8-1-1979

Linear groups

Marion Stinson
Atlanta University

LINEAR GROUPS

A THESIS

SUBMITTED TO THE FACULTY OF ATLANTA UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE

BY

MARION STINSON

DEPARTMENT OF MATHEMATICS

ATLANTA, GEORGIA

AUGUST 1979

ABSTRACT

Stinson, Marion B.S. Stillman College, 1976

Dr. Johnny L. Houston, Advisor

Master of Science degree conferred August, 1979

Linear groups represent an important study of Lie groups. These locally compact groups are important in the study of representation theory. They are noticeable in the study of vector spaces and topological spaces. Applications of linear groups can be seen in their relation to inner products and normal matrices.

The basic concepts of linear groups are investigated to give a foundation for further, more advanced, study in this field. Special attention is paid to unitary, orthogonal and hermitian matrices. Inner products, congruence and conjunctive equivalence in relation to the above matrices are also discussed.

ACKNOWLEDGEMENTS

I am truly grateful to the staff of the department of Mathematics for their unselfishness and consideration. I am especially grateful to Dr. Johnny L. Houston, my thesis advisor, for his patience, sincere attention, and expert guidance in preparation of this thesis.

TABLE OF CONTENTS

CHAPTER I

LINEAR GROUPS AND THEIR BASIC PROPERTIES

The study of linear groups is one of the main objects of Lie group theory. Linear groups give a strong starting point for representation theory, since in this case the group is a representation of itself. Linear groups make up a subclass of locally compact groups. The word "linear" is taken from linear algebra in which the study of matrices is the main theme.

The theory of linear groups began in 1876 with a paper by F. Klein on invariants. It was greatly extended by C. Jordan who used it to study similar problems. The theory developed repidly in the quarter-century around 1900. Mathematicans such as Frobenius, Schur, Blichfeldt and Burnside did work in this area during this time. This period is sometimes described as the "classical" period. Although the theory was used for studying finite groups, the linear groups considered were almost always over subfields of the complex numbers. An exception to this was work done by L.E. Dickson. The modern period began around 1950 when it was discovered how linear groups arise in the analysis of the structure of abstract groups.

The purpose of this paper is to develop the linear group and to study extensions of the group. We will look at some of the basic properties and theorems that describe this group, as well as, its relation to vector spaces and topological spaces. We will also look at some of the applications of linear groups as they relate to inner products and normal matrices.

The set of all nonsingular linear transformations of dimension "n" under the operation of multiplication has the following four distinct properties:

a) The set is closed under multiplication; that is, if A and B are nonsingular transformations, then so is their product C=AB;

b) The operation of multiplication of transformations is associative, that is, A(BC)=(AB)C;

c) The set contains an identity transformation I such that AI=IA=A for every A;

d) Every nonsingular transformation A has a unique inverse transformation A^{-1} such that $A^{-1}A=AA^{-1}=I$.

Thus the set of all nonsingular linear transformations of dimension "n" is a group denoted by $G_n(C)$ and called the full linear group of dimension "n".

We will now show that certain subsets of the full linear group $G_n(C)$ and not just the whole group $G_n(C)$, also form groups.

Example. Consider the set of all rotations of the plane L_2 about the origin of coordinates. This set is a group, since the product of any two rotations is a rotation, and the same is true of the transformation inverse to any rotation. To verify this by a formal calculation, let

$$A=\begin{pmatrix}\cos\alpha & -\sin\alpha\\ \sin\alpha & \cos\alpha\end{pmatrix}, \qquad B=\begin{pmatrix}\cos\beta & -\sin\beta\\ \sin\beta & \cos\beta\end{pmatrix}$$

be the matrices corresponding to rotations through the angles α and β, respectively. Then the matrix

$$AB=\begin{pmatrix}\cos\alpha\cos\beta-\sin\alpha\sin\beta & -\cos\alpha\sin\beta-\sin\alpha\cos\beta\\ \sin\alpha\cos\beta\ \ \cos\alpha\sin\beta & -\sin\alpha\sin\beta\ \ \cos\alpha\cos\beta\end{pmatrix}$$
$$=\begin{pmatrix}\cos(\alpha+\beta) & -\sin(\alpha+\beta)\\ \sin(\alpha+\beta) & \cos(\alpha+\beta)\end{pmatrix}$$

corresponds to a rotation through the angle $-\alpha$, while the matrix

$$A^{-1}=\begin{pmatrix}\cos\alpha & \sin\alpha\\ -\sin\alpha & \cos\alpha\end{pmatrix},$$ corresponding to rotations

through the angle $-\alpha$, clearly satisfies the relation $A^{-1}A=AA^{-1}=I$.

As another example, suppose the transformation A with matrix A does not change the absolute value of the parallelepiped constructed on an arbitrary triple of vectors. Then the absolute value of the determinant A equals unity, that is, $|A| = \pm 1$. The set of all transformations of this type forms a group called the unimodular group. In turn, the set of all transformations which preserve both the volume and the orientation of triples of vectors forms a subgroup of the unimodular group. For such transformations, we have $|A| =1$.

Next we will consider another important subgroup of the full linear group, the subgroup of orthogonal transformations.

(1.1) DEFINITION. A square matrix A with the property

$A^{-1}=A^T$ is said to be an <u>orthogonal</u> <u>matrix</u>.

The determinant of an orthogonal matrix equals ± 1. We now give an algebraic proof of this fact. It follows from the formula $A^TA=I$ and the theorem on multiplication of determinants that $|A^TA| = |A^T||A| = |A|^2 = 1$, since $|A^T| = |A|$ and $|I| = 1$. But then obviously $|A| = \pm 1$.

Another importnat subgroup of the full linear group is the group of unitary matrices.

(1.2) DEFINITION. A matrix A such that $AA^*=I$ is said to be a unitary matrix, where $A^*=$ the conjugate transpose of A.

We now present the general linear group and some of its interesting subgroups, from the perpective of **abstract** algebra:

NOTATION	COMMON NAME	DESCRIPTION
$G_n(C)$	General or Full linear group	The set of all regular nxn matrices (matrices with inverses). They form a group.
$U_n(C)$	Unitary group	A $G_n(C)$, $\bar{A}=A^*=A^{-1}$
$O_n(C)$	Orthogonal group	A $G_n(C)$, $A^{-1}=A^T$
$SL_n(C)$	Special Full	A $G_n(C)$, $\|A\|=1$
$SU_n(C)$	Special Unitary	A $G_n(C)$, $\|A\|=1$
$SO_n(C)$	Special Orthogonal	A $G_n(C)$, $\|A\|=1$

In chapter I we have dealt with the classical general linear group and some of its basic properties. Among the matrices highlighted were the unitary and **orthogonal matrices**, as some of its noteworthy matrices of interest.

Chapter II deals with linear groups as related to vector spaces. In this chapter we will look at vector spaces, their properties, their relationship to matrices and their relationship to the linear group.

Chapter III is concerned with linear groups as a topological group. We will look at some properties of topological spaces and the relationship between linear groups and topological spaces.

Chapter IV deals with the theory of inner product spaces. Basic definitions and theorems will be given to aid us in the development of the theory.

Chapter V will then deal with some solutions of problems involving linear groups and applications in inner product spaces and normal matrices.

CHAPTER II

LINEAR GROUPS AND VECTOR SPACES

Let K denote the field R or C of real or complex numbers. For a positive integer n, K^n denotes the direct product of K as a factor n-times, i.e., $x \in K^n$ if and only if, $x=(x_1,\ldots,x_n)$, where $x_i \in K$, $1 \leq i \leq n$. With the usual definition of addition and scalar multiplication, K^n is a vector over the field K. More specifically, the addition and scalar multiplication are defined as follows: For each pair $x,y \in K^n$,

$x+y=(x_1+y_1,\ldots,x_n+y_n)$, where $x=(x_1,x_2,\ldots,x_n)$ and $y=(y_1,y_2,\ldots,y_n)$; and for any $\lambda \in K$ and $x \in K^n$,

$$\lambda x=(\lambda x_1, \lambda x_2,\ldots, \lambda x_n).$$

(2.1) DEFINITION. A mapping f of K^n into itself is said to be an <u>endomorphism</u> of K^n if:

(i) $f(x+y)=f(x)+f(y)$, $x,y \in K^n$;

(ii) $f(\lambda x)=\lambda f(x), \lambda \in K$.

It is convenient to replace the endomorphisms of K^n by certain objects called matrices. For this we have the following:

Let a_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n$, be mn elements of K. The array

$$\begin{bmatrix} a_{11}, & a_{12}, & \ldots, & a_{1n} \\ a_{21}, & a_{22}, & \ldots, & a_{2n} \\ \ldots, & \ldots, & \ldots, & \ldots \\ \ldots, & \ldots, & \ldots, & \ldots \\ a_{m1}, & a_{m2}, & \ldots, & a_{mn} \end{bmatrix}$$

is called a matrix of m rows and n columns, and is denoted by (a_{ij}), $1 \leq i \leq m$, $1 \leq j \leq n$. A matrix of m rows and n columns is usually called a matrix of order mxn. If m=n (i.e., the number of columns equals the number of rows) the matrix (a_{ij}), $1 \leq i \leq m$, $1 \leq j \leq n$, is called a square matrix of order mxm or nxn.

Since the study of linear groups is the study of special kinds of matrices and the study of matrices involves greatly the study of vectors, we denote our attention in this chapter to some of the fundamental properties of matrices.

A vector space is an algebraic system based upon two sets F and V of elements. The elements of F are called scalars and constitute an algebraic system called a field. The elements of V are called vectors. Usually we take F to be the set R of real numbers or the set C of complex numbers. Both R and C are examples of an algebraic system that is called a field. This leads to the following definition.

(2.2) DEFINITION. A <u>vector</u> <u>space</u> over F is a nonempty set V such that (A) There is defined an operation of addition (that is, for all $x, y \in V$, $x+y \in V$). (B) There is defined an operation of scalar multiplication (that is, for all $\alpha \in F$,

$x \in V$, $\alpha x \in V$) satisfying the following axioms:

A1. Addition is commutative. For all vectors $x, y \in V$, $x+y=y+x$.

A2. Addition is associative. For all $x, y, z \in V$, $(x+y)+z=x+(y+z)$.

A3. Existence of 0. There exists a vector $0 \in V$ such that for all $x \in V$, $x+0=x=0+x$.

A4. Existence of additive inverse. For each vector $x \in V$, there is a vector $y \in V$ such that $x+y=0=y+x$.

A5. For all $\alpha \in F$, $x, y \in V$, $\alpha(x+y)=\alpha x+\alpha y$.

A6. For all $\alpha, \beta \in F$, $x \in V$, $(\alpha+\beta)x=\alpha x+\beta x$.

A7. For all $\alpha, \beta \in F$, $x \in V$, $(\alpha\beta)x=\alpha(\beta x)$.

A8. For all $x \in V$, $1x=x$.

(2.3) DEFINITION. A vector space E over the field K is said to be an algebra if the following conditions hold: For each pair $x, y \in E$ the product xy is defined and belongs to E. The product is associative and the distributive laws hold: $x(y+z)=xy+xz$, $(x+y)z=xz+yz$, $\lambda(xy)=(\lambda x)y=x(\lambda y)$ for all $x, y, z \in E$ and $\lambda \in K$. If in addition, $xy=yx$ for all $x, y \in E$ then the algebra is said to be commutative.

(2.4) PROPOSITION. The set E of all endomorphisms of K^n forms an algebra.

Proof. Let f, g be two endomorphisms of K^n. Define $f+g=h$ so that $h(x)=f(x)+g(x)$ for each $x \in K^n$. Then clearly h is an endomorphism of K^n. If 0 denotes the zero-endomorphism, i.e.,

$O(x)=O \in K^n$ for each $x \in K^n$ and $(-f)(x)=-f(x)$, then it is easily seen that E is an abelian additive group. Furthermore, if, for $\lambda \in K$, $\lambda f(x)=\lambda(f(x))$ for each $x \in K^n$, then it follows that E is a vector space. For each pair $f, g \in E$, define $fg(x)=f(g(x))$. Then with the composition as the product of elements in E, one verifies that E is a commutative algebra. Q.E.D.

(2.5) DEFINITION. Let V be a vector space and let W be a nonempty subset of V. We call W a <u>subspace</u> of V if and only if W is a vector space under the laws of addition and scalar multiplication defined on V.

(2.6) DEFINITION. Let V be a vector space. A vector of the form $a_1X_1+a_2X_2+\dots+a_rX_r$, where $X_i \in V$ and $a_1, a_2, \dots, a_r$ are arbitrary real numbers, is called a <u>linearly combination</u> of the vectors $X_1, \dots, X_r$.

Let us suppose that a vector space V is given in advance. If it is possible to find r vectors $X_1, X_2, \dots, X_r$ of V such that every vector X of V can be written as a linearly combination

$$X=a_1X_1+a_2X_2+\dots+a_rX_r,$$

we say that the vectors $X_1X_2, \dots, X_r$ are a <u>set of generators</u> of V, or that V is <u>generated</u> (spanned) by the vectors $X_1, X_2, \dots, X_r$.

We use the notation $V=L\{X_1, X_2, \dots, X_r\}$ to denote that the vector space V is generated by the vectors $X_1, X_2, \dots, X_r$.

Example. Let V be the subspace of $V_n(R)$ (where $V_n(R)$ is the set of all vectors with components: $X=(X_1,X_2,\ldots,X_n)$ over the field of real numbers) consisting of all vectors whose fourth component is zero. Every vector in V has the form $[a_1,a_2,a_3,0]$, and by using the properties of vector operations we have

$$[a_1,a_2,a_3,0] = a_1[1,0,0,0] + a_2[0,1,0,0] + a_3[0,0,1,0].$$

Hence,

$$V=L\{[1,0,0,0]\ ,\ [0,1,0,0]\ ,\ [0,0,1,0]\}.$$

Since $[1,0,0,0]= [1,1,1,0] - [0,1,1,0]$, we may also write

$$[a_1,a_2,a_3,0] = a_1\ [1,1,1,0] -a_1\ [0,1,1,0] + a_2[0,1,0,0] + a_3[0,0,1,0]$$

or $$V=L\{[1,1,1,0]\ ,\ [0,1,1,0]\ ,\ [0,1,0,0]\ ,[0,0,1,0]\}.$$

(2.7) DEFINITION. The vectors $X_1,X_2,X_3,\ldots,X_n$ of a vector space V are said to be <u>linearly dependent</u> if there exist real numbers $a_1,a_2,\ldots,a_n$, not all zero, such that $a_1X_1+a_2X_2+\ldots+a_nX_n=0$.

If the vectors $X_1,X_2,\ldots,X_n$ of a vector space V are said to be <u>linearly independent</u>, there exist real numbers $a_1,a_2,\ldots,a_n$, each equal to zero, such that $a_1X_1+a_2X_2+\ldots+a_nX_n=0$.

(2.8) DEFINITION. A <u>basis</u> for a vector space V is a linearly independent set that spans V. If V contains a finite basis, V is said to be <u>finite-dimensional</u>, otherwise, V is said to be <u>infinite- dimensional</u>.

The dimension of a vector space could have been defined as the maximum number of linearly independent vectors of V. In particular, $V_n(R)$ is an n-dimensional vector space.

Example. The vectors $[1,1,0,0]$, $[0,0,1,0]$, $[0,0,0,1]$ are easily shown to be linearly independent and they form a basis for the subspace V of $V_4(R)$ consisting of all vectors for which $X_1=X_2$. Thus the dimension of V is 3.

The following theorem provides a relation between linearly independent vectors, bases, and dimension of a vector space V.

(2.9) THEOREM. If $Y_1,\ldots,Y_n$ are linearly independent vectors of V, then either $V=L(Y_1,Y_2,\ldots,Y_n)$ or vectors $X_{n+1},\ldots,X_r$ contained in V may be found such that the vectors $Y_1,\ldots,Y_n; X_{n+1},\ldots,X_r$ form a basis for V.

Proof. First we know that there is a set of linearly independent vectors $X_1,\ldots,X_n$ such that $V=L(X_1,\ldots,X_r)$. Now consider $V=L\{Y_n,X_1,\ldots,X_r\}$ and reduce to a linearly independent basis. Now remove one X_i. Repeated application of this process gives a basis $V=L\{Y_1,\ldots,Y_n, X_{n+1},\ldots,X_r\}$, where $X_{k+1},\ldots,X_r$ are a subset of the X_i vectors , or it may happen that K=r. In this latter case $V=L\{Y_1,\ldots,Y_k\}$.

In other words, we say that any set of linearly independent vectors of V may be extended to a basis of V.

(2.10) DEFINITION. Let $\chi=(X^1,\ldots,X^n)$ be a basis for V. If $z\in V$ and $z=\sum_{i=1}^{n} x^i\gamma_i=\chi_c$, where

$$c=\begin{bmatrix}\gamma_1\\ \vdots\\ \gamma_n\end{bmatrix},$$

then we call c the representation of z with respect to the basis $\mathcal{X}$.

(2.11) THEOREM. Let $\mathcal{X}=(X^1,\ldots,X^n)$ be a basis for V. If $z \in V$, let Rz be the representation of z with respect to $\mathcal{X}$. Then

(1) $R(z^1+z^2)=Rz^1+Rz^2$.

(2) $R(\alpha z)=\alpha Rz$, if $\alpha \in F$.

(3) $z^1=z^2$ if and only if $Rz^1=Rz^2$.

(4) If $c \in F_{n,1}$, then there is a $z \in V$ such that $c=Rz$.

CHAPTER III

LINEAR GROUPS AND TOPOLOGICAL SPACES

In this chapter we shall look at linear groups as they relate to topological spaces. The properties associated with topological spaces hold for linear groups, since they are locally compact groups.

(3.1) DEFINITION. A <u>group</u> G is a nonempty set together with an operation (which we denote here by *) satisfying the following:

(1) For all $a,b \in G$, $a*b \in G$ (* is closed).

(2) For all $a,b,c \in G$, $a*(b*c)=(a*b)*c$ (* is associative).

(3) There is an element $e \in G$ such that for all $a \in G$, $a*e=e*a=a$ (G contains an identity element for *).

(4) If $a \in G$, there is an element $a^{-1} \in G$ such that $a*a^{-1}=a^{-1}*a=e$ (each element in G has an inverse in G relative to *).

(3.2) DEFINITION. Let X be a set. A <u>topology</u> on X is a collection of subsets of X that satisfies the following three conditions:

1) $\emptyset$ and X belong to $\mathfrak{J}$.

2) If $\mathfrak{a} \subset \mathfrak{J}$, then $\bigcup \mathfrak{a} \in \mathfrak{J}$.

3) If $\mathfrak{B}$ is finite and $\mathfrak{B} \subset \mathfrak{J}$, then $\bigcap \mathfrak{B} \in \mathfrak{J}$.

The pair $(X,\mathfrak{J})$ is called a <u>topological space</u>, and the subsets of X that belong to $\mathfrak{J}$ are called <u>open sets</u> in $(X, \mathfrak{J})$.

(3.3) DEFINITION. A topological space X is a Hausdorff space if x and y are two distinct points in the space and there exist open sets U and V such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$.

(3.4) DEFINITION. For a subset S of a topological space, whenever a collection of open sets U of the space is such that $S \subset \bigcup U$, we call U an open cover of S, and any subcollection of U whose union also contains S is called an open subcover of S contained in U.

(3.5) DEFINITION. A subset A of a topological space X is compact if every open cover of A contains a finite subcover.

(3.6) DEFINITION. A topological space is locally compact if each point in the space has a compact neighborhood.

(3.7) DEFINITION. A topological space is said to be 2nd countable (or is said to satisfy the second axiom of countability) if the topology on the space can be generated by a countable basis.

(3.8) NOTATION. $M_n(K)$ denotes the set of all nxn matrices with coeficients in K.

(3.9) PROPOSITION. $M_n(K)$ is a Hausdorff topological group which is locally compact, noncompact, and satifies the second axiom of countability.

Proof. First of all, we shall assign a topology on $M_n(K)$ by identifying it with K^m for some m. Let the coefficients of each matrix $A=(a_{ij})$ in $M_n(K)$, be arranged in a definite order. In that case, A can be regarded as a point in K^{n^2}. Let f be the mapping that to each A in $M_n(K)$, assigns the point of K^{n^2} obtained by arranging the coefficients of A in the fixed order. Then it is easy to see that f is 1:1 and onto. Let $M_n(K)$ be endowed with the same topology as that of K^{n^2}. In other words, a set T in $M_n(K)$ is open if, and only if, f(T) is an open subset of K^{n^2} under the usual Euclidean topology. Since K^{n^2} is an additive abelian topological group satisfying all the properties mentioned in the Proposition, so is $M_n(K)$. Q.E.D.

CHAPTER IV

LINEAR GROUPS AND INNER PRODUCTS

(4.1) INNER PRODUCTS

The notion of an inner product has many applications in mathematics and physics. In this paper we study inner product spaces in connection to unitary, orthogonal and hermitian matrices. These kinds of matrices are all representations of inner products, as we shall see later.

We first look at the general idea of inner product. Suppose we take two vectors:

$$x=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{and } y=\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

in the plane. By a standard result in analytic geometry it follows that x and y are orthogonal (at right angles) if and only if $x_1y_1+x_2y_2=0$. Here $x_1y_1+x_2y_2$ can be considered as an inner product between the vectors x and y that vanishes when x and y are orthogonal. Suppose that $x,y \in R_{n,1}$. We shall define an inner product in $R_{n,1}$.

(4.1.1) DEFINITION. Let $x,y \in R_{n,1}$. Then

$$x^T y=x_1y_1+x_2y_2+\dots+x_ny_n,$$

and x^Ty is called an inner product of x and y.

We also wish to define an inner product in a space over the complex numbers.

(4.1.2) DEFINITION. Let $x, y \in C_{n,1}$. Then we define

$$x^*y = \bar{x}_1 y_1 + \bar{x}_2 y_2 + \ldots + \bar{x}_n y_n,$$

where $x^* = [\bar{x}_1, \ldots, \bar{x}_n]$, to be an inner product of x and y.

From now on we shall state the analogous results for real spaces without proof. This can be easily seen how one reduces the real results from the complex since for real numbers x_i, $\bar{x}_i = x_i$.

If $\alpha = \beta + i\gamma$ is a complex number, where β and γ are real, then $\alpha\bar{\alpha} = \beta^2 + \gamma^2 \geq 0$. As usual we write $|\alpha|^2 = \alpha\bar{\alpha}$, where $|\alpha| \geq 0$. Clearly $|\alpha| = 0$ if and only if $\alpha = 0$. Let us examine some of the properties of our inner product.

(4.1.3) PROPOSITION. In $C_{n,1}$:

(1) $x^*x \geq 0$ and $x^*x = 0$ if and only if $x = 0$.

(2) $x^*y = \overline{y^*x}$.

(3) $x^*(\alpha y) = \alpha(x^*y)$.

Proof(1) If

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},$$

then $x^*x = \bar{x}_1 x_1 + \ldots + \bar{x}_n x_n = |x_1|^2 + \ldots + |x_n|^2 \geq 0$ and $|x_1|^2 + |x_n|^2 = 0$ if and only if each $x_i = 0$.

(2) We observe that $x^*y = \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n = \overline{\bar{y}_1 x_1 + \ldots + \bar{y}_n x_n} = \overline{y^*x}$.

(3) We observe that (3) is just a special case of properties on matrix multiplication. Q.E.D.

We now look at the notion of an inner product in an arbitrary vector space.

(4.1.4) DEFINITION. Let V be a vector space over the complex numbers. Suppose with each pair of vectors $x, y \in V$ there is associated a scalar (x,y). Then $(\, , \,)$ is called an inner product on V if and only if

(1) For all $x \in V$, $(x,x) \geq 0$, and $(x,x)=0$ if and only if $x=0$.

(2) $(x,y)=\overline{(y,x)}$.

(3) (a) $(x,y_1+y_2)=(x,y_1)+(x,y_2)$.

(b) $(x,\alpha y)=\alpha(x,y)$.

(4.1.5) Example. Let $C_2[t] = \{\alpha_0+\alpha_1 t \mid \alpha_0, \alpha_1 t \mid \alpha_0, \alpha_1 \in C\}$.

For $p(t)$, $q(t) \in C_2[t]$ define

$$(p(t),q(t))=\int_0^1 \overline{p(t)}q(t)\,dt.$$

We claim that $(\, , \,)$ is an inner product on $C_2[t]$. We shall not give the formal verification, but we shall illustrate (1) and (2) by an example. So let $p(t)=1+it$. Then

$$(p(t),p(t))=\int_0^1 (1-it)(1+it)\,dt=\int_0^1 (1+t^2)\,dt= -\frac{4}{3},$$

which is positive. Now let $q(t)=t$. Then

$$(p(t),q(t))=\int_0^1 (1-it)t\,dt=\int_0^1 (t-it^2)dt= \frac{1}{2} - \frac{i}{3}$$

$$(q(t),p(t))=\int_0^1 t(1+it)dt=\int_0^1 (t-it^2)dt= \frac{1}{2} + \frac{i}{3}= \overline{(p(t),q(t))}.$$

(4.2) REPRESENTATION OF LINEAR PRODUCTS

Let $V=C_{n,1}$. We already know that there is one inner product on $C_{n,1}$, that is, $(x,y)=x^*y$. It is interesting to investigate whether there are other inner products in $C_{n,1}$. Let us suppose that H is an nxn matrix, and let us see under what conditions

(4.2.1) $(x,y)=x^*Hy$

is an inner product.

It is easy to verify that for every matrix H, (x,y) defined by (4.2.1) satisfies conditions 3(a) and (b). By Proposition (4.1.3) (2) $\overline{y^*(Hx)}=(Hx)^*y$, whence $\overline{(y,x)}=\overline{y^*Hx}=x^*H^*y$. But $(x,y)=x^*Hy$. Thus (2) of (4.1.1) will be satisfied if $H=H^*$.

If $H=H^*$, then $\overline{x^*Hx}=x^*H^*x=x^*Hx$, and so, for all x, x^*Hx is real. For (y,x) to satisfy (1) of (4.1.4), we must require in addition that for $x\neq 0$, $x^*Hx>0$. In order to turn the remarks we have made into a respectable proposition, we make some definitions.

(4.2.2) DEFINITIONS.

(1) An nxn matrix H is called <u>hermitian</u> if and only if $H=H^*$.

(2) A hermitian matrix is called <u>positive definite</u> (hermitian) if and only if for all nonzero vectors $x, x^*Hx>0$.

(3) A heritian matrix is called <u>positive semidefinite</u> if and only if for all x, $x^*Hx\geq 0$.

Note that if H is positive definite, then $x^*Hx\geq 0$, since either $x\neq 0$ or $x=0$. Hence a positive definite matrix is positive semidefinite.

Example. Let

$$H=\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} \text{ and } x=\begin{bmatrix}x_1\\ x_2\end{bmatrix}.$$

Then $x^*Hx=2\bar{x}_1x_1+\bar{x}_1x_2+\bar{x}_2x_1+2\bar{x}_2x_2=2(x_1+\frac{1}{2}x_2)(x_1+\frac{1}{2}x_2)+\frac{3}{4}x_2x_2=$ $2|x_1+\frac{1}{2}x_2|^2+\frac{3}{2}|x_2|^2$, and it is easily seen that this is positive unless $x_1=x_2=0$. Thus H is a positive definite matrix.

(4.2.3) PROPOSITION. On $C_{n,1}$ let $(x,y)=x^*Hy$, where H is a positive definite matrix. Then (,) is an inner product on $C_{n,1}$.

(4.2.4) LEMMA. Suppose H and K are nxn matrices such that for all $x,y \in C_{n,1}$, $x^*Hy=x^*Ky$. Then H=K.

Proof. If $x=e^i$ and $y=e^i$, then $x^*Hy=h_{ij}$, and $x^*Ky=k_{ij}$. Hence $h_{ij}=k_{ij}$. Since this is true for all $i,j=1,\dots,n$, H=K.

We can interpret this lemma in terms of inner products. Thus both $(x,y)_1=x^*Hy$ and $(x,y)_2=x^*Ky$ may yield inner products. But our lemma asserts that these are equal only if H=K.

We obtain a generalization of Proposition (4.2.3). Let V be a finite-dimensional vector space and let $X=(X_1,\dots X_n)$ be a basis for V. We recall that (2.1) we wrote $z=\chi_c=\sum_{i=1}^{n} x^i r_i$, where

$$c=\begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix},$$

We also know that the mapping R given by c R_z is an isomorphism on V onto $C_{n,1}$; that is, R satisfies the conditions of Theorem (2.8). Q.E.D.

(4.2.5) PROPOSITION. Let H be an nxn positive definite matrix, and let V be an n-dimensional vector space. Let X be a basis for V. If z=Xc and y=Xb, then (,) defined by (z,y)=

c^*Hb is an inner product on V.

Proof. Although we shall prove this proposition in detail, we claim that this result is really obvious because R is an isomorphism, and, by Proposition (4.2.3), ((c,b)) defined by $((c,b))=c^*Hb$ yields an inner product on $C_{n,1}$. We turn to the proof. Thus to verify (1) of (4.1.4), note that $(z,z)=((Rz,Rz))$ ≥ 0 and is zero if and only if $Rz=0$. But Rz 0 if and only if $z=0$. To verify (2) note that $(z,y)=((Rz,Ry))=(y,z)$. For (3) we we use the linearity property of (2.6), and the details are obvious. Q.E.D.

We shall now show that every inner product is of the form given in this proposition.

(4.2.6) PROPOSITION. Let V be an n-dimensional vector space, and let (,) be an inner product on V. If $X=(x^1,\ldots,x^n)$ is a basis for V, let H be the matrix $h_{ij}=(x^i,x^j)$, $i,j=1,\ldots,n$. Then H is positive definite, and if $z=Xc$ and $y=Xb$, then $(x,y)=$ c^*Hb.

Proof. Since $z=\sum_{i=1}^{n} x^i \gamma_i$ and $y=\sum_{j=1}^{n} x^j \beta_j$, it follows from the axioms (1), (2), and (3) of (4.1.4) that

$$(z,y)=\left(\sum_{i=1}^{n} x^i \gamma_j, \quad \sum_{j=1}^{n} x^j \beta_j\right) \sum \bar{\gamma}_i (x_i, x^j)\beta_j = c^*Hb,$$

and the last part of the proposition is proved.

To show that H is hermitian observe that

$$h_{ji}=(x^j,x^i)=\overline{(x^i,x^j)}=h_{ij}.$$

Suppose that $c\neq 0$; then $z=Xc\neq 0$, whence $c^*Hc=(z,z)>0$. This proves that H is positive definite. Q.E.D.

(4.2.7) COROLLARY. With the same notation as in Proposition (4.2.6), there is a unique positive definite matrix H such that $(z,y)=c^*Hb$.

Proof. The existence of a positive definite matrix H is asserted by Proposition (4.2.6), and the uniqueness follows from (4.2.4).

As an example, consider the space $C_2[t]$ with the inner product as defined in (4.1.5). Let $p(t)=\gamma_0+\gamma_1 t$, and $q(t)=\beta_0+\beta_1 t$. Then

$$(p(t),q(t))=\int_0^1 \overline{p(t)}\,q(t)\,dt=\int_0^1 \bar{\gamma}_0\beta_0+(\bar{\gamma}_0\beta_1+\gamma_1\beta_0)t + \bar{\gamma}_1\beta_1 t^2 dt=\frac{\gamma\beta_0+\gamma_0\beta_1+\gamma_1\beta_0}{2}+\frac{\gamma_1\beta_1}{3}.$$

Let

$$c=\begin{bmatrix}\gamma_0\\ \gamma_1\end{bmatrix} \qquad b=\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} \qquad H=\begin{bmatrix}1 & 1/2\\ 1/2 & 1/3\end{bmatrix}.$$

Then also

$$c^*Hb=\frac{\gamma_0\beta_0+\bar{\gamma}_0\beta_1+\bar{\gamma}_1\beta_0}{2}+\frac{\bar{\gamma}_1\beta_1}{3}.$$

Hence $(p(t),q(t))=c^*Hb$. Q.E.D.

Now we have represented inner products as matrices. More formally we have:

(4.2.8) DEFINITION. If X is a basis for V, and $H=[h_{ij}]$, $h_{ij}=(x^i,x^j)$, H will be called the <u>representation</u> of (,) with respect to the basis X. We may write $H=P_X(\ ,\ )$.

Again consider the space $C_2[t]$ of (4.1.5) with the inner product defined there. If we take for the basis $X=(1,t)$, then by the above compution we see that

$$P_X(\ ,\ ) = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}.$$

We can now reformulate Proposition (4.2.6) thus: If $c=R_Xz$, $b=R_Xy$, and $H=P_X(\ ,\ )$, then $(z,y)\quad c^*Hb$.

Let us for a moment consider all possible inner products that can be defined on an n-dimensional vector space V. Let X be a basis for V. Proposition (4.2.6) assures us that to each inner product (,) there corresponds an nxn positive definite matrix $H=P_X(\ ,\ )$ satisfying $(z,y)=(Rz)^*H(Ry)$. Lemma (4.2.4) asserts that there is only one matrix H for which $(z,y)= (Rz)^*H(Ry)$.

Conversely, Proposition (4.2.5) asserts that we can define an inner product by putting $(z,y)\quad (Rz)^*H(Ry)$. Thus there is a one-to-one correspondence between inner products on V and positive definite matrices. Observe that the correspondence depends upon the basis.

To state the analogous results for inner products on real spaces we require some definitions.

(4.2.9) DEFINITIONS.

(1) A real nxn matrix S is called <u>real symmetric</u> if and only if $S=S^T$.

(2) A real symmetric nxn matrix is called <u>real positive definite</u> if and only if $x^TSx=0$ for all $x>0$ in $R_{n,1}$.

(3) A real symmetric nxn matrix is called <u>real positive semidefinite</u> if and only if $x^TSx\geq 0$ for all $x \in R_{n,1}$.

Note that a real hermitian matrix, $S^*=S^T$.

We state without proof the following theorem.

(4.2.10) THEOREM. Let V be an n-dimensional real vector space. Let $X=(x^1,\ldots,x^n)$ be a basis for V. If $z=Xc$, let $Rz=c$. If S is any real positive definite matrix, then (,) defined by $(z,y)=(Rz)^T S(Ry)$ is an inner product on V. Conversely, if (,) is an inner product on V, then there is a unique real positive definite matrix S for which $(z,y)=(Rz)^T S(Ry)$.

We shall now see how the representation of an inner product changes with a change of bases.

(4.2.11) PROPOSITION. Let V be a finite-dimensional vector space, and let X and $Y=XP$, where P is a nonsingular matrix, be bases for V. If (,) is an inner product on V and H and K are the representations of this inner product with respect to X and Y, respectively, then

$$K=P^*HP.$$

Proof. By definition, $k_{ij}=(y^i,y^j)$. Hence

$$k_{ij}=((XP))_{ij},(XP)_j))$$

$$\Big(\sum_{h=1}^{n} x^h p_{hi},\quad \sum_{k=1}^{n} x^k p_{kj}\Big)$$

$$\sum_{h=1}^{n}\quad \sum_{k=1}^{n} \overline{p}_{hi}(x^h,x^k)p_{kj}=(P^*)_{i*}HP_{*j},$$

since the ith row of P^* is $\left[\overline{p}_{1i},\overline{p}_{2i},\ldots,\overline{p}_{ni}\right]$. Hence $K=P^*HP$, as claimed.

Note that the family $Y=(1+t,1-t)$ is a basis for $C_2[t]$. Then $Y=XP$, where $X=(1,t)$ and

$$P=\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}.$$

It follows from (4.2.11) that

$$P_y(\ ,\ )\quad P^*HP=\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}1 & 1/2\\ 1/2 & 1/3\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}=\begin{bmatrix}7/3 & 2/3\\ 2/3 & 1/3\end{bmatrix}.$$

As another example let $V=C_{3,1}$, and let

$$H=\begin{bmatrix}6 & 5 & 5\\ 5 & 6 & 5\\ 5 & 5 & 6\end{bmatrix}\quad \text{where } P=\begin{bmatrix}2 & 1 & 1\\ 1 & 2 & 1\\ 1 & 1 & 2\end{bmatrix}.$$

H is in fact positive definite. Thus by (4.2.5) H defines an inner product on $C_{3,1}$ with respect to the basis χ see (4.2.8)

$$X=\begin{bmatrix}2\\1\\1\end{bmatrix},\ \begin{bmatrix}1\\2\\1\end{bmatrix},\ \begin{bmatrix}1\\1\\2\end{bmatrix}.$$

Let $=(e_{*1},e_{*2},e_{*3})$ be a second basis as in (4.2.9). Then

$$XQ,\ \text{where } Q=\begin{bmatrix}3/4 & -1/4 & -1/4\\ -1/4 & 3/4 & -1/4\\ 1/4 & -1/4 & 3/4\end{bmatrix}=P^{-1}.$$

It follows from (4.2.11) that the representation of the inner product with respect to the new basis is $Q^*HQ=I$. Q.E.D.

(4.3) ORTHOGONAL BASES

(4.3.1) DEFINITION. Let V be a vector space with inner product (,).

(1) If $x,y \in V$, then x and y are <u>orthogonal</u> if and only if $(x,y)=0$.

(2) A family $(x^1,\ldots,x^r)$ will be called an <u>orthogonal family</u> if and only if $(x^i,x^j)=0$ whenever $i \neq j$.

(3) A family $(x^1,\ldots,x^r)$ will be called an <u>orthonormal family</u> if and only if

$$(x^i,x^j)=\begin{cases}0 & \text{if } i\neq j\\ 1 & \text{if } i=j\end{cases}.$$

Example. Let $V=C_{3,1}$ with the usual inner product, (x,y) x^*y. If

$$x=\begin{bmatrix} i \\ -i \\ 0 \end{bmatrix} \text{ and } y=\begin{bmatrix} -i \\ -i \\ 0 \end{bmatrix},$$

then $x^*y=(-i)(-i)+(i)(-i)+(0)(0)=-1+1=0$, so that x and y are orthogonal. If

$$z=\begin{bmatrix} 0 \\ 0 \\ 1+i \end{bmatrix},$$

then we can easily check that (x,y,z) is an orthogonal family. For we have $(x,z)=(-i)(0)+(i)(0)+0(1+i)=0$, and $(y,z)=$ $(i)(0)+(i)(0)+0(1+i)=0$. Finally if

$$x^1=\begin{bmatrix} i/\sqrt{2} \\ -i/\sqrt{2} \\ 0 \end{bmatrix} \qquad x^2=\begin{bmatrix} -i/\sqrt{2} \\ -i/\sqrt{2} \\ 0 \end{bmatrix} \quad x^3=\begin{bmatrix} 0 \\ 0 \\ (1+i)/\sqrt{2} \end{bmatrix},$$

it follows from Proposition (4.1.4) that (x^1,x^2,x^3) is an orthogonal family. Computing $x^{i*}x^i$, we have

$$x^{1*}x^1= \left(\frac{-i}{\sqrt{2}}\right)\left(\frac{i}{\sqrt{2}}\right)+\left(\frac{i}{\sqrt{2}}\right)\left(\frac{-i}{\sqrt{2}}\right)+0.0=\tfrac{1}{2}+\tfrac{1}{2}=1$$

$$x^{2*}x^2= \left(\frac{i}{\sqrt{2}}\right)\left(\frac{i}{\sqrt{2}}\right)+\left(\frac{i}{\sqrt{2}}\right)\left(\frac{-i}{\sqrt{2}}\right)+0\cdot 0=\tfrac{1}{2}+\tfrac{1}{2}=1$$

$$x^{3*}x^3= 0\cdot 0+0\cdot 0+\left(\frac{1-i}{\sqrt{2}}\right)\left(\frac{1+i}{\sqrt{2}}\right)=\frac{1+1}{2}\quad 1 \ .$$

Thus (x^1,x^2,x^3) is an orthonormal family.

We shall now show that given any family $(x^1,\ldots,x^r)$ of linearly independent vectors it is possible to orthonormalize them thus: y^1 will be obtained from x^1 by x^2, and so forth.

(4.3.2) THEOREM(GRAM-SCHMIDT ORTHONORMALIZATION PROCESS). Let $(x^1,\ldots,x^r)$ be a linearly independent family of vectors in a vector space V. Then there is an orthonormal family $(y^1,\dots,y^r)$ such that $y^i \in \langle x^1,\dots,x^i \rangle$, $i=1,\dots,r$.

Proof. Since (x^1,x^1) 0, we put $y^1=x^1/\sqrt{(x^1,x^1)}$. Then $(y^1,y^1) \simeq 1$. We next put $z^2=x^2-\alpha_1 y^1$, where α_1 is to be chosen presently. Observe that $(y^1,z^2)=(y^1,x^2)-\alpha_1$. Hence $(y^1,z^2)=0$ if $\alpha_1=(y^1,x^2)$. So put $z^2=x^2-(y^1,x^2)y^1$. Since $y^1 \in \langle x^1 \rangle$, it follows that $z^2 \in \langle x^1,x^2 \rangle$. Further, $x^2 \notin \langle y^1 \rangle = \langle x^1 \rangle$. Hence $z^2 \neq 0$. Thus $(z^2,z^2)>0$. We now put $y^2=z^2/\sqrt{(z^2,z^2)}$. Then $(y^1,y^2)=(y^1,z^2)/\sqrt{(z^2,z^2)}=0$ and $(y^2,y^2)=1$. Thus the family (y^1,y^2) is orthonormal.

Suppose now we already have an orthonormal family $(y^1,\dots,y^{i-1})$ where $y^j \in \langle x^1,\dots,x^j \rangle$, $j=1,\dots,i-1$. For reasons similar to those above put

$$z^i=x^i-(y^1,x^i)y^1-(y^2,x^i)y^2-\dots-(y^{i-1},x^i)y^{i-1}.$$

Then for $j=1,\dots,i-1$,

$$(y^j,z^i)=(y^j,x^i)-(y^j,x^i)(y^j,y^j)=0,$$

since

$$(y^j,y^k)=\begin{cases}1, & j=k\\ 0, & j\neq k\end{cases}.$$

Now note that $y^j \in \langle x^1,\dots,x^{i-1} \rangle$ for $j=1,\dots,i-1$, whence $z^i \in \langle x^1,\dots,x^i \rangle$. Hence we may put $y^i=z^i/\sqrt{(z^i,z^i)}$. Then clearly $(y^1,\dots,y^i)$ is orthonormal and $y^j \in \langle x^1,\dots,x^j \rangle$, $j=1,\dots,i$. Proceeding thus we obtain an orthonormal family satisfying the conditions of the theorem. Q.E.D.

Observe that our proof was constructive. We have not merely proved the existence of orthonormal families satisfying

(4.2.12), but we have given a method for their construction.

Let V be a vector space. If $\mathcal{Y}=(y^1,\ldots,y^n)$ is a basis for V and an orthonormal family, then $\mathcal{Y}$ will be called an <u>orthonormal basis</u>. We shall state without proof the following theorem.

(4.3.3) THEOREM. Let V be a finite-dimensional vector space with an inner product. Then there is an orthonormal basis for V.

CHAPTER 5

APPLICATIONS

(5.1) EQUIVALENCE RELATIONS

In this chapter we shall look at some applications of linear groups, especially in relation to inner products. We shall first consider the concepts of congruence and conjunctive equivalence and normal matrices.

(5.1.1) DEFINITION. If $K=P^*HP$, where P is nonsingular and H and K are complex nxn matrices, we shall say K is <u>conjunctive</u> to H, and we shall write $K \overset{C}{\sim} H$.

(5.1.2) DEFINITION. Let H and K be real nxn matrices. If $K=P^THP$, where P is real nonsingular, then K will be called <u>congruent</u> to H, and we shall write $K \overset{R}{\sim} H$.

Both conjunctive and congruence are equivalence relations, i.e. for appropriate matrices H,J and K we have

$H \overset{C}{\sim} H \qquad (H \overset{R}{\sim} H)$

$H \overset{C}{\sim} K \Rightarrow K \overset{C}{\sim} H \qquad (H \overset{R}{\sim} K) \Rightarrow (K \overset{R}{\sim} H)$

$H \overset{C}{\sim} K \text{ and } K \overset{C}{\sim} J \Rightarrow H \overset{C}{\sim} J \qquad (H \overset{R}{\sim} K \text{ and } K \overset{R}{\sim} J \Rightarrow H \overset{R}{\sim} J)$

To illustrate (5.1.2) let

$$K=\begin{bmatrix} -40 & -25 & -5 \\ -25 & -15 & 0 \\ -5 & 0 & -3 \end{bmatrix} \quad H=\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \quad P=\begin{bmatrix} 1 & 2 & 1 \\ 3 & 1 & 2 \\ 4 & 3 & 0 \end{bmatrix}$$

Then a straightforward calculation shows that $K=P^THP$, where $K \overset{R}{\sim} H$.

$$K=\begin{bmatrix}-40 & -25 & -5\\ -25 & -15 & 0\\ -5 & 0 & -3\end{bmatrix}=\begin{bmatrix}1 & 3 & 4\\ 2 & 1 & 3\\ 1 & 2 & 0\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & -2\end{bmatrix}\begin{bmatrix}1 & 2 & 1\\ 3 & 1 & 2\\ 4 & 3 & 0\end{bmatrix}$$

$$=\begin{bmatrix}1 & -3 & -8\\ 2 & -1 & -6\\ 1 & -2 & 0\end{bmatrix}\begin{bmatrix}1 & 2 & 1\\ 3 & 1 & 2\\ 4 & 3 & 0\end{bmatrix}$$

$$=\begin{bmatrix}-40 & -25 & -5\\ -25 & -15 & 0\\ -5 & 0 & -3\end{bmatrix}$$

For the sake of brevity and clarity, for the theorems in this chapter, we do not give complete proofs (with a few exceptions) instead we sketch an outline of the proof and then give an illustrative example, or we state the theorem without proof and reference where the reader might find a proof.

The following theorem will be given without proof.

(5.1.3) THEOREM. Let V be an n-dimensional vector space over the real numbers. Let $\mathcal{X}$ and $\mathcal{Y}=\mathcal{X}P$ be bases, where P is a real nonsingular matrix. Let (,) be an inner product on V. If $H=P_{\mathcal{X}}$ (,) and $K=P_{\mathcal{Y}}$ (,), then $K=P^THP$.

(5.1.4) DEFINITION. Let A and B be nxn matrices with complex elements. We shall say that A and B are unitarily equivalent if and only if $B=U^*AU$ for some unitary matrix U. We shall write $A \overset{U}{\sim} B$.

For example, let

$$A=\begin{bmatrix}i\sqrt{2} & 2 & i\\ -i\sqrt{2} & 2 & -i\\ 1 & \frac{1}{2} & 1\end{bmatrix} \quad B=\begin{bmatrix}2 & -2 & 0\\ i\sqrt{2} & i\sqrt{2} & i\sqrt{2}\\ 3\sqrt{2}/4 & \sqrt{2}/4 & 1\end{bmatrix}$$

Then with the unitary matrix U,

$$U=\begin{bmatrix}\sqrt{2}/2 & \sqrt{2}/2 & 0\\ \sqrt{2}/2 & -\sqrt{2}/2 & 0\\ 0 & 0 & 1\end{bmatrix}.$$

$A \overset{U}{\sim} B$, since

$$U^*AU=\begin{bmatrix}\sqrt{2}/2 & \sqrt{2}/2 & 0\\ \sqrt{2}/2 & -\sqrt{2}/2 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}i\sqrt{2} & 2 & i\\ -i\sqrt{2} & 2 & -i\\ 1 & \frac{1}{2} & 1\end{bmatrix}\begin{bmatrix}\sqrt{2}/2 & \sqrt{2}/2 & 0\\ \sqrt{2}/2 & \sqrt{2}/2 & 0\\ 0 & 0 & 1\end{bmatrix}$$

$$=\begin{bmatrix}2 & -2 & 0\\ i\sqrt{2} & i\sqrt{2} & i\\ 3\sqrt{2}/4 & \sqrt{2}/4 & 1\end{bmatrix}=B.$$

(5.2) CONGRUENCE AND CONJUNCTIVE EQUIVALENCE

In this section we shall find canonical forms for hermitian matrices under the relation $\overset{C}{\sim}$ defined in (5.1.1).

(5.2.1) DEFINITION. A nxn matrix of the form $I_s \oplus (-I_t) \oplus 0_{n-s-t}$ is said to be in <u>conjunctive</u> <u>canonical</u> form.

For example, if n=5, s=2, and t=1, we have

$$I_2 \oplus (I_1) \oplus 0_2 = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} + [-1] \oplus \begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix} = \begin{bmatrix}1 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & -1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}.$$

Next we state and prove an important theorem.

(5.2.2) SYLVESTER'S THEOREM. Let H be a hermitian matrix. Then there is precisely one matrix in conjunctive canonical form that is conjunctively equivalent to H.

Proof. Step 1.

(a) If H=0, then H is in conjunctive canonical form, and there is no more to be done. Suppose $H \neq 0$, then we are finished with step (a). If not, choose some $h_{ji} \neq 0$. Thus we let $h_{i*} \longrightarrow$

$h_{i*}-\lambda h_{j*}$, and we call the matrix thus obtained H'. We then perform $h'_{*i}\rightarrow h'_{*i}-\bar{\lambda}h'_{*j}$, and we shall call the new matrix H . We shall choose our λ presently. We thus have $h'_{ii}=h_{ii}-\lambda h_{ji}=-\lambda h_{ji}$, and $h'_{ij}=h_{ij}-\lambda h_{jj}=h_{ij}$. Hence $h''_{ii}=h'_{ii}-\bar{\lambda}h'_{ij}=-\lambda h_{ji}-\bar{\lambda}h_{ij}=-(\lambda h_{ji}+\bar{\lambda}\bar{h}_{ji})$ since $h_{ji}=\bar{h}_{ij}$. If we put $\lambda=\bar{h}_{ji}$, it follows from $h_{ji}\bar{h}_{ji}=|h_{ji}|^2$ that $h''_{ii}=2|h_{ji}|^2\neq 0$. We shall now write H for H , and we may assume that some $h_{ii}\neq 0$.

(b) If $h_{11}\neq 0$, then we are finished with step (b). If not, then we know some $h_{ii}\neq 0$ with $i\neq 1$. First interchange $h_{i*}\leftrightarrow h_{1*}$, then $h_{*i}\leftrightarrow h_{*1}$. This interchanges h_{ii} and h_{11}, and hence in the new matrix $h_{11}\neq 0$. Since H is hermitian, h_{11} is real, so either $h_{11}=-\kappa^2$. Multiply h_{1*} by κ^{-1}, then multiply h_{*1} by κ^{-1}. We obtain either $h_{11}=1$ or $h_{11}=-1$.

(c) For $j=2,\ldots,n$, perform $h_{j*}\rightarrow h_{j*}-(h_{11}h_{j1})h_{1*}$ and $h_{*j}\rightarrow h_{*j}-(h_{11}h_{j1})h_{*1}$. Note that $h_{11}=h_{11}^{-1}$. We obtain a hermitian matrix H in which $h_{21}=\ldots=h_{n1}=0$, and also $h_{12}=\ldots=h_{1n}=0$. Thus $H=h_{11}\oplus H_1$.

Step 2. We repeat the above argument on H_1, and we obtain that $H=h_{11}\oplus h_{22}\oplus H_2$, where $h_{11}=\pm 1$, $h_{22}=\pm 1$, and H_2 is hermitian.

This argument either terminates after r steps, and we obtain that $H=\mathrm{diag}(h_{11},\ldots,h_{rr})\oplus 0_{n-r}$, where each $h_{ii}=1$, or else after n steps we obtain $H=\mathrm{diag}(h_{11},\ldots,h_{nn})$, $h_{ii}=1$. Suppose s of the h_{ii} are 1 and t of the h_{ii} are -1, where $s+t=r$. Then we may permute rows and columns simultaneously so that $h_{11}=\ldots=h_{ss}=1$, $h_{s+1,s+1}=\ldots=h_{s+t,s+t}=-1$. Thus we have reduced H to $I_s\oplus(I_t)\oplus$

O_{n-s-t}, a matrix in conjunctive canonical form.

Notice that the last step of our algorithm can be effected by premultiplication by a permutation matrix Q and postmultiplication by $Q^T=Q^*$. Each of the other operations can be performed by premultiplication by an elementary matrix and postmultiplication by E^*. Hence we have replaced H by P*HP, where P is a product of elementary matrices and a permutation matrix and hence is nonsingular. We have shown that P*HP is in conjunctive canonical form such that $K \overset{c}{\sim} H$. If also $K' \overset{c}{\sim} H$, where K is in conjunctive canonical form, then $K' \overset{c}{\sim} K$, whence $K=K'$ and this proves uniqueness. Q.E.D.

(5.3) NORMAL MATRICES

In this section we have collected a number of results relating to normal matrices. However, instead of giving the proofs we have given hints for some of the more difficult theorems.

(5.3.1) DEFINITION. Let A be a complex nxn matrix. Then A is <u>normal</u> if and only if AA*=A*A.

The matrix

$$A=\begin{bmatrix} 5 & -i & 1+i \\ i & 2 & -i \\ 1-i & i & 1 \end{bmatrix}$$

is normal.

(5.3.2) THEOREM. A matrix A is normal if and only if it is unitarily equivalent to a diagonal matrix. Hence a hermitian matrix is normal.

HINTS. Suppose $A=U\Delta U^*$. Show that $A^*A=AA^*$. Now suppose that $AA^*=A^*A$. Thus $U^*AU=T$, an upper triangular matrix. Show that $T^*T=TT^*=B$. By computing b_{11} in two ways show that $t_{1j}=0$, $j=2,\ldots,n$. Proceed by induction.

(5.3.3) DEFINITION. Let A be an nxn matrix with complex elements. If H is positive semidefinite hermitian and W is unitary, then we call $A=HW$ a <u>polar decomposition</u> of A. Similarly, if K is positive semidefinite and X is unitary, then $A=XK$ is called a polar decomposition of A.

If a is a complex number, we can write $a=hw$, where $h=|a|$ and $|w|=1$. The polar decomposition for matrices generalizes the decomposition for complex numbers.

(5.3.4) PROPOSITION. For every positive definite matrix H, there is a unique positive definite matrix Q such that $Q^2=H$.

HINT. Let $H=U^*\Delta U$. Find D such that $D^2=\Delta$. For example, the matrix

$$H=\tfrac{1}{2}\begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix}$$

is hermitian(in fact, symmetric) positive definite. If

$$U=\left[\frac{\sqrt{2}}{2}\begin{matrix} 1 & -1 \\ 1 & 1 \end{matrix}\right],$$

then $U^THU=\mathrm{diag}(4,1)=\Delta$. If we take $D=\mathrm{diag}(2,1)$, then

$$D^2=\Delta \quad Q=UDU^T= 1/2\begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \quad Q^2=H.$$

Frequently the matrix Q is called the square root of H.

(5.3.5) THEOREM. Let A be an nxn matrix. Then A has polar decomposition $A=HW=XK$ as in (5.3.3). If A is nonsingular, then

H,K,W, and X are unique and W=X.

HINT. Write A=U*DV. Let H=U*DU and W=U*V. For uniqueness consider A*A and use the uniqueness of the square root.

(5.3.6) LEMMA. Let D be a diagonal matrix,

$$D=d_1I_{n1}\oplus d_2I_{n2}\oplus\cdots\oplus d_pI_{np},$$

where I_j is the jxj identity matrix and the d_j are distinct. If X commutes with D prove that

$$X=X_1\oplus X_2\oplus\cdots\oplus X_p,$$

where the X_i are distinct and of order n_i.

As an example of this lemma let us consider D=diag(2,2,1) and

$$X=\begin{bmatrix}2&1&0\\0&2&0\\0&0&3\end{bmatrix}=\begin{bmatrix}2&1\\0&2\end{bmatrix}\oplus[3]=X_1\oplus X_2.$$

Then $D=2I_2\oplus 1I_1$. Direct computation shows that

$$XD=\begin{bmatrix}2&1&0\\0&2&0\\0&0&3\end{bmatrix}\begin{bmatrix}2&0&0\\0&2&0\\0&0&1\end{bmatrix}=\begin{bmatrix}4&2&0\\0&4&0\\0&0&3\end{bmatrix}$$

$$DX=\begin{bmatrix}2&0&0\\0&2&0\\0&0&1\end{bmatrix}\begin{bmatrix}2&1&0\\0&2&0\\0&0&3\end{bmatrix}=\begin{bmatrix}4&2&0\\0&4&0\\0&0&3\end{bmatrix}.$$

Thus DX=XD.

BIBLIOGRAPHY

1. Chevalley, C. Theory of Lie Groups. Princeton, New Jersey: Princeton University Press, 1946.

2. Finbeiner, Daniel T. Elements of Linear Algebra. San Francisco: W.H. Freeman and Company, 1972.

3. Husain, Taqdir. Introduction to Topological Groups. Philadelphia: W.B. Saunders Company, 1966.

4. Pettofrezzo, Anthony J. Elements of Linear Algebra. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1970.

5. Schneider, Hans and George P. Barker. Matrices and Linear Algebra. New York: Holt, Rinehart and Winston, Inc., 1968.