GEOMETRIES OF MATRICES. I$_1$ ARITHMETICAL CONSTRUCTION

BY

LOO-KENG HUA

A discussion of the redundancy of the conditions involved in the generalizations of von Staudt's Theorem should be preceded by a study of the involutions. As an illustration and a supplement to part I, we give here a discussion of the geometry of 2-rowed symmetric matrices; we intend to treat the general case at a later occasion. More definitely, the condition concerning the harmonic separation is a consequence of the invariance of arithmetic distance. For the real case, even continuity (as well as the condition concerning the signature) is redundant (a proof for even order symmetric matrices has been obtained). As to the general discussion, some knowledge concerning involutions seems to be indispensable; the author will come back to it later.

Throughout the paper, the notations in I are taken over and we assume that $n=2$.

1. Normal subspaces.

THEOREM 1. *Given two matrices Z_1 and Z_2 with arithmetic distance $r(Z_1, Z_2)=1$, the points Z satisfying*

$$r(Z, Z_1) \leqq 1, \qquad r(Z, Z_2) \leqq 1$$

form a normal subspace.

Proof. Without loss of generality, we may take

$$Z_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad Z_2 = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Let

$$Z = \begin{pmatrix} x & y \\ y & z \end{pmatrix}.$$

Then we have

$$(x \pm 1)z - y^2 = 0,$$

that is, $z=y=0$. The theorem is now evident.

DEFINITION. The normal subspace obtained in Theorem 1 is said to be spanned by Z_1 and Z_2.

DEFINITION. Two normal subspaces are said to be *complementary* if there is one and only one pair of matrices, one from each subspace, with arithmetic distance less than 2.

Presented to the Society, April 28, 1945; received by the editors January 2, 1945.

THEOREM 2. *Two complementary subspaces may be carried simultaneously to*

$$\begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 \\ 0 & y \end{pmatrix}.$$

Proof. We may let

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$$

be the matrices to span the first subspace, and

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad X$$

be the matrices to span the second subspace. In fact, by a theorem of I, we may carry any three points A, B, C with

$$r(A, B) = 1, \qquad r(A, C) = r(B, C) = 2,$$

into

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad B_1 = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad C_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Since

$$r\left(\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, X\right) = 1,$$

we have

$$X = \begin{pmatrix} a^2 & ab \\ ab & b^2 + 1 \end{pmatrix}.$$

If $a \neq 0$, the matrix

$$\begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix}, \qquad x = a^2/(b^2 + 1),$$

of the first subspace is at distance 1 from X. This contradicts our hypothesis. Thus the second is spanned by

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 \\ 0 & b^2 + 1 \end{pmatrix}.$$

The theorem follows.

Consequently two complementary subspaces have a unique matrix in common.

THEOREM 3. *Let* $\mathfrak{S}$ *denote the set of matrices* S *such that*

$$r(S, P) = 2,$$

where P is the common matrix of two complementary normal subspaces. Then $\mathfrak{S}$ is transitive under the subgroup leaving the two subspaces invariant.

Proof. Let the two complementary subspaces be

$$\begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 \\ 0 & y \end{pmatrix}$$

and S be the points in $\mathfrak{S}$. By the hypothesis with $P=0$, we know that S is nonsingular. The transformation

$$Z_1 = Z(-S^{-1}Z + I)^{-1}$$

carries S into ∞, and it leaves the subspaces

$$\begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 \\ 0 & y \end{pmatrix}$$

invariant.

2. **Direct sum of complementary subspaces.**

DEFINITION. The subspace formed by points of the form

$$\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}$$

is called a (completely) *reducible subspace* which is the direct sum of the subspaces

$$\begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 \\ 0 & y \end{pmatrix}.$$

Now we shall give its arithmetic construction. We take, by Theorem 2, I as any point of the set $\mathfrak{S}$. Find P satisfying

$$r\left(P, \begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix}\right) = r\left(P, \begin{pmatrix} 0 & 0 \\ 0 & y \end{pmatrix}\right) = r(P, I) = 1.$$

Let

$$P = \begin{pmatrix} a^2+1 & ab \\ ab & b^2+1 \end{pmatrix};$$

we have

$$(a^2+1-x)(b^2+1) = a^2b^2, \qquad (a^2+1)(b^2+1-y) = a^2b^2.$$

Then, $a^2 = y(1-x)/(xy-x-y)$, $b^2 = x(1-y)/(xy-x-y)$.

Thus we have two solutions:

$$P_{\pm} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{xy-x-y}\begin{pmatrix} y(1-x) & \pm(xy(1-x)(1-y))^{1/2} \\ \pm(xy(1-x)(1-y))^{1/2} & x(1-y) \end{pmatrix}.$$

Finally, we find all K satisfying

$$r(K, P_+) = r(K, P_-) = 1.$$

Putting

$$K = I + \frac{1}{(xy - x - y)}\begin{pmatrix} k_1 & k_2 \\ k_2 & k_3 \end{pmatrix},$$

we have

$$(k_1 - y(1 - x))(k_3 - x(1 - y)) - (k_2 \pm (xy(1 - x)(1 - y))^{1/2})^2 = 0.$$

Then $k_2 = 0$, $k_1 = (1+\rho)y(1-x)$, $k_3 = (1+1/\rho)x(1-y)$.

The matrices of the form

$$K = \begin{pmatrix} 1 + (1 + \rho)y(1 - x) & 0 \\ 0 & 1 + (1 + 1/\rho)x(1 - y) \end{pmatrix}$$

run over all matrices of the reducible space.

Since we use the arithmetical notion only, we have the following general definition.

DEFINITION. Given two complementary subspaces $\mathfrak{X}$ and $\mathfrak{Y}$, let Q be a point of the set $\mathfrak{S}$. The reducible space (or direct sum of both subspaces) is defined by the aggregate of points K such that

$$r(P_+, K) = r(P_-, K) = 1,$$

where P_+ and P_- are both solutions of

$$r(Q, P) = r(P, X) = r(P, Y) = 1,$$

and X and Y belong to $\mathfrak{X}$ and $\mathfrak{Y}$ respectively.

As a consequence of Theorems 2 and 3 we have the following theorem.

THEOREM 4. *The aggregate of reducible spaces is transitive.*

3. **Involutions.** Given

$$Z = \begin{pmatrix} z_1 & z_2 \\ z_2 & z_3 \end{pmatrix},$$

we wish to find all matrices

$$\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}$$

of a reducible space such that

$$r\left(\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}, \begin{pmatrix} z_1 & z_2 \\ z_2 & z_3 \end{pmatrix}\right) = 1,$$

consequently

$$(x - z_1)(y - z_3) - z_2^2 = 0.$$

This set is denoted by Σ. To each matrix Z we have a set Σ. Conversely, to each Σ, we have two matrices

$$Z \quad \text{and} \quad \begin{pmatrix} z_1 & -z_2 \\ -z_2 & z_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} Z \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Thus we obtain a transformation

$$Z_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} Z \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{1}$$

which is called an *involution of the first kind.* Further, each point of the reducible space is a fixed point of (1); there are no other fixed points.

Since

$$\frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

we have an equivalent involution

$$Z_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} Z \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \tag{2}$$

and, since

$$\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}^{-1}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix},$$

we have another equivalent involution

$$Z_1 = -\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} Z \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{3}$$

It may be shown that the most general form of involutions of the first kind(1) is

$$Z_1 = (PZ - K_1)(K_2 Z + P')^{-1}, \tag{4}$$

where K_1 and K_2 are skew symmetric and

$$\mathfrak{F} = \begin{pmatrix} P & -K_1 \\ K_2 & P' \end{pmatrix}$$

(1) The general definition of an involution of the first or second kind is that $\mathfrak{F}^2 = \mathfrak{T}$ or $-\mathfrak{T}$. It may be verified easily that they are equivalent symplectically to (1) or (5) respectively. A detailed study of involutions forms the subject of II, which will appear soon. For this reason, the author omits some details of the discussion. Certainly, for $n = 2$, the properties used can all be verified directly and easily.

is symplectic. We use $\mathfrak{F}_1$, $\mathfrak{F}_2$, $\mathfrak{F}_3$ to denote the matrices of (1), (2) and (3) respectively. We may easily verify that

$$\mathfrak{F}_i\mathfrak{F}_j = -\mathfrak{F}_j\mathfrak{F}_i,$$

and

$$\mathfrak{F}_1\mathfrak{F}_2\mathfrak{F}_3 = \begin{pmatrix} iI & 0 \\ 0 & -iI \end{pmatrix},$$

that is,

$$Z_1 = -Z. \tag{5}$$

This is called an involution of the second kind.

4. **Commutative involutions.** Let $\mathfrak{S}$ and $\mathfrak{S}_0$ be two reducible subspaces associated with two involutions of the first kind, $\mathfrak{F}$ and $\mathfrak{F}_0$ respectively. (It may be shown easily that, if $\mathfrak{F}\mathfrak{F}_0 = \mathfrak{F}_0\mathfrak{F}$, then $\mathfrak{F}_0$ carries $\mathfrak{S}$ into itself pointwise.)

If $\mathfrak{F}$ carries $\mathfrak{S}_0$ into itself, but not pointwise, then

$$\mathfrak{F}_0\mathfrak{F} = -\mathfrak{F}\mathfrak{F}_0.$$

In fact, $\mathfrak{F}^{-1}\mathfrak{F}_0\mathfrak{F}$ leaves $\mathfrak{S}_0$ fixed pointwise. Thus

$$\mathfrak{F}^{-1}\mathfrak{F}_0\mathfrak{F}_1 = \pm\mathfrak{F}_0;$$

the upper sign is ruled out, since $\mathfrak{F}$ does not leave $\mathfrak{S}_0$ pointwise fixed.

THEOREM 5. *Any pair of commutative involutions of the first kind may be carried into* $\mathfrak{F}_1$ *and* $\mathfrak{F}_2$ *simultaneously.*

Proof. The first one may be assumed to be $\mathfrak{F}_1$. Let the second be given by (4). Then

$$\begin{pmatrix} P & -K_1 \\ K_2 & P' \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = - \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} P & -K_1 \\ K_2 & P' \end{pmatrix},$$

and we have

$$P = \begin{pmatrix} 0 & p_1 \\ p_2 & 0 \end{pmatrix}, \qquad K_1 = \begin{pmatrix} 0 & k_1 \\ -k_1 & 0 \end{pmatrix}, \qquad K_2 = \begin{pmatrix} 0 & k_2 \\ -k_2 & 0 \end{pmatrix},$$

where

$$p_1p_2 + k_1k_2 = 1.$$

If

$$\begin{pmatrix} p_1 - \lambda & -k_1 \\ k_2 & p_2 - \lambda \end{pmatrix}$$

has only simple elementary divisors, we have a, b, c, d such that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} p_1 & -k_1 \\ k_2 & p_2 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}, \qquad ad - bc = 1.$$

Then

$$\begin{pmatrix} 0 & a & 0 & b \\ a & 0 & b & 0 \\ 0 & c & 0 & d \\ c & 0 & d & 0 \end{pmatrix}\begin{pmatrix} P & -K_1 \\ K_2 & P' \end{pmatrix}\begin{pmatrix} 0 & d & 0 & -b \\ d & 0 & -b & 0 \\ 0 & -c & 0 & a \\ -c & 0 & a & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & \lambda & 0 & 0 \\ 1/\lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/\lambda \\ 0 & 0 & \lambda & 0 \end{pmatrix}.$$

(Notice that the transformation carries $\mathfrak{F}_1$ into $-\mathfrak{F}_1$, but they denote the same transformation.) Further

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/\lambda \end{pmatrix}$$

carries $\mathfrak{F}$ into $\mathfrak{F}_2$. If

$$\begin{pmatrix} p_1 - \lambda & -k_1 \\ k_2 & p_2 - \lambda \end{pmatrix}$$

has a double elementary divisor, we may take

$$\begin{pmatrix} p_1 & -k_1 \\ k_2 & p_2 \end{pmatrix} = \pm\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Now we have to consider the case

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 \end{pmatrix}.$$

The transformation

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

carries $\mathfrak{F}$ into $\mathfrak{F}_2$. Thus, we have the theorem.

THEOREM 6. *Any triple of commutative involutions of the first kind may be carried into* $\mathfrak{F}_1$, $\mathfrak{F}_2$ *and* $\mathfrak{F}_3$ *simultaneously.*

Proof. We may assume that the first two are $\mathfrak{F}_1$ and $\mathfrak{F}_2$. Let the third one be $\mathfrak{F}$. Since $\mathfrak{F}_1\mathfrak{F} = -\mathfrak{F}\mathfrak{F}_1$, we have

$$\mathfrak{F} = \begin{pmatrix} 0 & p_1 & 0 & -k_1 \\ p_2 & 0 & k_1 & 0 \\ 0 & k_2 & 0 & p_2 \\ -k_2 & 0 & p_1 & 0 \end{pmatrix}, \qquad p_1p_2 + k_1k_2 = 1.$$

Since $\mathfrak{F}_2\mathfrak{F} = -\mathfrak{F}\mathfrak{F}_2$, we have $p_1 = -p_2$. We may change it into $\mathfrak{F}_3$, since

$$\begin{vmatrix} p_1 - \lambda & -k_1 \\ k_2 & -p_1 - \lambda \end{vmatrix} = \lambda^2 + 1.$$

5. **Involution of the second kind.** An involution of the second kind has two isolated fixed points with arithmetic distance 2. Conversely, any two given points, with arithmetic distance 2, will serve as the isolated fixed points of an involution of the second kind, which is uniquely determined by them. In fact, let 0 and ∞ be fixed points, then the transformation takes the form

$$Z_1 = -AZA', \qquad A^2 = I.$$

We have P such that

$$PAP' = \begin{cases} \pm I, \\ \pm \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{cases}$$

The latter case cannot happen, since then 0 would not be an isolated fixed point. Thus we have

$$Z_1 = -Z.$$

Now we may define harmonic ranges. Four points Z_1, Z_2, Z_3, Z_4, no two of them with arithmetic distance less than 2, are said to form a harmonic range, if the involution determined by Z_1, Z_2 permutes Z_3 and Z_4.

Analytically, we let

$$(Z_1, Z_2, Z_3, Z_4) = ((Z_1 - Z_3)(Z_1 - Z_4)^{-1})((Z_2 - Z_3)(Z_2 - Z_4)^{-1})^{-1}.$$

The involution

$$(Z - Z_1)(Z - Z_2)^{-1} = -(Z^* - Z_1)(Z^* - Z_2)^{-1}$$

carries Z_3 into Z_4. Evidently

$$(Z_1, Z_2, Z_3, Z_4) = -I.$$

This condition is sufficient as well as necessary.

Thus the "invariance of arithmetic distances" implies the invariance of "harmonic range."

NATIONAL TSING HUA UNIVERSITY,
KUNMING, YUNNAN, CHINA.