PAIRS OF MATRICES WITH PROPERTY L. II[1]

BY

T. S. MOTZKIN AND OLGA TAUSSKY

This note is concerned, for matrices with elements in an algebraically closed field of arbitrary characteristic p, with pencils generated by pairs of matrices with property L. A pair of n by n matrices is said to have property L if for a special ordering of the characteristic roots α_i of A and β_i of B, the characteristic roots of $\lambda A+\mu B$ are $\lambda\alpha_i+\mu\beta_i$ for all values of λ and μ. (See [1–5].)

In §§1–5 another characterization of pairs of matrices with property L is given for a large class of such pairs. The method employed for this purpose is used in §6 for the study of pencils (not necessarily with property L) of diagonable matrices, i.e., matrices which are similar to a diagonal matrix. (These matrices are also called nondefective.) It is shown that for $p=0$, as well as for $n\leq p$, such pencils are *always generated by commutative matrices*. In §7 the significance of this result for general pencils of commutative matrices is investigated.

1. **The ν-discriminant.** The new characterization of pairs A, B of matrices with property L is obtained by considering those ratios λ/μ for which $\lambda A+\mu B$ has a multiple characteristic root. We see as follows that this is the case either for at most $n(n-1)$ ratios or for every λ/μ.

The characteristic roots of $\lambda A+\mu B$ are the solutions $\nu=\nu_1, \cdots, \nu_n$ of the determinantal equation

$$f(\lambda, \mu, \nu) \equiv \left| \nu I - \lambda A - \mu B \right| = 0$$

where I is the unit matrix. This equation has a multiple root if and only if the ν-discriminant Δ of $f(\lambda, \mu, \nu)$ vanishes. The ν-discriminant $\Delta(g)$ of a polynomial $g=\sum_{i=0}^{n} g_i\nu^i=g_n\prod_{i=1}^{n}(\nu-\nu_i)$ is defined as the Sylvester resultant

$$g_n^{2n-2}\prod_{i<k}(\nu_i-\nu_k)^2 = \left|\begin{matrix} g_0 & \cdots & \cdot & \cdot & g_{n-1} & g_n & 0 & \cdots & 0 \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & \cdots & 0 & g_0 & g_1 & \cdot & \cdot & \cdots & g_n \\ g_1 & \cdots & \cdot & (n-1)g_{n-1} & ng_n & 0 & \cdot & \cdots & 0 \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & \cdots & \cdot & 0 & g_1 & 2g_2 & \cdot & \cdots & ng_n \end{matrix}\right| \Big/ g_n$$

Presented to the Society, December 29, 1952; received by the editors July 14, 1954.

[1] This work was sponsored (in part) by the Office of Scientific Research, USAF.

We are indebted to H. Wielandt for discussion of our first draft which also led him to evolve, for $p=0$, alternative proofs and generalizations of our Theorems 3 and 4 [6].

See also T. S. Motzkin and Olga Taussky, *Pairs of matrices with property* L, Proc. Nat. Acad. Sci. U.S.A. vol. 39 (1953) pp. 961–963.

of g and $dg/d\nu$. Since $\Delta=\Delta(f)$ is a form (homogeneous polynomial) in λ and μ of degree $n(n-1)$ the assertion follows.

Call the ratios λ/μ for which $\Delta=0$ the *discriminant roots* of the pencil $\lambda A+\mu B$. These roots need not, of course, be simple. We have, e.g., the following fact.

THEOREM 1. *Let A, B be a pair of matrices with property* L. *Then all discriminant roots of the pencil $\lambda A+\mu B$ are of even order, and the number of different discriminant roots is therefore at most $n(n-1)/2$, unless every matrix $\lambda A+\mu B$ has a multiple characteristic root.*

Proof. Because of property L we have

$$\left| \nu I - \lambda A - \mu B \right| = \prod_{i=1}^{n} (\nu - \lambda\alpha_i - \mu\beta_i)$$

whence

$$\Delta = \prod_{i<k} [\lambda\alpha_i + \mu\beta_i - (\lambda\alpha_k + \mu\beta_k)]^2.$$

2. Property D.

DEFINITION. A pair of matrices A, B has property D if $\Delta\equiv 0$ or if all discriminant roots are of even order, i.e., Δ is the square of a form in λ and μ.

Note that for characteristic $p=2$ the discriminant Δ is always a square. To show this expand $f(\lambda, \mu, \nu)$ in powers of ν:

$$f = \sum_{i=0}^{n} f_i \nu^i$$

where f_i are forms in λ and μ. We then have for $p=2$:

$$\frac{\partial f}{\partial \nu} = \sum f_{2i+1}\nu^{2i}.$$

Hence Δ is of the form

$$\begin{vmatrix} f_0 & f_1 & \cdot & \cdot & \cdot & \cdot \\ 0 & f_0 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ f_1 & 0 & f_3 & 0 & \cdot & \cdot \\ 0 & f_1 & 0 & f_3 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix}$$

which after subtracting the $(n+1)$st row from the first, the $(n+2)$nd from the second, and so on, turns out equal to

$$\begin{vmatrix} f_0 & 0 & f_2 & 0 & \cdot & \cdot \\ 0 & f_0 & 0 & f_2 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ f_1 & 0 & f_3 & 0 & \cdot & \cdot \\ 0 & f_1 & 0 & f_3 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix}.$$

The last column contains only zeros apart from f_n in the $(n-1)$st or $(2n-1)$st row, for even or odd n respectively. Expanding with respect to the last column, we see that after a suitable permutation of rows and columns

$$\Delta = f_n \begin{vmatrix} \Delta_1 & 0 \\ 0 & \Delta_1 \end{vmatrix} = (f_n^{1/2} \Delta_1)^2$$

where

$$\Delta_1 = \begin{vmatrix} f_0 & f_2 & \cdot & \cdot & \cdot \\ 0 & f_0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ f_1 & f_3 & \cdot & \cdot & \cdot \\ 0 & f_1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix}.$$

This proves the assertion.

Though property D does not imply property L as we shall see later, a partial converse of the fact that property L implies property D is the following.

THEOREM 2. *Let A, B be a pair of matrices with property* D. *Assume that no matrix $\lambda A+\mu B$ (for λ, $\mu\neq 0$, 0) in the pencil has a characteristic root of multiplicity $\geqq 3$ nor two different double characteristic roots. Except if the characteristic of the field is* 2 *it then follows that A, B have property* L.

REMARK. Instead of property D it suffices to assume that there is no discriminant root of order 1.

3. **The characteristic curve.** In order to prove Theorem 2 a few concepts from the theory of algebraic curves will be used. Interpret the equation $f(\lambda, \mu, \nu)=0$ as the equation of an algebraic curve C of order n in the projective λ, μ, ν-plane. We call C the *characteristic curve* of the pencil.

To say that A, B have property L is obviously equivalent to saying that the characteristic curve C splits into straight lines.

For no A and B does C pass through the point $P=(0, 0, 1)$. A straight line $\mu_0\lambda-\lambda_0\mu=0$ through P meets C in n points $(\lambda_0, \mu_0, \nu_i)$ where ν_i are the characteristic roots of $\lambda_0 A+\mu_0 B$. Defining a tangent of C at a point T of C, called point of contact, as a straight line having at T an intersection multiplicity(2) $m>1$ we see that $\mu_0\lambda-\lambda_0\mu=0$ is a tangent if and only if $\lambda_0 A+\mu_0 B$ has a multiple characteristic root, and that to every multiple characteristic root there corresponds a point of contact such that the intersection multiplicity there equals the multiplicity of the root.

For a point T of C, the smallest intersection multiplicity at T of a straight line through T is called the multiplicity m of T. (For $T=(1, 0, 0)$, this is m if λ^{n-m} is the highest power of λ actually appearing in $f(\lambda, \mu, \nu)$). For $m=n$, C splits into straight lines through T.) A point T with $m>1$ is singular, and every straight line through T is a tangent at T in the sense defined. The intuitive notion of tangent is better represented by the concepts of 0-tangent and 1-tangent (see [7]). Let $T=(1, 0, 0)$, and let a branch B of C be given by power series

$$\lambda = 1, \qquad \mu = \sum_{k=0}^{\infty} \alpha_k t^k, \qquad \nu = \sum_{k=0}^{\infty} \beta_k t^k,$$

or equivalently by series in another parameter τ obtained from the above by substituting $t=\sum_{k=1}^{\infty} \gamma_k \tau^k$, $\gamma_1 \neq 0$ (substitutions with $\gamma_1=0$ give inadmissible representations of the branch); the branch B belongs to T if $\alpha_0=\beta_0=0$. If $k_0 \leqq m$ is the smallest k for which $(\alpha_{k_0}, \beta_{k_0}) \neq (0, 0)$, then $\beta_{k_0}\mu-\alpha_{k_0}\nu=0$ is the *0-tangent*, and is a straight line through T whose intersection multiplicity with B is $>k_0$; hence its intersection multiplicity with C at T (added up from the different branches at T) is $>m$. Denoting by $\mu'=\sum \alpha_k k t^{k-1}$ and $\nu'=\sum \beta_k k t^{k-1}$ the derivatives of μ and ν, the generic tangent at (λ, μ, ν) is, in line coordinates, $(\mu\nu'-\nu\mu', -\nu', \mu')$, and by specialization to $t=0$ we obtain the *1-tangent* $(0, -\nu', \mu')_{t=0}$, that is, $\beta_{k_1}\mu-\alpha_{k_1}\nu=0$ where $k_1 \geqq k_0$ is the smallest k for which $(\alpha_{k_1}k_1, \beta_{k_1}k_1) \neq (0, 0)$. *The 0-tangent and 1-tangent coincide certainly if k_0 is not divisible by the characteristic p of the field*, hence in particular if $p=0$ or $n \leqq p$ (note that for $m=n=p$, $k_0=1$); they will then be called *proper tangent*.

All 1-tangents of C define, in the so-called dual plane, the *dual* C^* of C. The dual C_1^* of an irreducible component C_1 of C is an irreducible variety and cannot be the whole plane; if its dimension is 1 then it has a point in common with every straight line, and thus *there is a 1-tangent to C through every point* of the original plane. If the dimension is 0 there is only one 1-tangent to C_1 and C_1 is a straight line(3).

Returning to a tangent $\mu_0\lambda-\lambda_0\mu=0$ at a point T of multiplicity m we note that there will be more than two coinciding ν_i at T, that is, $\lambda_0 A+\mu_0 B$ will

(2) The intersection multiplicity of a curve C and a straight line L at a point is defined as the multiplicity of the corresponding root of the resultant of their equations and as ∞ if L belongs to C.

(3) C^{**} need, for finite characteristic, not be C, even if C contains no straight line [cf. [8], where $p=2$ in the fifth line should be $p\neq 2$].

have a corresponding characteristic root of multiplicity three at least if and only if either $m \geqq 3$, or $m \geqq 2$ and the tangent is a 0-tangent, or $m \geqq 1$ and T is an inflexion point(4).

Since the tangents from P to C correspond to the matrices with a multiple characteristic root, they also correspond to the discriminant roots. A multiple discriminant root, however, does not imply a characteristic root of higher multiplicity than two. Necessary and sufficient conditions for the occurrence of a multiple discriminant root are given in the following lemma.

4. Double discriminant roots.

LEMMA. *Let $f(\lambda, \mu, \nu)=0$ be the equation of an algebraic curve C in the projective plane over an algebraically closed field F. Then the ratio λ_0/μ_0 is a multiple root of the ν-discriminant Δ of f if and only if the straight line $\mu_0\lambda-\lambda_0\mu=0$ either passes through a singular point of C or is the tangent at an inflexion point of C or is a tangent with at least two points of contact or is a tangent at $P=(0, 0, 1)$, or if the characteristic of F is 2.*

Proof. Since λ_0/μ_0 is a root of Δ, the line $\mu_0\lambda-\lambda_0\mu=0$ is a tangent of C. If a point of contact T is $\neq P$ we can assume, without loss of generality, $T=(1, 0, 0)$. We then want to see under what circumstances $\mu_0=0$ is a multiple root of Δ. Since $\mu=0$ has a point of contact at T, we have $f=\mu g+\nu^2 h$ where g is a form in λ, μ, ν and h is a form in λ, ν. Let

$$g = \sum g_i\nu^i, \qquad g_i = g_i(\lambda, \mu)$$

and

$$h = \sum h_i\nu^i, \qquad h_i = h_i(\lambda).$$

The ν-discriminant of f is then

$$\left|\begin{matrix} \mu g_0 & \mu g_1 & \mu g_2 + h_0 & \cdots & h_{n-2} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mu g_1 & 2(\mu g_2 + h_0) & 3(\mu g_3 + h_1) & \cdots & n h_{n-2} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix}\right| \Big/ h_{n-2}.$$

Dividing the first column by μ we may write $\Delta/\mu \equiv \Delta_1 \pmod{\mu}$ where

$$\Delta_1 = \left|\begin{matrix} g_0 & 0 & h_0 & h_1 & \cdot & \cdot \\ 0 & 0 & 0 & h_0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ g_1 & 2h_0 & 3h_1 & 4h_2 & \cdot & \cdot \\ 0 & 0 & 2h_0 & 3h_1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix}\right| \Big/ h_{n-2} = \pm 4g_0h_0^2 \left|\begin{matrix} h_0 & h_1 & \cdot & \cdot & \cdot \\ 0 & h_0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 2h_0 & 3h_1 & \cdot & \cdot & \cdot \\ 0 & 2h_0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix}\right| \Big/ h_{n-2}.$$

(4) A point T with $m=1$ is called inflexion point if the tangent at T has at T an intersection multiplicity $\geqq 3$.

Subtracting the double of every row starting with an h_0 from the row starting in the same column with $2h_0$ we obtain

$$\Delta_1 = \pm 4g_0h_0^3\Delta_2,\ \Delta_2 = \left|\begin{matrix} h_0 & h_1 & \cdots \\ 0 & h_0 & \cdots \\ \cdots & \cdots & \cdots \\ h_1 & 2h_2 & \cdots \\ 0 & h_1 & \cdots \\ \cdots & \cdots & \cdots \end{matrix}\right| \Big/ h_{n-2}.$$

We see that, for $T \neq P$, Δ is divisible by μ^2 if and only if: either $2=0$, or g_0 is divisible by μ (T is singular since $f(1, \mu, \nu)$ has no constant term and no linear terms), or g_0 is not divisible by μ and $h_0=0$ (T is an inflexion point), or $h_0 \neq 0$ and $\Delta_2=0$ (existence of another point of contact). For $T=P$ we can assume $\mu_0=0$ whence, in similar notation as before, $f=\mu g+\lambda^2 h$ and

$$\Delta = \begin{array}{r} \\ (n-1)\text{st row} \\ \\ (2n-1)\text{st row} \end{array}\left|\begin{matrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0\cdots 0 & \mu g_0+\lambda^2 h_0 & \cdots & \mu g_{n-2}+\lambda^2 h_{n-2} & \mu g_{n-1} & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0\cdots 0 & 0 & \mu g_1+\lambda^2 h_1 & \cdots & (n-1)\mu g_{n-1} & n \end{matrix}\right|$$

since the coefficient of ν^n is 0. Dividing the last column but one by μ we have $\Delta/\mu \equiv \Delta_1 \pmod{\mu}$ with

$$\Delta_1 = \begin{array}{r} \\ (n-1)\text{st row} \\ \\ (2n-1)\text{st row} \end{array}\left|\begin{matrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0\cdots 0 & \lambda^2 h_0 & \cdots & \lambda^2 h_{n-2} & g_{n-1} & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0\cdots 0 & 0 & \lambda^2 h_1 \cdots & (n-2)\lambda^2 h_{n-2} & (n-1)g_{n-1} & n \end{matrix}\right| = 0$$

since the last three columns have entries $\neq 0$ only in two rows.

If *every* matrix $\lambda_0 A+\mu_0 B$ has a multiple characteristic root it follows that C has a component which has to be counted double, or for characteristic $p \neq 0$, a component for which the exponents of ν with nonvanishing coefficients are all divisible by p. Indeed, from $\Delta \equiv 0$ it follows that every straight line through P is a tangent to C and therefore since P is not on C that C, if without multiple components, has an irreducible component whose generic tangent passes through P. We now show under what circumstances this is possible. If $g(\lambda, \mu, \nu)=0$ is the equation of the irreducible component and if all tangents pass through $P=(0, 0, 1)$ this implies $\partial g/\partial\nu=0$ along C. Since C is an irreducible curve this is not possible unless $\partial g/\partial \nu \equiv 0$. (This conclusion follows e.g. from the Hilbert Nullstellensatz.) The fact that $\partial g/\partial\nu \equiv 0$ implies the assertion.

5. Property D resumed.

Proof of Theorem 2. We want to prove that under the conditions assumed the curve C splits into n straight lines. Suppose C has an irreducible component not a straight line and consider a 1-tangent t from P to this component. As the matrix which corresponds to t has only one multiple characteristic root, t has only one point of contact with the whole curve C. Since there are no triple characteristic roots the point of contact is not an inflection point of C nor a singular point of multiplicity >2, and if it were a point of multiplicity 2 for $p \neq 2$ then t would be a 0-tangent and yield a characteristic root of multiplicity >2. By the lemma it follows for $\Delta \not\equiv 0$ that t corresponds to a discriminant root of multiplicity 1 the existence of which was excluded. For $\Delta \equiv 0$ and a double component we get a characteristic root of multiplicity $\geqq 4$. There remains the case $\Delta \equiv 0$, $p \neq 0$ with no double component. The multiplicity of the characteristic root is then at least p, whence $p = 2$.

The condition that no matrix in the pencil has a triple characteristic root is necessary. We can give two matrices A, B with property D, but without property L:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & 2 \\ -1 & -1 & -2 \end{pmatrix}.$$

The matrices A and B have the triple characteristic root 0. The matrix $A+B$, however, does not have the characteristic root 0, hence A, B do not have property L. The discriminant of the pencil is $-27\lambda^4\mu^2$.

Another example, valid for matrices with elements in a field with arbitrary characteristic $p \neq 2$, is

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

The matrices A, B, $A-B$ all have a triple characteristic root; A, B do not have property L, since A has all its characteristic roots zero and B is singular, while $A+B$ is not singular. Here all discriminant roots are exactly double roots, $\Delta = -27\lambda^2\mu^2(\lambda+\mu)^2$.

An example without triple characteristic roots, where some (six) matrices have two double characteristic roots, valid for $p \neq 2, 3, 5$, is

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & -2 & 0 \end{pmatrix}.$$

Here $|\nu I-\lambda A-\mu B| = (\nu^2+\mu^2-\lambda^2)(\nu^2+4\mu^2-4\nu\lambda)$, $\Delta = (8(\mu^2-\lambda^2)(\mu^2+3\lambda^2) \times(9\mu^2-5\lambda^2))^2$, whence D holds but not L.

For characteristic $p=2$ property D does not imply property L; as noted all pairs have property D, but already for $n=2$, e.g.

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

do not have property L.

According to Theorem 2, for $p\neq 2$ and $n=2$ property D alone with $\Delta\neq 0$ implies property L. Hence we have *for $n=2$: a pencil with, but for scalar multiples, only one matrix with a double characteristic root is generated by a pair of matrices with property* L.

A direct proof of this fact can easily be obtained: Assume A in triangular form and that it is the only matrix with a double root, which is further assumed to be zero. These assumptions do not constitute a restriction. Assume then

$$A = \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}.$$

If the matrix

$$\lambda A + B = \begin{pmatrix} \alpha & \beta + \lambda b \\ \gamma & \delta \end{pmatrix}$$

has no double roots for all finite values of λ then

$$(\alpha + \delta)^2 - 4(\alpha\delta - \gamma(\beta + \lambda b)) \neq 0$$

for all λ. This implies $\gamma b=0$. If $b=0$ the pencil contains a scalar matrix; in this case it is generated by a pair of commutative matrices. If $\gamma=0$ then A, B are both triangular matrices, and hence have property L.

6. Pencils of diagonable matrices.

THEOREM 3. *Let $\lambda A+\mu B$ be a pencil in which all matrices are diagonable and, for $p\neq 0$, let $n\leqq p$. Then A, B have property* L.

Proof. Consider again the characteristic curve C of the pencil. Suppose C had an irreducible component C' which is not a straight line. Let t be a proper tangent from $P=(0, 0, 1)$ to C. Without loss of generality we may assume the point of contact $T=(1, 0, 0)$. This means that A has a multiple characteristic root 0, say of multiplicity m. Hence we may suppose, by hypothesis, that the first m rows and columns of A vanish. Expanding $|\nu I-\lambda A-\mu B|$ into powers of λ we see that no higher power than λ^{n-m} occurs; T is therefore of multiplicity m, so that t is not a proper tangent at T, and this contradiction proves the theorem.

For characteristic 0, the conclusion of Theorem 3 still holds if (but for scalar multiples) only a single matrix in the pencil is not diagonable. Indeed, the line (0, 0, 1) is, for $p=0$, not a tangent to the dual C_1^* of any irreducible component C_1 of C, and has therefore, if C_1 is not a straight line, at least two points in common with C_1^*. To these there correspond two proper tangents from the point (0, 0, 1) to C_1, and hence two nonproportional nondiagonable matrices.

With C_1 of order 2, pencils without property L are easily constructed which have only two nonproportional nondiagonable matrices.

For $n=kp+1$, $k\geqq 1$, an example(5) of a pencil without property L formed entirely by diagonable matrices is (shown for $n=4$):

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

with $f=|\nu I-\lambda A-\mu B|=\nu^{n-1}(\nu+\lambda-\mu)-(\lambda+\nu)^{n-1}\lambda=\nu^n-\nu^{n-1}\mu-\lambda^n$, $\partial f/\partial\nu=\nu^{n-1}$. Here f is irreducible and $\lambda A+\mu B$ has no multiple characteristic root except for $\lambda=0$.

The proof of Theorem 3 shows that every 0-tangent from (0, 0, 1) belongs to a nondiagonable matrix. The converse is not true: for

$$A=\begin{pmatrix}0&1\\0&0\end{pmatrix},\qquad B=\begin{pmatrix}1&1\\0&0\end{pmatrix},$$

we have $f=(\nu-\mu)\nu$, and the nondiagonable matrix A belongs to the tangent $\mu=0$ which is not a 0-tangent.

A stronger result than Theorem 3 will now be established.

THEOREM 4. *Let $\lambda A+\mu B$ be a pencil in which all matrices are diagonable and, for $p\neq 0$, assume $n\leqq p$ or that A and B have property* L. *Then A, B can be diagonalized by the same similarity and therefore they commute.*

As in Theorem 3 it is essential that the field of elements be algebraically closed.

$$A=\begin{pmatrix}1&0\\0&0\end{pmatrix}\quad\text{and}\quad B=\begin{pmatrix}0&1\\1&0\end{pmatrix}$$

do not have property L, but $\lambda A+\mu B$ is diagonable and has different characteristic roots for all real λ and $\mu(\neq(0,0))$.

Theorem 4 fails to hold if a single matrix in the pencil is not diagonable. As an example take

(5) Developed in a discussion with I. Kaplansky in 1955, to whom we are also grateful for an earlier remark that led us to Theorems 3 and 4.

$$A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

We shall use the following

LEMMA. *Consider the determinant*

$$\Delta = \begin{vmatrix} x_{11} & x_{12} & \cdots & x_{1r} \\ x_{21} & x_{22} - x_2\lambda & \cdots & x_{2r} \\ \vdots & \vdots & & \vdots \\ x_{r1} & x_{r2} & \cdots & x_{rr} - x_r\lambda \end{vmatrix}$$

and assume that Δ and all its principal minors of order $\geqq r-s$ containing x_{11} vanish for $\lambda=\lambda_0$. Then Δ is a polynomial in λ which has λ_0 as a root of multiplicity $s+1$.

Proof. We may assume $\lambda_0=0$, for otherwise we put $\lambda=\lambda'+\lambda_0$ and consider Δ as a polynomial in λ'. Express now Δ as a polynomial in λ. It is easily seen that the constant term as well as the coefficients of all powers of λ up to λ^s vanish. This establishes the lemma.

Return now to the *proof of Theorem* 4. By Theorem 3 the characteristic roots of $\lambda A+B$ are $\lambda\alpha_i+\beta_i$. If every matrix in the pencil $\lambda A+B$ has a multiple characteristic root, then there must be some value of i, k $(i\neq k)$ for which the equation $\lambda\alpha_i+\beta_i=\lambda\alpha_k+\beta_k$ is satisfied for all values of λ. This implies $\alpha_i=\alpha_k$, $\beta_i=\beta_k$. Assume the characteristic roots so numbered that $\alpha_1=\alpha_2=\cdots=\alpha_t$ ($t=1$ in case not every matrix $\lambda A+B$ has a multiple characteristic root), $\alpha_{t+i}\neq\alpha_1$, $i>0$, $\beta_1=\beta_2=\cdots=\beta_s$ $(s\geqq t)$. For $t>s$ interchange A and B. It follows that $\lambda\alpha_1+\beta_1$ is a characteristic root of multiplicity t of $\lambda A+B$ for all values of λ, but of multiplicity $>t$ only for special values of λ. We assume A in diagonal form. The matrix $\lambda A+B$ is then of the form

$$\begin{pmatrix} \lambda\alpha_1+b_{11} & b_{12} & \cdot & b_{1t} & \cdot & \cdot & b_{1n} \\ b_{21} & \lambda\alpha_1+b_{22} & \cdot & b_{2t} & \cdot & \cdot & b_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{t1} & b_{t2} & \cdot & \lambda\alpha_1+b_{tt} & \cdot & \cdot & b_{tn} \\ b_{t+1,1} & \cdot & \cdot & \cdot & \lambda\alpha_{t+1}+b_{t+1,t+1} & \cdot & \cdot \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{n1} & \cdot & \cdot & \cdot & \cdot & \cdot & \lambda\alpha_n+b_{nn} \end{pmatrix}.$$

Consider the matrix $C_\lambda=(\lambda\alpha_1+\beta_1)I-\lambda A-B$. Since $\lambda A+B$ has the characteristic root $\lambda\alpha_1+\beta_1$ with multiplicity t for all values of λ the $(n-t+1)$-dimensional minors of C_λ must vanish for all values of λ, because the same is true for the similar diagonal matrix. Now consider an $(n-t+1)$-dimensional minor formed from the last $n-t$ rows and columns and the ith row and kth column ($i, k=1, 2, \cdots, t$). The coefficient of λ^{n-t} is $\pm b_{ik}\cdot\prod_{j>t}(\alpha_j-\alpha_1)$ for $i\neq k$ and $\pm(b_{ii}-\beta_1)\cdot\prod_{j>t}(\alpha_j-\alpha_1)$ for $i=k$. Since this coefficient vanishes, but $\alpha_j-\alpha_1\neq 0$ it follows that $b_{ik}=0$ for $i\neq k$, i, $k=1, \cdots, t$, and $b_{ii}=\beta_1$, $i=1, \cdots, t$.

In order to prove that all other $b_{ik}=0$, $i\neq k$, consider values of λ for which $\lambda\alpha_1+\beta_1$ is a characteristic root of higher multiplicity than t. This will occur when for some $i>t$ we have

$$\lambda_{1i}\alpha_1+\beta_1=\lambda_{1i}\alpha_i+\beta_i.$$

If such a characteristic root is of multiplicity $t+s$ then the $(n-t-s+1)$-dimensional minors of C_λ must all vanish for $\lambda=\lambda_{1i}$. Consider the minor Δ formed by the first column and the last $n-t-1$ columns and by the last $n-t$ rows. The determinant Δ is of the form mentioned in the lemma with $r=n-t$ and s replaced by $s-1$. Hence λ_{1i} is a zero of Δ of multiplicity s. However, Δ is a polynomial in λ of degree $n-t-1$ which vanishes for all values of λ_{1i}, hence for $n-t$ values. Although these values need not be different the lemma shows that they have to be counted with their full multiplicities. This is only possible if the polynomial vanishes identically. The coefficient of λ^{n-t-1} is $\pm b_{t+1,1}\prod_{i>1}(\alpha_{t+i}-\alpha_1)$, hence $b_{t+1,1}=0$. Similarly we can show that $b_{t+i,k}=0$, $i=1, \cdots, n-t$, $k=1, \cdots, t$. The same argument further applies to the columns, and we obtain $b_{k,t+i}=0$, $i=1, \cdots, n-t$, $k=1, \cdots, t$. The matrix $\lambda A+B$ is therefore of the form

$$\begin{pmatrix} \lambda\alpha_1+\beta_1 & & & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots & & \vdots \\ 0 & & \lambda\alpha_1+\beta_1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \lambda\alpha_{t+1}+b_{t+1,t+1} & \cdots & b_{t+1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & b_{n,t+1} & \cdots & \lambda\alpha_n+b_{nn} \end{pmatrix}.$$

The pencil in the lower right corner consists again of diagonable matrices with property L. By induction the theorem is proved.

7. **Pencils of commutative matrices.** It is well known that every matrix with complex numbers as elements can be regarded as the limit of a sequence of diagonable matrices, e.g., as the limit of a sequence of matrices with only simple characteristic roots. This idea can be extended to pairs of commutative matrices. We have

THEOREM 5. *Every pair of commutative n by n matrices A, B with complex numbers as elements is the limit of a sequence of pairs of (eo ipso, simultaneously) diagonable commuting matrices.*

This theorem can be interpreted as a converse of Theorem 4.

Proof. Assume that the theorem holds for smaller n. We consider several cases: 1. The case where A has at least two different characteristic roots; 2. The case when A has only one characteristic root, but more than one characteristic vector corresponding to it; 3. The case when A has only one characteristic root with only one characteristic vector corresponding to it.

In all three cases we assume A in its Jordan normal form. This is no restriction, for apply the same similarity transformation which transforms A to Jordan normal form to both A and B. We obtain again a pair of commutative matrices. If we show that this pair is the limit of a sequence of pairs of diagonable commuting matrices then the original pair is the limit of the sequence of pairs obtained by applying the inverse similarity.

CASE 1. In this case A can be assumed in the form

$$\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$$

where A_1, A_2 are square matrices with no characteristic roots in common. Since B commutes with A, it must be of the form

$$\begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}$$

where B_1, B_2 are square matrices of the same dimensions as A_1, A_2. This can easily be ascertained for A in normal form (see [9, p. 148]). The result then holds by induction hypothesis.

CASE 2. Since A splits up into several blocks, let $m(<n)$ be the dimension of the first one. Denote by C an auxiliary diagonal matrix (c_{ii}) with $c_{11} = c_{22} = \cdots = c_{mm} \neq c_{m+1,m+1} = c_{m+2,m+2} = \cdots = c_{nn}$. The matrices A and C commute and hence A and any matrix $\lambda B + \mu C$ commute. Since C does not have all its characteristic roots equal there must (by continuity) be some matrix $\lambda B + \mu C$, $\lambda \neq 0$ which does not have all its characteristic roots equal. Hence the pair of matrices A, $B + (\mu/\lambda)C$ satisfies the conditions of Case 1 and is therefore the limit of a sequence of pairs of diagonable matrices. The same is therefore true for the pair A, B.

CASE 3. A is of the form

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 \end{pmatrix} + \alpha I,$$

hence B is of the form

$$B = \begin{pmatrix} 0 & \beta_1 & \beta_2 & \cdots & \beta_{n-1} \\ 0 & 0 & \beta_1 & \cdots & \beta_{n-2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots 0 & \beta_1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} + \beta I.$$

The matrix $B-\beta_1 A$ has more than one characteristic vector which reduces Case 3 to Case 2.

This completes the proof of Theorem 5.

For matrices with elements in an arbitrary algebraically closed field an analogous proof shows:

THEOREM 6. *A generic pair of commutative n by n matrices A, B is diagonable.*

This means that the commutative pairs form an irreducible variety V (in $2n^2$-dimensional affine space) on which almost every point corresponds to a diagonable pair. Here "almost every" is the algebrico-geometrical "almost all": all but a proper, and thus lower-dimensional subvariety of V. For matrices with complex elements this implies that every commutative pair A, B is the limit of a sequence of pairs of diagonable matrices. If we already knew that V is irreducible the theorem would follow easily by remarking that a general A is not only diagonable, but has only simple characteristic roots, and that when this A is in diagonal form, B is also diagonal. As it is, the irreducibility of V is a by-product of the *proof*, in which we assume that the theorem is true for every smaller degree (for $n=1$ it holds trivially). We want mainly to show that V is generated by, that is, is the smallest variety containing, all diagonable pairs. This implies the irreducibility of V; for let W be the ostensibly irreducible variety of all pairs of diagonal matrices and T any nonsingular matrix: then if f and g are polynomials in the $2n^2$ elements of a pair of n by n matrices such that $fg=0$ on V then, for every T, either $f=0$ or $g=0$ on TWT^{-1}, and since the T's generate the irreducible variety of all n by n matrices, it follows that either $f=0$ on TWT^{-1}, for every T, and thus $f=0$ on V, or else $g=0$ on V, which means that V is irreducible.

We now proceed as in the proof of Theorem 5. In the first case where A has at least two different characteristic roots let

$$A = T\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} T^{-1}$$

where A_1, A_2 are square matrices with no characteristic roots in common. Since B commutes with A it must be of the form

$$T\begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix} T^{-1}$$

where B_1, B_2 are square matrices of the same dimensions as A_1, A_2. Since A_i and B_i $(i=1, 2)$ commute the pair A_i, B_i is by induction hypothesis in the variety generated by diagonable pairs of the corresponding dimension, and thus the same is true for A, B.

Secondly, if A has only the characteristic root α but more than one characteristic vector define m and C as in Case 2 of the preceding proof. Again A and every $\lambda B+\mu C$ commute. Now $\lambda B+\mu C$ has, for general λ/μ, different characteristic roots, whence by Case 1 the pair $A, \lambda B+\mu C$ is in the variety generated by diagonable pairs, and the same is therefore true for the specialization A, B.

Finally if A has only one characteristic vector then

$$A = T\begin{pmatrix} 0 & 1 & 0 & \cdot & 0 \\ 0 & 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & 1 \\ 0 & 0 & 0 & \cdot & 0 \end{pmatrix} T^{-1} + \alpha I$$

and B must be of the form

$$B = T\begin{pmatrix} 0 & b_1 & b_2 \cdots b_{n-1} \\ 0 & 0 & b_1 \cdots b_{n-2} \\ \cdot & \cdot & \cdot \cdots \cdot \\ 0 & 0 & 0 \cdot\cdot 0 \quad b_1 \\ 0 & 0 & 0 \cdots 0 \end{pmatrix} T^{-1} + \beta I.$$

The matrix $B-b_1A$ has more than one characteristic vector, so that by Case 2 the pair A, $B-b_1A$ belongs to the variety generated by diagonable pairs, and since the set of diagonable pairs, with C, D, also contains C, $D+b_1C$, the same holds for the variety generated by it.

THEOREM 7. *If A, B commute and $n>2$ then the pencil $\lambda A+\mu B$ $(\lambda, \mu \neq 0, 0)$ contains either only diagonable matrices or none or exactly one (but for scalar multiples). If $n=2$ there is no commutative pencil which contains only nondiagonable matrices.*

Proof. If the pencil contains two nonproportional diagonable matrices A, B then all are diagonable. For let A be in diagonal form with equal characteristic roots arranged adjoining each other so that

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & A_r \end{pmatrix}.$$

Here the A_i are scalar matrices such that A_i and A_k have different diagonal elements when $i \neq k$. It then follows that

$$B = \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & B_r \end{pmatrix}$$

where B_i is a square matrix of the same order as A_i. Since B is diagonable each B_i is. The similarity which diagonalizes B_i leaves A_i invariant, hence A, B are simultaneously diagonable and every matrix in the pencil is diagonable.

That all three cases can occur for $n > 2$ is evident from the case $n = 3$. The pencil generated by

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

does not contain any diagonable matrix apart from the zero matrix.

For $n = 2$ the pencil generated by

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} \beta & \beta_1 \\ 0 & \beta \end{pmatrix}, \qquad \beta_1 \neq 0$$

contains a diagonable matrix. It is easy to see that any two commutative 2 by 2 nondiagonable matrices can be reduced to this form by similarities and by adding scalars.

References

1. T. S. Motzkin and Olga Taussky, *Pairs of matrices with property* L, Trans. Amer. Math. Soc. vol. 73 (1952) pp. 108–114. (On p. 111, l. 1, delete the reference to k-commutativity, which does not even imply reducibility of the characteristic curve.)

2. ———, *On representations of finite groups*, Neder. Akad. Wetensch. Ser. A vol. 55 (1952) pp. 511–512.

3. N. A. Wiegmann, *A note on pairs of normal matrices with property* L, Proc. Amer. Math. Soc. vol. 4 (1953) pp. 35–36.

4. H. Wielandt, *Pairs of normal matrices with property* L, Journal of Research of the National Bureau of Standards vol. 51 (1953) pp. 89–90.

5. I. Kaplansky, *Completely continuous normal operators with property* L, Pacific Journal of Mathematics vol. 3 (1953) pp. 721–724.

6. H. Wielandt, *Remarks on diagonable matrices*, manuscript.

7. T. S. Motzkin, *L'Hospital's rule for nonzero characteristic*, to appear.

8. T. S. Motzkin, *The dual curve for $p \neq 0$*, Bull. Amer. Math. Soc. Abstract 56-1-25.

9. H. W. Turnbull and A. C. Aitken, *Canonical matrices*, London and Glasgow, 1932.

University of California,
Los Angeles, Calif.
National Bureau of Standards,
Washington, D. C.