ldots,s-1$, $j = 1,\ldots,u$. As no two distinct eigenvalues of $A(x)$ touch at ζ we get that $\lambda_{1s}(\zeta) \neq \lambda_{js}(\zeta)$, for $j = 2,\ldots,u$. So

$$\prod_{j=2}^{u} [C_1(v^s+\zeta) - \lambda_j(y^s+\zeta)I]/[\lambda_1(y^s+\zeta) - \lambda_j(y^s+\zeta)]$$

$$= \prod_{j=2}^{u} [C_3(y^s+\zeta) - (\sum_{k=s}^{\infty} \lambda_{jk}(\zeta)y^{k-s})I]/\{\sum_{k=s}^{\infty} [\lambda_{1k}(\zeta) - \lambda_{jk}(\zeta)]y^{k-s}\} \quad .$$

Also $(\lambda_1(y^s+\zeta) - \lambda_j(y^s+\zeta))^{-1} \in H_\zeta$ for $j = u+1,\dots,\ell$. This shows that $P_1(y^s+\zeta)$ is analytic in H_0 for any finite ζ. By considering $A(x) = x^m A(\frac{1}{x})$ we transform ∞ to 0 and the same result applies to $\zeta = \infty$. In particular we have that $P_1(x)$ is bounded on the Riemann sphere. Put $P_1(x) = (p_{ij}^{(1)}(x))$. Let $\max|p_{ij}^{(1)}(x)| = |p_{ij}^{(1)}(\zeta_{ij})|$, ($\zeta_{ij}$ may be ∞). As $p_{ij}^{(1)}(\zeta_{ij} + y^s)$ is analytic in y the maximum principle implies that $p_{ij}^{(1)}(\zeta_{ij} + y^s)$ is constant in the neighborhood of the origin (e.g. Rudin [1974]). The analytic continuation principle yields that $p_{ij}^{(1)}(x)$ is constant. Hence $P_1 = P_1(x)$ is constant and in the same manner we deduce that all $P_j(x)$ are constant. Define the matrix X as in the proof of Theorem 1.35.11 to deduce (1.35.18).

□

Corollary 1.35.20. *Let* $A(x)$ *be of the form* (1.35.1). *Assume that the matrices* $A_0,\dots,A_m$ *are diagonable and* $A_iA_j \neq A_jA_i$ *for some* $0 \leq i < j \leq m$. *Then either there exist* $\zeta_0,\zeta_1 \in \mathbb{C}$ *such that* $A(\zeta_0,\zeta_1)$ *is not diagonable or there exists a point* ζ_0 *on the Riemann sphere (possible* $\zeta_0 = \infty$) *and two distinct eigenvalues of* $A(x)$ *which touch at* ζ_0.

It can be shown that for the matrix (1.35.14) the two distinct eigenvalues of $A(x)$ touch at ∞. (Problem 1.35.25.) However if $A(x)$ is a pencil, i.e., $A(x) = A_0 + xA_1$ then two conditions of Theorem 1.35.17 are redundant. More precisely we have

Theorem 1.35.21. *Let* $A(x) = A_0 + xA_1$ *be a pencil in* $M_n(\mathbb{C}[x])$. *Assume that for any* $\zeta \in \mathbb{C}$, $A(\zeta)$ *is a diagonable matrix. Then the eigenvalues of* $A(x)$ *are linear functions in* x

$$\lambda_j(x) = \alpha_j + \beta_j x, \quad j = 1,\dots,n \quad . \tag{1.35.22}$$

In particular no two distinct eigenvalues of $A(x)$ *intersect at any point of the Riemann sphere.*

Proof. Consider a multivalued function $\lambda_j(x)$ which has the Puiseaux series (1.35.3) clearly $\lambda_j'(x)$ is also multivalued functions which is given by

$$\lambda_j'(x) = \sum_{k=0}^{\infty} \frac{k}{s} \lambda_{jk}(\zeta)(x-\zeta)^{k/s-1} \quad .$$

As $A(\zeta)$ is a diagonable matrix any $\lambda_j(\zeta)$ is a geometrically simple. According to Theorem 1.32.1 $\lambda_{jk} = 0$, $k = 1,\ldots,s-1$. So $\lambda_j'(x)$ is bounded in the neighborhood of ζ. Let $\zeta = \infty$. Then the Puiseaux series of $\lambda_j(x)$ are of the form

$$\lambda_j(x) = \sum_{k=0}^{\infty} \lambda_{jk}(\infty)x^{(s-k)/s} \quad .$$

So $\lambda_j'(x)$ is also bounded at the neighborhood of ∞. Now use the arguments of the last part of the proof of Theorem 1.35.17 to deduce that $\lambda_j'(x) = \beta_j$, $j = 1,\ldots,n$. This of course implies (1.35.22). In view of (1.35.22) the equalities (1.35.16) imply that $\alpha_p = \alpha_q$, $\beta_p = \beta_q$, i.e. $\lambda_p(x) = \lambda_q(x)$ for all x.

□

Problems

(1.35.23) Let $A(x) \in M_n(\mathbf{C}[x])$. Assume that there exists an infinite sequence of distinct points $\{\zeta_k\}_1^{\infty}$ such that $A(\zeta_k)$ is diagonable, for $k = 1,2,\ldots$. Show that $A(x)$ is diagonable for all but a finite number of points. (Hint - Consider the rational canonical form of $A(x)$ over the field of rational functions.).

(1.35.24) Let $A(x_0,x_1) = x_0A_0 + x_1A_1 \in M_n(\mathbf{C}[x_0,x_1])$. Show that if $A(x_0,x_1)$ is diagonable for any $x_0,x_1 \in \mathbf{C}$ then (1.35.18) holds $(m = 1)$.

(1.35.25) Consider the mat ix (1.35.14). Show that the eigenvalues of $A(x_0,x_1)$ are $\lambda_1 = x_1^2$, $\lambda_2(x) = x_0^2 + x_1^2$. Prove for $x_0 \neq 0$ $A(x_0,x_1)$ is diagonable. As $A(0,x_1) = x_1^2I$ $A(x_0,x_1)$ is diagonable for all $x_0,x_1 \in \mathbf{C}$. Show that the eigenvalues of $A(1,x)$ touch at $\zeta = \infty$.

1.36. Strict similarity of pencils.

Let $A(x)$ and $B(x)$ be two linear pencils

$$A(x) = A_0 + xA_1,\ B(x) = B_0 + xB_1 \in M_n(C[x]) \ .$$

Assume that $A(x)$ and $B(x)$ are strictly similar. That is

$$B_0 = PA_0P^{-1},\ B_1 = PA_1P^{-1} \ , \tag{1.36.1}$$

for some non-singular $P \in M_n(C)$. From (1.22.5) it follows

Lemma 1.36.2. Let $A(x)$ and $B(x)$ be two pencils in $M_n(C[x])$ which are strictly similar. Then the three pencils in (1.30.4) are strictly equivalent.

Using the Kronecker's result (Theorem 1.14.23) we can determine whether the pencils in (1.30.4) are strictly equivalent. We now study the implications of the assumption that the three pencils in (1.30.4) are strictly equivalent. More precisely we have

Lemma 1.36.3. Let $A(x)$ and $B(x)$ be two pencils in $M_n(C[x])$ such that the first two pencils in (1.30.4) are strictly equivalent. Then there exist two constant non-zero matrices $U,V \in M_n(C)$ such that

$$A(x)U - UB(x) = 0,\ VA(x) - B(x)V = 0 \ . \tag{1.36.4}$$

In particular

$$A_0 \ker(V),\ A_1 \ker(V) \subseteq \ker(V),\ B_0 \ker(U),\ B_1 \ker(U) \subseteq \ker(U) \ . \tag{1.36.5}$$

Proof. As

$$A(x)I - IA(x) = 0$$

the strict equivalence of the first two pencils in (1.30.4) implies that $C(A,B)$ and $C(B,A)$ contains at least one non-zero matrix as the first row and column index of $I \otimes A - A^t \otimes I$ is 0. Assume that $\xi \in \ker(U)$. That is $U\xi = 0$. From the first equality in (1.36.4) we deduce $U(B(x)\xi) = 0$. So $B(x) \ker U \subseteq \ker U$ which is equivalent to the second part of the inequality (1.36.5). The first part of (1.36.5) is established in a similar way.

□

<u>Theorem</u> 1.36.6. <u>Let</u> $A(x) = A_0 + xA_1$, $B(x) = B_0 + xB_1$, $A_i, B_i \in M_n(\mathbb{C})$, $i = 1,2$ <u>be two given pencils. Suppose that either</u> A_0, A_1 <u>or</u> B_0, B_1 <u>do not have a common invariant subspace different from</u> $[0]$ <u>or</u> $\mathbb{C}^n$ <u>(the trivial subspaces). Then</u> $A(x) \underset{s}{\approx} B(x)$ <u>if and only if the first two pencils in</u> (1.30.4) <u>are strictly equivalent.</u>

<u>Proof.</u> Assume that A_0 and A_1 do not have in common non-trivial invariant subspace. Then the matrix $V \neq 0$ in (1.36.4) must be non-singular in view of (1.36.5). So $A(x) \underset{s}{\approx} B(x)$. In case that B_0 and B_1 do not have a common non-trivial invariant subspace we get that $|U| \neq 0$.

□

A simple criterion for A_0 and A_1 not have a common non-trivial subspace is that the polynomial $|\lambda I - A(x)|$ is irreducible over $\mathbb{C}[x,\lambda]$. (Problem 1.36.17.)

Next we show the connection between the notions of analytic similarity of matrices over H_0 and strict similarity of pencils. Let $A(x), B(x) \in M_n(H_0)$ and assume that $\eta(A,A) = 1$. Suppose that $A(x) \underset{r}{\approx} B(x)$. According to Theorem 1.33.3 $A(x) \approx B(x)$ if and only if there exist two matrices T_0, T_1, $|T_0| \neq 0$ such that

$$A_0T_0 = T_0B_0, \quad A_1T_0 + A_0T_1 = T_0B_1 + T_1B_0 .$$

Let

$$(1.36.7) \qquad F(A_0,A_1) = \begin{pmatrix} A_0 & A_1 \\ 0 & A_0 \end{pmatrix} \in M_{2n}(\mathbb{C}) .$$

Then (1.33.4) in this case is equivalent to

$$(1.36.8) \qquad F(A_0,A_1)F(T_0,T_1) = F(T_0,T_1)F(B_0,B_1) \ .$$

As $|F(T_0,T_1)| = |T_0|^2$, T_0 is non-singular if and only if $F(T_0,T_1)$ is non-singular.

Definition 1.36.9. Let $A_i,B_i \in M_n(\mathbf{C})$, $i = 1,2$. Then $F(A_0,A_1)$ and $F(B_0,B_1)$ are called strongly similar $(F(A_0,A_1) \approxeq F(B_0,B_1))$ if there exists a non-singular matrix $F(T_0,T_1)$ which satisfies (1.36.8).

Clearly if $F(A_0,A_1) \approxeq F(B_0,B_1)$ then $F(A_0,A_1) \approx F(B_0,B_1)$. It can be shown that the notion of the strong similarity is stronger than the ordinary notion of similarity. (Problem 1.36.24.)

Lemma 1.36.10. The matrices $F(A_0,A_1)$ and $F(B_0,B_1)$ are strongly similar if and only if the pencils

$$A(x) = F(0,I) + x\,F(A_0,A_1), \ B(x) = F(0,I) + x\,F(B_0,B_1)$$

are strictly similar.

Proof. Let $P = (P_{ij})$, $P_{ij} \in M_n(\mathbf{C})$, $i,j = 1,2$. Then $F(0,I)P = PF(0,I)$ if and only if $P_{11} = P_{22}$, $P_{21} = 0$. That is $P = F(P_{11},P_{12})$ and the lemma follows.

Clearly if $F(A_0,A_1)$ and $F(B_0,B_1)$ are strongly similar then $A_0 \approx B_0$. Without the restriction in generality we may assume that $A_0 = B_0$. (See Problem 1.36.19.) Consider all matrices T_0,T_1 satisfying (1.36.7). For $B_0 = A_0$ (1.36.8) reduces to $A_0T_0 = T_0A_0$, $A_0T_1 - T_1A_0 = T_0B_1 - A_1T_0$. According to Theorem 1.23.3 the set of all matrices T_0 which satisfies the above requirements is of the form

$$(1.36.11) \qquad P(A_1,B_1) = \{T_0 \mid T_0 \in C(A_0),\ \mathrm{tr}(V(T_0B_1 - A_1T_0)) = 0,\ V \in C(A_0)\} \ .$$

We also observe

<u>Lemma</u> 1.36.12. <u>Suppose that</u> $F(A_0,A_1) \approx F(A_0,B_1)$. <u>Then</u>

$$\dim \Gamma(A_1,A_1) = \dim \Gamma(A_1,B_1) = \dim \Gamma(B_1,B_1) \quad . \tag{1.36.13}$$

As in Theorem 1.22.9 for a fixed A_0,A_1 there exists a neighborhood $D(A_1,\rho)$ such that the first two equalities in (1.36.13) imply that $F(A_0,A_1) \approx F(A_0,A_1)$ for $B_1 \in D(A,\rho)$ (Problem 1.36.18).

Next we consider a splitting result analogous to Theorem 1.29.19.

<u>Theorem</u> 1.36.14. <u>Assume that</u>

$$A_o = \mathrm{diag}(A_{11}^{(0)},A_{22}^{(0)}),\ A_{ii}^{(0)} \in M_{n_i}(\mathbf{C}),\ i = 1,2 \quad , \tag{1.36.15}$$

<u>where</u> $A_{11}^{(0)}$ <u>and</u> $A_{22}^{(0)}$ <u>do not have a common eigenvalue. Let</u>

$$A_1 = \left(A_{ij}^{(1)}\right)_1^2,\ B_1 = \left(B_{ij}^{(1)}\right)_1^2$$

<u>be the conformal partition of</u> A_1 <u>and</u> B_1 <u>with</u> A_0. <u>Then</u>

$$\Gamma(A_1,B_1) = \Gamma(A_{11}^{(1)},B_{11}^{(1)}) \oplus \Gamma(A_{22}^{(1)},B_{22}^{(1)}) \quad . \tag{1.36.16}$$

<u>Moreover,</u> $F(A_0,A_1) \approx F(A_0,B_1)$ <u>if and only if</u> $F(A_{ii}^{(0)},A_{ii}^{(1)}) \approx F(A_{ii}^{(0)},B_{ii}^{(1)})$ <u>for</u> $i = 1,2$.

<u>Proof.</u> According to Problem (1.21.22)

$$C(A_0) = C(A_{11}^{(0)}) \oplus C(A_{22}^{(0)}) \quad .$$

Then the trace condition in (1.36.11) reduces to

$$\operatorname{tr}[V_1(T_1^{(0)}B_{11}^{(1)} - A_{11}^{(1)}T_1^{(0)}) + V_2(T_2^{(0)}B_{22}^{(1)} - A_{22}^{(1)}T_2^{(0)})] = 0 \quad .$$

Here

$$V = V_1 \oplus V_2, \; T_0 = T_1^{(0)} \oplus T_2^{(0)} \in C(A_{11}^{(0)}) \oplus C(A_{22}^{(0)}) \quad .$$

Choosing either $V_1 = 0$ or $V_2 = 0$ we obtain (1.36.16). As $|T_0| = |T_1^{(0)}|\,|T_2^{(0)}|$, T_0 is non-singular if and only if $T_0^{(0)}$ and $T_2^{(0)}$ are non-singular. This establishes the last claim of the theorem.

□

Thus, the classification of strong similarity classes for the matrices $F(A_0,A_1)$ reduces to the case where A_0 is nilpotent (Problem 1.36.20). In case that $A_0 = 0$ the notion of the strong similarity reduces to the standard notion of similarity. In case that $A_0 = H(n)$ the strong similarity classes of $F(A_0,A_1)$ are classified completely (Problem 1.36.23). This case corresponds to the case discussed in Theorem 1.29.17. The case $A_0 = H(n) \oplus H(n)$ can be also classified completely using the results of Problem 1.33.20 (Problem 1.36.25).

<u>Problems</u>

(1.36.17) Let $A(x) \in M_n(\mathbf{C}[x])$ and assume that $A(x)\mathbf{U} \subseteq \mathbf{U}$ where $\mathbf{U}$ is a subspace of $\mathbf{C}^n$, $1 \leq \dim \mathbf{U} \leq n-1$. Let $p(\lambda,x) \in \mathbf{C}[\lambda,x]$ be the minimal polynomial of the restriction of $A(x)$ to $\mathbf{U}$. Thus $\deg p(\lambda,x) \leq n-1$. Prove that $p(\lambda,x)$ divides $|\lambda I - A(x)|$. That is $|\lambda I - A(x)|$ is reducible over $\mathbf{C}[\lambda,x]$.

(1.36.18) Modify the proof of Theorem 1.22.9 to show that for fixed A_0,A_1 then there exist $\rho > 0$ such that the first two equalities in (1.36.13) for $B \in D(A,\rho)$ imply that $F(A_0,A_1) \approx F(A_0,B_1)$.

(1.36.19) Prove that $F(A_0,A_1) \approx F(B_0,B_1)$ if and only if $F(A_0,A_1) \approx F(PB_0P^{-1},PB_1P^{-1})$ for any non-singular P. Suppose that $F(A_0,A_1) \approx F(B_0,B_1)$. Show that it is possible to choose P such that $A_0 = PB_0P^{-1}$.

(1.36.20) Prove that $F(A_0,A_1) \approx F(B_0,B_1)$ if and only if $F(A_0 - \lambda I, A_1) \approx F(B_0 - \lambda I, B_1)$ for any λ.

(1.36.21) Let $F(A_0,\ldots,A_{s-1}) = (F_{ij})$, $F_{ij} \in M_n(\mathbf{C})$, $i,j = 1,\ldots,s$, $F_{ij} = A_{j-i}$ for $1 \leq i \leq j \leq s$, $F_{ij} = 0$ for $1 \leq j < i \leq s$. $F(A_0,\ldots,A_{s-1})$ and $F(B_0,\ldots,B_{s-1})$ are called strongly similar $(F(A_0,\ldots,A_{s-1}) \approx F(B_0,\ldots,B_{s-1}))$ if there exist $F(T_0,\ldots,T_{s-1})$ such that $F(A_0,\ldots,A_{s-1})F(T_0,\ldots,T_{s-1}) = F(T_0,\ldots,T_{s-1})F(B_0,\ldots,B_{s-1})$. Prove that $F(A_0,\ldots,A_{s-1}) \approx F(B_0,\ldots,B_{s-1})$ if and only if the equalities (1.33.4) hold for $k = 0,1,\ldots,s-1$ where $|T_0| \neq 0$.

(1.36.22) Let

$$Z = H(n)\oplus\cdots\oplus H(n),\ X = (X_{pq}),\ Y = (Y_{pq}) \in M_m(\mathbf{C}),\ m = sn\ ,$$

$$X_{pq} = (x_{ij}^{(pq)}),\ Y_{pq} = (y_{ij}^{(pq)}) \in M_n(\mathbf{C}),\ p,q = 1,\ldots,s\ .$$

Define

$$A_r = (a_{pq}^{(r)}),\ B_r = (b_{pq}^{(r)}) \in M_s(\mathbf{C})\ ,$$

$$a_{pq}^{(r)} = \sum_{i=1}^{r+1} x_{(n-r+i-1)i}^{(pq)},\ b_{pq}^{(r)} = \sum_{i=1}^{r+1} y_{(n-r+i-1)i}^{(pq)},\ r = 0,\ldots,n-1\ .$$

Using Theorem 1.21.9 prove that $F(Z,X) \approx F(Z,Y)$ if and only if $F(A_0,\ldots,A_{n-1}) \approx F(B_0,\ldots,B_{n-1})$. (To do that one needs the following auxiliary result. Consider $X = (X_{pq})$ of the above form. Assume that each X_{pq} is an upper triangular matrix. Expand the determinant of X by the rows $n,2n,\ldots,sn$ and use the induction to show

$$|X| = \prod_{r=1}^{n} |(x_{rr}^{(pq)})_{p,q=1}^{s}|).$$

(1.36.23) Use the two preceding problems to prove that $F(H(n),X) \approx F(H(n),Y)$, $X = (x_{ij})$, $Y = (y_{ij}) \in M_n(\mathbf{C})$ if and only if

$$\sum_{i=1}^{r} x_{(n-r+i)i} = \sum_{i=1}^{r} y_{(n-r+i)i} \quad \text{for} \quad r = 1,\ldots,n \quad .$$

(1.36.24) Let $X = (x_{ij}) \in M_2(\mathbf{C})$. Prove that if $x_{21} \neq 0$ then $F(H(2),X) \approx H(4)$. Combine this result with Problem 1.36.23 to show the existence of $Y \in M_2(\mathbf{C})$ such that $F(H(2),X) \approx F(H(2)Y)$ but $F(H(2),X) \not\approx F(H(2),Y)$.

(1.36.25) Assume in Problem (1.36.22) $s = 2$. Let

$$A(x) = \sum_{i=0}^{n-1} A_i x^i, \quad B(x) = \sum_{i=0}^{n-1} B_i x^i \in M_2(H_0) \quad .$$

Using the results of Problems (1.36.21) - (1.36.22), Section 1.33.3 and Problem 1.33.20 prove that $F(Z,X) \approxeq F(Z,X)$ if and only if the three matrices in (1.30.4) have the same local invariant polynomials up to the degree $n-1$.

1.37. Notes

Most of the material in Sections 1.1 - 1.8 is standard. See Lang [1967] and van der Waerden [1959] for the algebraic concepts. Consult Gunning-Rossi [1965] and Rudin [1974] for the material concerning the analytic functions. See Kaplansky [1949] for the properties of elementary divisor domains. It is an open problem whether there exists a Bezout domain which is not an elementary divisor domain. Theorem 1.5.6 for $\Omega = \mathbf{C}$ is due to Helmer [1940]. A nice introduction to the theory of algebraic varieties can be found in Lange [1958].

Section 1.9 is standard, e.g. Curtis and Reiner [1962] and MacDuffee [1933]. Most of the content of Section 1.10 is well know, e.g. MacDuffee [1933]. Perhaps Lemma 1.10.3 is not common. The content of Section 1.11 seems to be new since the underlying ring is assumed to be only a Bezout domain. In case that the underlying is **EDD**, i.e., A is equivalent to a diagonal matrix. Theorems 1.11.7 and 1.11.12 are well know. It would be interesting to generalize Theorem 1.11.7 for $D = F[x_1,\ldots,x_p]$, for $p > 2$. The fact that the Smith normal form can be achieved for the elementary divisor domain is due to Helmer [1943]. Consult also Kaplansky [1949].

Most of the results of Section 1.13 are from Friedland [1979b]. It is an open problem whether the results of Problem 1.13.26 hold for any $\Omega \subseteq \mathbf{C}^p$. In case that $\Omega = D(0,\rho)$, $\{\zeta \mid \zeta = (\zeta_1,\ldots,\zeta_p),\ \sum_{j=1}^{p} |\zeta_j|^2 < \rho\}$ the results of Problem 1.13.26 apply. This follows from the Cartan theorem b, e.g. Gunning and Rossi [1965]. This result is due to J. Mather (unpublished).

The exposition of Section 1.14 is close to Gantmacher [1959]. The content of Section 1.15 is standard. Theorem 1.16.7 is well known (e.g. Gantmacher [1959]). Other results of Section 1.16 are not common and some of them may be new. Section 1.17 is standard and its exposition is close to Gantmacher [1959]. Theorem 1.18.5 is probably know of **EDD** (see Leavitt [1948] for the case $D = H(\Omega)$, $\Omega \subseteq \mathbf{C}$). Perhaps it is new for Bezout domains. The results of Section 1.19 are standard. Theorem 1.20.10 appears implicitly in Friedland [1979b]. The exposition of Section 1.21 is close to Gantmacher [1959]. For additional properties of the tensor product of matrices see, for example, Marcus and Minc [1964]. Problem 1.21.28 is close to the results of Faddeev [1966] for necessary and sufficient conditions for the similarity of A and B over $\mathbf{Z}$. See also Guralnick [1980] for a arbitrary integral domain $\mathbf{D}$. The results of Section 1.22 are recent. Theorems 1.22.3 and 1.22.7 are taken from Friedland [1979b]. See Gauger and Byrnes [1977] for a weaker version of Theorem 1.22.7. Some of the results of Section 1.23 seem to be new. Theorem 1.23.3 was taken out of Friedland [1979a]. Theorem 1.24.1 is due to Friedland [1979b].

The exposition of Section 1.25 is close to Gantmacher [1959]. The results of Section 1.26 were inspired by the paper of Rothblum [1980]. The notions of local indices can be found in Friedland-Schneider [1980]. The content of Section 1.27 is standard. Theorem 1.27.9 can be found for example in Wielandt [1967] and Problem 1.27.12 in Gantmacher [1959]. The use of the Cauchy integration formula to study the properties of the analytic functions of A is well accepted. See for example Kato [1976]. The results of Section 1.29 are due to Arnold [1971]. See also Wasow [1977]. See Wasow [1963], [1977] and [1978] for the notions of analytic and pointwise similarity and their importance in theory of differential equations in the neighborhood of singularities. Theorem 1.30.8 in case of one complex variable appears in Friedland [1979b]. Corollary 1.30.11 goes back to Wasow [1963]. Theorem 1.31.1 for simply connected domains is due to Gingold [1978]. See Wasow [1978] for the extension of Theorem 1.31.1 to certain domains $\Omega \subseteq \mathbf{C}^p$. It is shown there that Theorem 1.31.1 fails even for some simply connected domains in $\mathbf{C}^3$.

Theorem 1.32.1 can be found in Kato [1976] or Friedland [1978]. The results of Sections 1.33 - 1.34 were taken from Friedland [1979b]. It is worthwhile to mention the conjecture stated there that $A(x)$ and $B(x)$ are analytically similar over H_0 if the three matrices in 1.30.4 are equivalent over H_0. Theorem 1.35.11 is new. Theorem 1.35.17 is taken form Friedland [1980]. Theorem 1.35.21 and Problem 1.35.24 are due to Motzkin-Taussky [1955]. See also Kato [1976] for a proof of these results using the method of analytic functions. Most of the results of Section 1.36 are taken from Friedland

[1979a-b]. Some results and references on the problem of strict similarity of pairs (A,B) of matrices under the simultaneous similarity can be found in Brenner [1975]. See also Procesi for the extensive treatise on the invariants of pairs (A,B) under the strict similarity.

REFERENCES

Arnold, V. I. [1971], On matrices depending on parameters, Usp. Math. Nauk 26, 101-114.

Berman, A. and Plemmons, R. J. [1979], Nonnegative Matrices in the Mathematical Sciences, New York, Academic Press.

Brenner, S. [1975], Quivers with commutativity conditions and some phenomenology of forms, Lecture Notes Math. 488, Berlin, Springer-Verlag.

Curtis, C. W. and Reiner, I. [1962], Representation Theory of Finite Groups and Associative Algebras, New York, Interscience.

Faddeev, D. K. [1966], On the equivalence of systems of integral matrices, Izv. Akad. Nauk SSSR, Ser. Mat. 30, 449-454.

Friedland, S. [1978], Extremal eigenvalue problems, Bol. Soc. Bras. Mat. 9, 13-40.

____________ [1979a], On pointwise and analytic similarity of matrices, to appear in Israel J. Math.

____________ [1979b], Analytic similarities of matrices, preprint, to appear in the Proceedings of A.M.S. Summer Seminar on Algebraic and Geometric Methods in Linear Systems Theory, Harvard University 1979.

____________ [1980], A generalization of the Motzkin-Taussky theorem, preprint, to appear in Lin. Alg. Appl.

Friedland S. and Schneider, H. [1980], The growth of powers of non-negative matrix, SIAD to appear.

Gantmacher, F. R. [1959], <u>The Theory of Matrices</u>, Vol. I and II, New York, Chelsea Publ. Co.

Gauger, M. A. and Byrnes, C. I. [1977], Characteristic free, improved decidability criteria for the similarity problem, Lin. Multilin. Alg. 5, 153-158.

Gingold, H. [1978], A method of global block diagonalization for matrix-valued functions, SIAM J. Math. Anal.

Gunning, R. . and Rossi, H. [1965], <u>Analytic Functions of Complex Variables</u>, Englewood Cliffs, N.J., Prentice-Hall.

Guralnick, R. M. [1980], A note on the local-global principle for similarity of matries, Letters Lin. Alg. (to appear).

Hardy, G. [1949], Divergent Series, Oxford, Clarendon Press.

Helmer, O. [1940], Divisibility properties of integral functions, Duke Math. J. 6, 345-356.

__________ [1943], The elementary divisor theorems for certain rings without chain condition, Amer. Math. Soc. 49, 225-236.

Kaplansky, I. [1949], Elementary divisors and modules, Trans. Amer. Math. Soc. 66, 464-491.

Kato, T. [1976], Perturbation Theory for Linear Operators, 2d ed., New York, Springer-Verlag.

Kronecker, L. [1890], Algebraische reduction der schaaren bilinear formen, S-B Akad. Berlin, 763-776.

Lang, S. [1958], Introduction to Algebraic Geometry, New York, Interscience.

________ [1967], Algebra, Reading, Mass., Addison-Wesley.

Leavitt, W. G. [1948], A normal form for matrices whose elements are holomorphic functions, Duke Math. J. 15, 463-472.

MacDuffee, C. C. [1933], The Theory of Matrices, Berlin, Springer, 1933.

Marcus, M. and Minc, H. [1964], A Survey of Matrix Theory and Matrix Inequalities, Boston, Prindle, Weber and Schmidt.

Motzkin, T. S. and Taussky, O. [1955], Pairs of matrices with property L. II, Trans. Amer. Math. Soc. 80, 387-401.

Ostrowski, A. [1966], Solutions of Equations and System of Equations, 2nd ed., New York, Academic Press.

Procesi, C. [1976], The invariant theory of $n \times n$ matrices, Advances in Math. 19, 306-381.

Rothblum, U. [1980], Expansions of sums of matrix powers,preprint, SIAM Review (to appear).

Rudin, W. [1974], Real and Complex Analysis, 2nd ed., New York, McGraw Hill.

Van der Waerden, B. L. [1959], Algebra, vol. I and II, Berlin, Springer.

Wasow, W. [1963], On holomorphically similar matrices, J. Math. Anal. Appl. 4, 202-206.

________ [1977], Arnold's canonical matrices and asymptotic simplification of ordinary differential equations, Lin. Alg. Appl. 18, 163-170.

________ [1978], Topics in Theory of Linear Ordinary Differential Equations Having Singularities with Respect to a Parameter, IRMA, Universite Louis Pasteur, Strasbourg.

Weierstrass, K. [1867], Zur theorie der bilinearen und guadratischen formen, Monatsh. Akad. Wiss. Berlin, 310-338.

Wielandt, H. [1967], Topics in Analytic Theory of Matrices, Lecture Notes, Univ. of Wisconsin-Madison.

SF/jvs

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER #2082	2. GOVT ACCESSION NO. AD-A089635	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Spectra. Theory of Matrices I. General Matrices		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Shmuel Friedland		8. CONTRACT OR GRANT NUMBER(s) DAAG29-75-C-0024 DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 1 - Applied Analysis
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE May 1980
		13. NUMBER OF PAGES 174
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Analytic similarity of matrices, canonical forms.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This is the first of our survey reports on the spectral theory of matrices. The report is self contained from the matrix point of view. The main subject of this paper is the theory of analytic similarity of matrices. This topic of matrix theory is very important in the study of systems of ordinary differential equations having singularities either in time or a parameter.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)