

ON THE NUMERICAL RANGE OF A MATRIX

TRANSLATED FROM THE GERMAN BY PAUL F. ZACHLIN[†]
AND MICHIEL E. HOCHSTENBACH[‡]

ABSTRACT. This is an English translation of the paper “ÜBER DEN WERTEVORRAT EINER MATRIX” BY RUDOLF KIPPENHAHN, MATHEMATISCHE NACHRICHTEN 6 (1951), 193–228. This paper is often cited by mathematicians who work in the area of numerical ranges, thus it is hoped that this translation may be useful. Some notation and wording has been changed to make the paper more in line with present papers on the subject written in English.

In Part 1 of this paper Kippenhahn characterized the numerical range of a matrix as being the convex hull of a certain algebraic curve that is associated to the matrix. More than 55 years later this “boundary generating curve” is still a topic of current research, and “ÜBER DEN WERTEVORRAT EINER MATRIX” is almost always present in the bibliographies of papers on this topic.

In Part 2, the author initiated the study of a generalization of the numerical range to matrices with quaternion entries. The translators note that in Theorem 36, it is stated incorrectly that this set of points in 4-dimensional space is convex. A counterexample to this statement was given in 1984.^[1] In the notes at the end of this paper the translators pinpoint the flaw in the argument. In the opinion of the translators, this error does not significantly detract from the overall value and significance of this paper.

In the translation, footnotes in the original version are indicated by superscript Arabic numerals, while superscript Roman numerals in brackets are used to indicate that the translators have a comment about the original paper. All of these comments appear at the end of this paper, and the translators also have corrected some minor misprints in the original without comment.

[†] Department of Mathematics, Case Western Reserve University, Cleveland, OH 44106-7058, USA paul.zachlin@case.edu.

[‡] Department of Mathematics and Computing Science, Eindhoven University of Technology, PO Box 513, 5600 MB, The Netherlands. m.e.hochstenbach@tue.nl.

On the Numerical Range of a Matrix.¹

by RUDOLF KIPPENHAHN in Bamberg.

(Published September 13, 1951.)

Introduction

Let $A = (a_{\mu\nu})(\mu, \nu = 1, \dots, n)$ be a square matrix with complex number entries. The *numerical range* $W(A)$ of the matrix A is defined as the set of all complex numbers which can be assumed by the form²

$$(1) \quad \Phi(A, x) = x^*Ax$$

when the vector x with complex components x_1, \dots, x_n varies only over all vectors with two-norm 1, so we must add to (1) the side condition that $x^*x = 1$.

The numerical range of a complex matrix is a subset of the Gaussian plane. Since the region from which x is taken is closed, and since $\Phi(A, x)$ is a continuous function of x , it follows that the set of points $W(A)$ is also closed. TOEPLITZ [9]³ and HAUSDORFF [4] have proven that the region $W(A)$ is convex.

The goal of this work is to investigate the geometric properties of the numerical range of a matrix. Geometric and analytic methods can be applied to numerical ranges of matrices, since for each matrix of dimension n a curve of class n can be found explicitly, the *boundary generating* curve, and its convex hull coincides with the numerical range of the matrix (§ 3). The characteristic curve is a curve without points of inflection (§ 4). For the cases $n = 2$ and $n = 3$ each possible type of curve can be completely described (§ 7). A general examination of curve types should be based on a classification of the curves of class n . However, at the present day this has not yet been completed. From the representation of the boundary generating curve in the form of an equation one may estimate the width, diameter, and area of the numerical range, as well as deduce the length of the boundary of a numerical range (§ 9).

In the second part of this work the numerical ranges of matrices of dimension n whose elements are quaternions are investigated. These numerical ranges can be described as convex sets in a four-dimensional vector space. They always lie rotationally symmetric with respect to the "central axis". Their theory can be reduced to the theory of numerical ranges of complex matrices of dimension $2n$.

Part 1. Complex Matrices

1. SIMPLEST PROPERTIES OF THE NUMERICAL RANGE

Theorem 1. *The numerical range of a matrix A is invariant under unitary transformations.*

¹Dissertation Erlangen 1951; Referees: PROF. DR. WILHELM SPECHT, PROF. DR. GEORG NÖBELING. – I thank MR. PROF. SPECHT for the idea for this work as well as for much essential advice.

²To form the matrix M^* from M , we replace each element of the transpose matrix M^T by its complex conjugate. A vector x should be considered as a matrix with one column; x^* is therefore a row vector.

³Numbers in square brackets refer to the bibliography.

Proof. If U is a unitary matrix⁴, then

$$W(A) = W(U^*AU),$$

since if x runs over the set of all normalized n -dimensional complex vectors then so does $y = Ux$. \square

Theorem 2. *The numerical range of a Hermitian matrix H is a closed interval on the real axis, whose endpoints are formed by the extreme eigenvalues of H .*

Proof. Since the numerical range is a unitary invariant, we may assume that the Hermitian matrix H is in (real) diagonal form⁵:

$$H = \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \alpha_n \end{pmatrix} \text{ with } \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n.$$

If we form $\Phi(H, x)$, then we may also assume that each component α_ν of x is real, since each value which $\Phi(H, x)$ assumes for a complex x , $\Phi(H, x)$ also assumes when the component x_ν in x is replaced by its absolute value $\xi_\nu = |x_\nu|$, because $\Phi(H, x) \equiv x^*Hx \equiv \sum_{\nu=1}^n \alpha_\nu x_\nu \bar{x}_\nu$. Therefore let x be the vector with real components ξ_1, \dots, ξ_n with $\sum_{\nu=1}^n \xi_\nu^2 = 1$. This implies

$$\Phi(H, x) = \sum_{\nu=1}^n \alpha_\nu \xi_\nu^2.$$

This expression, however, can yield all values on the closed interval on the real axis bounded by the extreme eigenvalues α_1, α_n of H . \square

$\Phi(H, x)$ assumes the extreme eigenvalues α_1, α_n of the diagonal matrix H only for the eigenvectors

$$x_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad x_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

For arbitrary unitary U from the equivalence of the equations

$$Hx = \alpha x, \quad (U^*HU)U^*x = \alpha U^*x$$

it follows now that the extreme eigenvalues of U^*HU are assumed only for these eigenvectors in the function $\Phi(U^*HU, x)$. By this we have proved that the function $\Phi(K, x)$ of an arbitrary Hermitian matrix K assumes the extreme eigenvalues only for eigenvectors.

⁴A matrix U is called *unitary* if the equation $UU^* = U^*U = I_n$ holds with I_n the identity matrix of dimension n .

⁵To the extent that it is possible, Latin letters (except indices) are characterized by complex entries, Greek letters are characterized by real number entries.

It is well-known that every complex matrix A can be uniquely split into two components so that

$$A = H_1 + iH_2,$$

where H_1 and H_2 are Hermitian matrices which hold the following relation with A :

$$H_1 = \frac{A + A^*}{2}, \quad H_2 = \frac{A - A^*}{2i}.$$

From this we have

$$\Phi(A, x) = \Phi(H_1, x) + i\Phi(H_2, x),$$

where $\Phi(H_1, x)$ and $\Phi(H_2, x)$ are real for any vector x . This splitting of the matrix A into Hermitian components corresponds to splitting the function $\Phi(A, x)$ into real and imaginary parts.

If A is *normal*, i.e., $AA^* = A^*A$, then A may be put into diagonal form through a unitary transformation. The following applies to normal matrices:

Theorem 3. *If A is a normal matrix with eigenvalues a_1, \dots, a_n , then $W(A)$ is the convex hull of the points in the complex plane corresponding to the eigenvalues.*

Proof. This theorem is a generalization of the theorem for Hermitian matrices. One proves it in a similar fashion, by assuming that A is in diagonal form:

$$A = \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_n \end{pmatrix}.$$

Once again the vector x can be assumed to be real, hence

$$\Phi(A, x) = \sum_{\nu=1}^n a_\nu \xi_\nu^2 \quad \text{with} \quad \sum_{\nu=1}^n \xi_\nu^2 = 1.$$

In fact $W(A)$ is also bounded by the smallest convex polygon which encloses the points a_1, \dots, a_n . \square

2. AFFINE TRANSFORMATIONS OF NUMERICAL RANGES

If $W(A)$ is the numerical range of the matrix $A = H_1 + iH_2$, then the region of the complex plane which is the image of $W(A)$ under an affine transformation is once again the numerical range of a matrix. The general affine transformation of a point $z = \xi + i\eta$ in the Gaussian plane is represented by

$$z = \xi + i\eta \quad \rightarrow \quad z' = a\xi + ib\eta + c \quad (ab \neq 0, ab^{-1} \text{ not purely imaginary}).$$

If we denote this element of the affine group by $\tau = \tau_{abc}$, so that

$$\tau(z) = \tau_{abc}(z) = a\xi + ib\eta + c,$$

and if we define an affine transformation of a matrix by

$$\tau(A) = \tau_{abc}(A) = aH_1 + ibH_2 + cI_n,$$

then we have:

Theorem 4.

$$\tau(W(A)) = W(\tau(A)).$$

Proof. If $z = \xi + i\eta \in W(A)$, then $z = x_0^*Ax_0$ for a certain vector x_0 of norm one, then we have

$$\xi = x_0^*H_1x_0, \quad \eta = x_0^*H_2x_0$$

which gives us

$$\begin{aligned} \tau(z) &= ax_0^*H_1x_0 + ibx_0^*H_2x_0 + c \\ &= x_0^*(aH_1 + ibH_2 + cI_n)x_0 = x_0^*(\tau(A))x_0. \end{aligned}$$

□

If we say two matrices A, B are *affine equivalent*s when there exists τ such that $A = \tau(B)$, then the numerical ranges of affine equivalent matrices are affine transformations of each other; however, the converse is not true in general.

Only in the special case of a Hermitian matrix we have:

Theorem 5. *The numerical range $W(A)$ of a matrix A is a line segment exactly when A and a Hermitian matrix are affine equivalent.*s.

Proof. When A is an affine equivalent of a Hermitian matrix, then according to Theorems 2 and 4, $W(A)$ is a line segment. Conversely if $W(A)$ is a line segment, then A is an affine equivalent of a matrix B , where the numerical range $W(B)$ is a segment on the real axis. If we decompose B into its real and imaginary parts: $B = H_1 + iH_2$, then for any vector x of norm one:

$$x^*H_2x = 0$$

so

$$H_2 = 0 \quad \text{and} \quad H_1 = B.$$

□

Special affine transformations are a *rotation about the origin of an angle φ* , which we denote by τ_{aa0} with $a = e^{i\varphi}$, and the *parallel translation in the plane by the vector (γ_1, γ_2)* , which is represented by the element τ_{11c} with $c = \gamma_1 + i\gamma_2$.

The connection between unitary and affine transformations explains the following “swapping rule”:

Theorem 6.

$$\tau(U^*AU) = U^*(\tau(A))U$$

Proof. Indeed, we have

$$\begin{aligned} U^*\tau(A)U &= U^*(aH_1 + ibH_2 + cI_n)U \\ &= aU^*H_1U + ibU^*H_2U + cI_n = \tau(U^*AU). \end{aligned}$$

□

A square matrix A is said to have a *unitary decomposition* when by a unitary matrix U it can be put in the form

$$U^*AU = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$$

with square submatrices A_1, A_2 . This property of a matrix, to have a unitary decomposition, is invariant under affine equivalence, that is:

Theorem 7. *If a matrix A has a unitary decomposition, then all of its affine equivalent*s also have unitary decompositions.

Proof. A matrix $A = H_1 + iH_2$ has a unitary decomposition⁶ exactly when there exists a matrix $V \neq aI$, such that:

$$AV = VA, \quad A^*V = VA^*$$

or also

$$H_1V = VH_1, \quad H_2V = VH_2.$$

Now let A have a unitary decomposition, and let $V \neq aI$ commute with H_1 and H_2 . This implies

$$\begin{aligned} V\tau_{abc}(A) &= aVH_1 + ibVH_2 + cVI_n = aH_1V + ibH_2V + cI_nV = \tau_{abc}(A)V, \\ V(\tau_{abc}(A))^* &= \bar{a}VH_1 - i\bar{b}VH_2 + \bar{c}VI_n \\ &= \bar{a}H_1V - i\bar{b}H_2V + \bar{c}I_nV = (\tau_{abc}(A))^*V. \end{aligned}$$

Therefore the matrix $\tau_{abc}(A)$ also has a unitary decomposition. \square

In general an affine transformation of a matrix is not simultaneously a unitary transformation, since in general for a matrix A no element τ of the affine group exists such that

$$U^*AU = \tau(A).$$

Nevertheless, when this situation does occur, the numerical range $W(A)$ satisfies a certain symmetry condition, since for each transformation of the cyclic subgroup of the affine group generated by τ , $W(A)$ is mapped to itself.

For example, the numerical ranges of all matrices for which

$$U^*AU = \tau_{1,-1,0}(A) = \bar{A}$$

lie symmetric with respect to the real axis. Of these matrices, the class of matrices of dimension $2n$, for which⁷

$$V^*AV = \bar{A} \quad \text{with} \quad V = I_n \otimes \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

holds is particularly important for the theory of quaternion matrices.

Theorem 8. *If A is a unitary equivalent to both $\tau(A)$ and B , then also the matrices B and $\tau(B)$ are unitary equivalents.*

Proof. From

$$\tau(A) = UAU^* \quad \text{and} \quad B = W^*AW \quad (UU^* = 1, WW^* = 1)$$

it follows

$$\tau(B) = W^*\tau(A)W \quad \text{resp.} \quad \tau(A) = W\tau(B)W^*,$$

therefore, also

$$A = U^*W\tau(B)W^*U.$$

On the other hand,

$$A = WBW^* \quad \text{and from this} \quad B = W^*U^*W\tau(B)W^*UW.$$

\square

⁶See SPECHT [8].

⁷ $A \otimes B$ means the Kronecker product of two matrices A, B .

3. THE BOUNDARY GENERATING CURVE

Theorem 9. *If $A = H_1 + iH_2$ with $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_n$ the eigenvalues of H_1 and $\beta_1 \leq \beta_2 \leq \dots \leq \beta_n$ the eigenvalues of H_2 , then the points of $W(A)$ lie in the interior or on the boundary of the rectangle constructed by the lines $\xi = \alpha_1, \xi = \alpha_n; \eta = \beta_1, \eta = \beta_n$ positioned parallel to the axes. The sides of the rectangle share either one point (possibly with multiplicity > 1) or one closed interval with the boundary of $W(A)$.*

Proof. For the proof one notices that $\Phi(H_1, x)$ and $\Phi(H_2, x)$ are the real and imaginary parts of points in $W(A)$. The range of $\Phi(H_1, x)$ is the interval $[\alpha_1, \alpha_n]$ on the real axis, the range of $i\Phi(H_2, x)$ is the interval $[i\beta_1, i\beta_n]$ on the imaginary axis. From this the first part of our theorem follows immediately. The second part follows from the fact that the boundary of $W(A)$ shares with each side of the rectangle at least one point, since $\Phi(H_1, x)$ assumes the extreme values α_1, α_n of the numerical range of H_1 and $\Phi(H_2, x)$ assumes the extreme values β_1, β_n of the numerical range of H_2 . \square

A line in the complex plane is defined as a *support line of the region $W(A)$* , if it shares with the boundary of $W(A)$ either one point (possibly with multiplicity > 1) or one whole interval⁸. Therefore in particular the sides of the rectangle mentioned above are support lines.

In general if we denote the largest eigenvalue of the real part of a matrix A by $g(A)$, then we get a support line of $W(A)$ by

$$\xi = g(A).$$

Now if we rotate the numerical range by an angle $-\varphi$ while we switch to the matrix $e^{-i\varphi}A$, then for each value φ

$$\xi = g(e^{-i\varphi}A)$$

is a support line of $W(e^{-i\varphi}A)$. However, now $g(e^{-i\varphi}A)$ is the largest eigenvalue of the real part of

$$e^{-i\varphi}A = (\cos \varphi H_1 + \sin \varphi H_2) + i(\cos \varphi H_2 - \sin \varphi H_1)$$

and therefore the largest eigenvalue of

$$\cos \varphi H_1 + \sin \varphi H_2.$$

The eigenvalues of the latter matrix are obtained from the equation

$$|\cos \varphi H_1 + \sin \varphi H_2 - \lambda I_n| = 0.$$

The largest among these is $\lambda_n = g(e^{-i\varphi}A)$. If we rotate back the numerical range by the angle $+\varphi$, then $W(e^{-i\varphi}A)$ goes back to $W(A)$, but the line

$$\xi = g(e^{-i\varphi}A)$$

goes to the line

$$(2) \quad \xi \cos \varphi + \eta \sin \varphi - g(e^{-i\varphi}A) = 0.$$

Thus this line is a support line of $W(A)$. If φ is varied over all values between 0 and 2π then (2) yields every support line of $W(A)$. From this it follows:

⁸For all the concepts of the theory of convex regions see Bonneson-Fenchel [1].

Theorem 10. *To every complex matrix $A = H_1 + iH_2$ through the equation*

$$f_A(u, v, w) \equiv |H_1u + H_2v + I_n w| = 0$$

is associated a curve of class n in homogenous line coordinates in the complex plane. The convex hull of this curve is the numerical range of the matrix A .

Hereby we consider the points of the complex plane as finite points in the projective plane (i.e., points on the plane not lying on the line $u = 0, v = 0, w = 1$).^[II]

Proof. The curve is supported by, in particular, the line (2) with line coordinates

$$(\cos \varphi, \sin \varphi, -g(e^{-i\varphi} A))$$

for arbitrary φ . Thus the set of all support lines of $W(A)$ are generating elements of the curve. Thereby, each of these lines is characterized with respect to the generating elements of the curve parallel to itself by the fact that it is extreme, i.e., it does not lie between two elements of the curve parallel to itself. From this the proof follows immediately. \square

The curve of class n associated to the matrix A in this way may be called the *boundary generating curve of the matrix A* .

If the values $u = 1, v = i$ (respectively $u = 1, v = -i$) are taken in the equation of the boundary generating curve of the matrix A , then the solution which is obtained for w is the negative eigenvalue $-a_\nu$ of A (respectively, the negative eigenvalue $-\bar{a}_\nu$ of A^*). On the other hand, the lines with line coordinates $g_\nu : (1, i, -a_\nu)$ respectively $\bar{g}_\mu : (1, -i, -\bar{a}_\mu)$ ($\mu, \nu = 1, \dots, n$) represent lines through one of the two circular points. But the points of intersection of g_ν with \bar{g}_ν ($\nu = 1, \dots, n$) are the n real foci of the boundary generating curve. The point coordinates of these points of intersection are

$$(\Re(a_\nu), \Im(a_\nu), 1) \quad (\nu = 1, \dots, n).$$

From this it follows:

Theorem 11. *The real foci of the characteristic curve of the matrix A are the eigenvalues of A .*

One can say more about the position of the eigenvalues in the numerical range:

Theorem 12. *If the matrix A does not have a unitary decomposition, then the eigenvalues lie in the interior of $W(A)$.*

Proof. Through an eigenvalue $a = \alpha + i\beta$ of A , which lies on the boundary of the numerical range, goes a support line of the convex set $W(A)$, which is why in the notation before Theorem 10 for a certain angle φ_0 the equation

$$\alpha \cos \varphi_0 + \beta \sin \varphi_0 - g(e^{-i\varphi_0} A) = 0$$

holds. To the eigenvalue a corresponds a normalized eigenvector x_0 :

$$Ax_0 = ax_0 \quad \text{and} \quad a = x_0^* Ax_0 = x_0^* H_1 x_0 + ix_0^* H_2 x_0 = \alpha + i\beta.$$

Now if we switch to $e^{-i\varphi_0} A$ and consider only the real parts, then we have

$$\alpha \cos \varphi_0 + \beta \sin \varphi_0 = x_0^* (H_1 \cos \varphi_0 + H_2 \sin \varphi_0) x_0 = g(e^{-i\varphi_0} A).$$

Hereby $\alpha \cos \varphi_0 + \beta \sin \varphi_0$ is an extreme eigenvalue of the Hermitian matrix $H_1 \cos \varphi_0 + H_2 \sin \varphi_0$, thus x_0 is also an eigenvector of $H_1 \cos \varphi_0 + H_2 \sin \varphi_0$ associated to this eigenvalue:

$$(H_1 \cos \varphi_0 + H_2 \sin \varphi_0)x_0 = (\alpha \cos \varphi_0 + \beta \sin \varphi_0)x_0.$$

Consequently x_0 is simultaneously an eigenvector of H_1 and H_2 . If one complements $x_0 = x_1$ to a unitary matrix $U = (x_1, \dots, x_n)$, then

$$U^*AU = \begin{pmatrix} a & 0 \\ 0 & A_1 \end{pmatrix}$$

contradicts the assumption that A does not have a unitary decomposition. \square

Theorem 13. *Each singular point a on the boundary of the numerical range $W(A)$ of the matrix A is an eigenvalue of the matrix, and there exists a unitary matrix U corresponding to a , such that A may be decomposed in the form*

$$U^*AU = \begin{pmatrix} a & 0 \\ 0 & A_1 \end{pmatrix}.$$

Proof. If there are support lines through $a = \alpha + i\beta$ in different directions, then there is an entire interval $[\varphi_0, \varphi_1]$ ($\varphi_0 \neq \varphi_1$), so that for each value φ in that interval

$$\alpha \cos \varphi + \beta \sin \varphi - g(e^{-i\varphi}A) = 0,$$

or, respectively

$$\alpha u + \beta v + w = 0 \quad \text{with} \quad u = \cos \varphi, v = \sin \varphi, w = -g(e^{-i\varphi}A).$$

Moreover,

$$|H_1u + H_2v + I_nw| = 0,$$

from which follows an identity of the form

$$|H_1u + H_2v + I_nw| \equiv (\alpha u + \beta v + w)F(u, v, w),$$

in which $F(u, v, w)$ is homogenous of order $n - 1$. If we set $u = -1, v = -i$, then we find

$$|I_nw - A| \equiv (w - (\alpha + i\beta))F(-1, -i, w).$$

Therefore $a = \alpha + i\beta$ is an eigenvalue of the matrix A . From the proof of Theorem 12 it also follows directly that for an eigenvalue which lies on the boundary of $W(A)$, the matrix has a unitary decomposition of the given form. \square

In particular from this it follows:

Theorem 14. *The boundary of the numerical range of a matrix without a unitary decomposition is smooth.*

4. PROPERTIES OF THE BOUNDARY GENERATING CURVE

The boundary generating curve of class n associated with a matrix A of dimension n is given by the equation in line coordinates

$$(3) \quad f_A(u, v, w) \equiv |H_1u + H_2v + I_nw| = 0.$$

From this follows one important property of the boundary generating curve:

Theorem 15. *The boundary generating curve has n real tangents in each arbitrarily given real direction.*

Proof. If a real direction is given through $(\cos \varphi, \sin \varphi)$, then the equation (3) yields n real values for w , since the eigenvalues of the Hermitian matrix $H_1 \cos \varphi + H_2 \sin \varphi$ are real. It follows that the curve possesses n real tangents in every (real) direction⁹. \square

From this property of the boundary generating curve can be deduced directly:

Theorem 16. *The boundary generating curve does not have any real inflectional tangents.*

Theorem 17. *The real points of the boundary generating curve are all finite.*

Theorem 18. *If the matrix $\tau(A)$ is an affine equivalent to A , then the boundary generating curve of $\tau(A)$ is derived from the boundary generating curve of A , in the sense that the points of the boundary generating curve of A are subject to the affine transformation τ .*

Proof of Theorem 18. Let $A = H_1 + iH_2$ and $a = \alpha_1 + i\alpha_2, b = \beta_1 + i\beta_2, c = \gamma_1 + i\gamma_2$, $\tau(A) = aH_1 + ibH_2 + cI_n = (\alpha_1H_1 - \beta_2H_2 + \gamma_1I_n) + i(\alpha_2H_1 + \beta_1H_2 + \gamma_2I_n)$. Then the boundary generating curve of $\tau(A)$ is given by

$$|(\alpha_1H_1 - \beta_2H_2 + \gamma_1I_n)u + (\alpha_2H_1 + \beta_1H_2 + \gamma_2I_n)v + I_nw| = 0.$$

If τ is considered as a special projective transformation and also the line coordinates u, v, w are changed under the transformation contragredient to τ , then the boundary generating curve of $\tau(A)$ has the equation

$$|H_1u' + H_2v' + I_nw'| = 0,$$

where u', v', w' are the transformed line coordinates. But from this it follows that the coordinates of the boundary generating curve of A are affected in the transition to $\tau(A)$ only by the affine transformation τ . \square

Theorem 19. *Through each real point in the plane goes an even number or an odd number of real lines tangent to the boundary generating curve, depending on whether n is even or odd.*

Proof. It is enough to prove that the origin $w = 0$ has this property, since any finite point can be brought into the origin by a parallel translation. For points at infinity the assertion is contained in Theorem 15. But if we set $w = 0$, then (3) gives us an equation in u and v of degree n with real coefficients. The real solutions of this equation are even in number exactly when n is an even number. \square

Theorem 20. *The number of real cusps of an irreducible boundary generating curve of a matrix of dimension n is even or odd, depending on whether n is even or odd.*

Proof. The tangents in the cusps are given in line coordinates by simultaneously solving equation (3) and the equation of the curve determined by

$$\begin{vmatrix} \frac{\partial^2 f_A}{\partial u^2} & \frac{\partial^2 f_A}{\partial u \partial v} & \frac{\partial^2 f_A}{\partial u \partial w} \\ \frac{\partial^2 f_A}{\partial v \partial u} & \frac{\partial^2 f_A}{\partial v^2} & \frac{\partial^2 f_A}{\partial v \partial w} \\ \frac{\partial^2 f_A}{\partial w \partial u} & \frac{\partial^2 f_A}{\partial w \partial v} & \frac{\partial^2 f_A}{\partial w^2} \end{vmatrix} = 0,$$

⁹Brunn [2] has thoroughly studied unicursal curves with this property. In the following part his theorems on curves of class n are restated.

and dual to the Hessian curve. For an irreducible boundary generating curve there is a finite number of points that simultaneously solve both equations, in fact the number is $3n(n-2)$. Since both equations have real coefficients, the number of real solutions is even or odd, depending on whether n is even or not. However, the proposition need not be true in the case where the boundary generating curve is reducible. \square

5. THE SINGULAR DIRECTIONS OF THE BOUNDARY GENERATING CURVE

The boundary generating curve of a matrix is given by an equation of the form

$$(4) \quad f_A(u, v, w) \equiv |H_1u + H_2v + I_nw| \equiv w^n + C_1(u, v)w^{n-1} + \dots + C_n(u, v) = 0.$$

Line singularities are present exactly when for a pair of values u, v this equation has multiple solutions for w . For the presence of double roots of equation (4) it is necessary and sufficient that the discriminant $D(u, v)$ of (4) vanishes. This discriminant is homogeneous in u and v of order $n^2 - n$. Therefore there will be $n^2 - n$ directions in which the boundary generating curve has singular tangents. However, there exists yet another connection with the discriminant $D(u, v)$. Namely, we consider the matrix $H = -(H_1u + H_2v)$ and the number of matrices V which commute with it, i.e., for which there is an equation of the form

$$(5) \quad VH = HV \quad (V = V(u, v)).$$

In general the Hermitian matrix H has exactly n linearly independent matrices that commute with it, and the number of matrices commuting with H is larger than n if and only if H has eigenvalues of multiplicity greater than one. That is, for a given pair of numbers u, v there are more than n matrices commuting with H exactly when (4) has a root of multiplicity greater than one.

Equation (5) can be considered as a linear system for the n^2 elements of the matrix V . The matrix of coefficients of this system is

$$M = I \otimes H^T - H \otimes I.$$

Therefore in general the matrix M will have rank $n^2 - n$. The subdeterminants of dimension $n^2 - n$ are zero if and only if H has more than n commuting matrices. But from this it follows that the subdeterminants of dimension $n^2 - n$ vanish if and only if the discriminant $D(u, v)$ of (4) vanishes. Since both the subdeterminants of dimension $n^2 - n$ and $D(u, v)$ are homogeneous in u, v of order $n^2 - n$, it follows that the subdeterminants of dimension $n^2 - n$ actually only differ from the discriminant $D(u, v)$ of (4) each by a constant factor.

Thus the $n^2 - n$ singular directions are also determined by the vanishing of the subdeterminants of dimension $n^2 - n$ of the matrix $I \otimes H^T - H \otimes I$.

But they may also be determined in yet another manner:

The matrix $A = H_1 + iH_2$ can be modified by a unitary transformation U_1 so that the Hermitian part D_1 of

$$B = U_1^* A U_1 = U_1^* H_1 U_1 + i U_1^* H_2 U_1 = D_1 + iK_2$$

is a diagonal matrix. Therefore, we have

$$D_1 = \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \alpha_n \end{pmatrix}, \quad K_2 = U_1^* H_2 U_1.$$

Hereby $\alpha_1, \dots, \alpha_n$ are the (real) eigenvalues of H_1 . Similarly the matrix A can be changed by means of a unitary matrix U_2 into another matrix C , whose skew-Hermitian part is represented by a diagonal matrix D_2 . Therefore

$$C = U_2^* A U_2 = U_2^* H_1 U_2 + i U_2^* H_2 U_2 = K_1 + i D_2$$

with

$$K_1 = U_2^* H_1 U_2, \quad D_2 = \begin{pmatrix} \beta_1 & 0 & \cdots & 0 \\ 0 & \beta_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \beta_n \end{pmatrix}.$$

Here β_1, \dots, β_n are the (real) eigenvalues of H_2 . The form

$$f_A \equiv f_A(u, v, w) \equiv |H_1 u + H_2 v + I_n w|$$

is not changed by a unitary transformation:

$$\begin{aligned} f_{U^* A U}(u, v, w) &\equiv |U^* H_1 U u + U^* H_2 U v + U^* I_n U w| \\ &\equiv |U^* |f_A(u, v, w)| U| \equiv f_A(u, v, w). \end{aligned}$$

Consequently this gives us

$$(6) \quad f_A \equiv f_B \equiv f_C.$$

The principal submatrices of B of order $n-1$, produced by deleting the ν th row and column (for $\nu = 1, \dots, n$), are denoted by B_ν , and similarly the principal submatrices of C are denoted C_ν . The boundary generating curves of the submatrices B_ν and, respectively, C_ν , yield two real families, \mathcal{B} respectively \mathcal{C} , of curves of class $n-1$ with real parameters λ_ν :

$$\sum_{\nu=1}^n \lambda_\nu f_{B_\nu} = 0 \quad \text{respectively} \quad \sum_{\nu=1}^n \lambda_\nu f_{C_\nu} = 0.$$

The generic curve of the family \mathcal{B} respectively \mathcal{C} is denoted by $b(\lambda_1, \dots, \lambda_n)$ respectively $c(\lambda_1, \dots, \lambda_n)$. Then we have:

Theorem 21. *The families \mathcal{B} , \mathcal{C} have at least one curve in common, namely*

$$b(1, \dots, 1) = c(1, \dots, 1).$$

Proof. If we differentiate f_A with respect to w , then it follows on account of (6)

$$\frac{\partial f_A}{\partial w} \equiv \frac{\partial f_B}{\partial w} \equiv \sum_{\nu=1}^n f_{B_\nu} \quad \text{and} \quad \frac{\partial f_A}{\partial w} \equiv \frac{\partial f_C}{\partial w} \equiv \sum_{\nu=1}^n f_{C_\nu}.$$

Thus

$$\sum_{\nu=1}^n f_{B_\nu} \equiv \sum_{\nu=1}^n f_{C_\nu} \quad \text{respectively} \quad b(1, \dots, 1) = c(1, \dots, 1).$$

□

Theorem 22. *Let $\alpha_1, \dots, \alpha_n$ be the eigenvalues of H_1 , let β_1, \dots, β_n be the eigenvalues of H_2 , and let $A = H_1 + iH_2$. Then the singular tangents of the boundary generating curve of A are contained in the set of lines tangent to both of the curves*

$$b(\alpha_1, \dots, \alpha_n), \quad c(\beta_1, \dots, \beta_n).$$

Proof. We have

$$\frac{\partial f_A}{\partial u} \equiv \frac{\partial f_B}{\partial u} \equiv \sum_{\nu=1}^n \alpha_\nu f_{B_\nu}, \quad \frac{\partial f_A}{\partial v} \equiv \frac{\partial f_C}{\partial v} \equiv \sum_{\nu=1}^n \beta_\nu f_{C_\nu}.$$

For the curve to have a singular line through $f_A = 0$ it must be true that

$$\frac{\partial f_A}{\partial u} = 0, \quad \frac{\partial f_A}{\partial v} = 0.$$

The assertion follows from this. \square

Theorem 23. *Any common tangent to the curves $b(\alpha_1, \dots, \alpha_n)$, $c(\beta_1, \dots, \beta_n)$ is a singularity of the curve defined by $f_A(u, v, w) = 0$, exactly when it is also an element of the curve*

$$b(1, \dots, 1) = c(1, \dots, 1).$$

Proof. For such an element we have namely

$$\frac{\partial f_A}{\partial u} = \frac{\partial f_A}{\partial v} = \frac{\partial f_A}{\partial w} = 0,$$

whence by Euler's theorem for homogeneous functions it follows also that $f_A = 0$; thus the element belongs to the curve. But each element for which

$$\frac{\partial f_A}{\partial u} = \frac{\partial f_A}{\partial v} = \frac{\partial f_A}{\partial w} = f_A = 0$$

holds, is singular. Conversely it follows from the singularity of an element of the curve that

$$\frac{\partial f_A}{\partial u} = \frac{\partial f_A}{\partial v} = f_A = 0$$

holds. From this follows according to the theorem for homogeneous functions

$$\frac{\partial f_A}{\partial w} = 0.$$

\square

Theorem 24. *A real tangent of the boundary generating curve is singular exactly when it is simultaneously a generating element of the curve*

$$b(1, \dots, 1) \equiv c(1, \dots, 1).$$

Proof. The coordinates of a point on the boundary generating curve are of the form

$$\frac{\partial f_A}{\partial u}, \quad \frac{\partial f_A}{\partial v}, \quad \frac{\partial f_A}{\partial w}.$$

Since the boundary generating curve does not possess a real ideal point with coordinates $(\alpha, \beta, 0)$, for a real element of the curve defined by $f_A = 0$, the equation

$$\frac{\partial f_A}{\partial w} = 0$$

is always followed in turn by the equations

$$\frac{\partial f_A}{\partial u} = 0, \quad \frac{\partial f_A}{\partial v} = 0.$$

Thus the real elements of the curve $f_A = 0$, which also belong to the curve $b(1, \dots, 1) \equiv c(1, \dots, 1)$, are simultaneously elements of the curves $b(\alpha_1, \dots, \alpha_n)$, $c(\beta_1, \dots, \beta_n)$ and are thereby singular. \square

6. EXAMPLES

1. For the matrix

$$A = H_1 + iH_2 = \begin{pmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & 1 & 1 & 0 \\ 0 & 0 & 0 & \cdot & 0 & 1 & 1 \end{pmatrix}$$

of dimension n with no unitary decomposition we have

$$H_1 = \begin{pmatrix} 1 & \frac{1}{2} & 0 & \cdot & \cdot & \cdot & 0 \\ \frac{1}{2} & 1 & \frac{1}{2} & \cdot & \cdot & \cdot & 0 \\ 0 & \frac{1}{2} & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & \cdot & \cdot & 0 & \frac{1}{2} & 0 \end{pmatrix}, \quad H_2 = \begin{pmatrix} 0 & \frac{i}{2} & 0 & \cdot & \cdot & \cdot & 0 \\ -\frac{i}{2} & 0 & \frac{i}{2} & \cdot & \cdot & \cdot & 0 \\ 0 & -\frac{i}{2} & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & -\frac{i}{2} & 0 & \frac{i}{2} \\ 0 & 0 & \cdot & \cdot & 0 & -\frac{i}{2} & 0 \end{pmatrix}$$

Thus it follows

$$f_A(u, v, w) \equiv f_n \equiv |H_1 u + H_2 v + I_n w| \equiv \begin{vmatrix} u+w & \frac{u+iv}{2} & 0 & \cdot & \cdot & 0 \\ \frac{u-iv}{2} & u+w & \frac{u+iv}{2} & \cdot & \cdot & 0 \\ 0 & \frac{u-iv}{2} & u+w & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \frac{u-iv}{2} & u+w \end{vmatrix},$$

thereby the recursive formula

$$f_n = (u+w)f_{n-1} - \frac{1}{4}(u^2 + v^2)f_{n-2}$$

with

$$f_1 = u+w, \quad f_2 = (u+w)^2 - \frac{1}{4}(u^2 + v^2).$$

If we set

$$u+w = C_1, \quad \frac{1}{4}(u^2 + v^2) = C_2,$$

then with this notation $f_n(u, v, w) = g_n(C_1, C_2) = g_n$ becomes

$$g_n = C_1 g_{n-1} - C_2 g_{n-2}$$

and, since because the boundary generating curve is finite for all generating elements, $C_2 = \frac{1}{4}(u^2 + v^2)$ is always different from zero,

$$\frac{g_{2m}}{C_2^m} = \frac{C_1}{C_2} \frac{g_{2m-1}}{C_2^{m-1}} - \frac{g_{2m-2}}{C_2^{m-1}},$$

$$\frac{g_{2m+1}}{C_2^m} = C_1 \frac{g_{2m}}{C_2^m} - \frac{g_{2m-1}}{C_2^{m-1}}.$$

Now g_{2m} is a homogeneous polynomial of degree m in C_1^2, C_2 , so

$$\frac{g_{2m}}{C_2^m} = h_m \left(\frac{C_1^2}{C_2} \right)$$

is a polynomial in $\frac{C_1^2}{C_2}$ of the same degree; likewise

$$\frac{g_{2m+1}}{C_2^m} = C_1 k_m \left(\frac{C_1^2}{C_2} \right).$$

Let the zeros of $h_m(x)$ respectively $k_m(x)$ be $\gamma_1, \dots, \gamma_m$ respectively $\delta_1, \dots, \delta_m$, thus

$$\begin{aligned} f_{2m}(u, v, w) &\equiv (C_1^2 - \gamma_1 C_2) \dots (C_1^2 - \gamma_m C_2), \\ f_{2m+1}(u, v, w) &\equiv C_1 (C_1^2 - \delta_1 C_2) \dots (C_1^2 - \delta_m C_2), \end{aligned}$$

where each quadratic factor is an expression of the form

$$C_1^2 - \alpha C_2 \equiv (u + w)^2 - \alpha(u^2 + v^2),$$

which corresponds to a circle around the point 1 with the radius $\sqrt{\alpha}$.

Thus the boundary generating curve of the matrix A in the case of even dimension $n = 2m$ can be decomposed into m concentric circles, while in the case of odd dimension $n = 2m + 1$ it can be decomposed into m concentric circles and a point. This example shows that the boundary generating curve may be decomposable even in the case of matrices without unitary decompositions.

2. Let

$$A = \begin{pmatrix} 0 & -\frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{\sqrt{2}} & -\frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} \end{pmatrix} + i \begin{pmatrix} 0 & \frac{i}{2} & 0 \\ -\frac{i}{2} & 0 & \frac{i}{2} \\ 0 & -\frac{i}{2} & 0 \end{pmatrix}.$$

The boundary generating curve becomes, when we set $w = 1$, defined by the equation

$$\frac{1}{2}\sqrt{2}uv^2 - 2v^2 - 2u^2 + 4 = 0.$$

The eigenvalues of A are

$$z_i = -\sqrt[3]{\frac{1}{4\sqrt{2}}} \epsilon_i \quad (i = 1, 2, 3),$$

where the ϵ_i are the cube roots of unity. The curve possesses two components, an oval and a closed tricuspid curve in its interior (see Fig. 1).

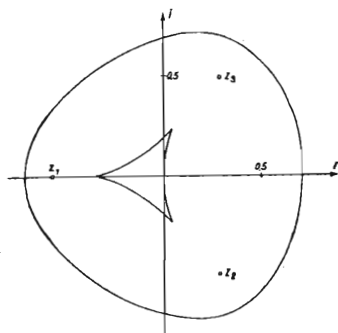


Fig. 1.

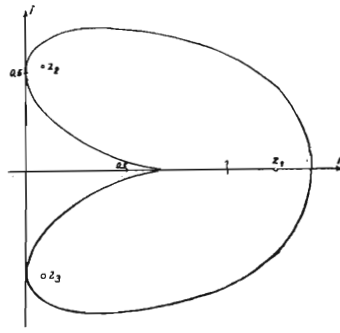


Fig. 2.

3. Let

$$A = \begin{pmatrix} 0 & -\frac{1}{2} & 0 \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & \frac{1}{2} & \sqrt{2} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix} + i \begin{pmatrix} 0 & \frac{i}{2} & 0 \\ -\frac{i}{2} & 0 & \frac{i}{2} \\ 0 & -\frac{i}{2} & 0 \end{pmatrix}.$$

For the boundary generating curve we get the equation

$$uv^2 - \sqrt{2}v^2 - 4u - 2\sqrt{2} = 0.$$

The eigenvalues z_i of A are the roots of the equation

$$z^3 - \sqrt{2}z^2 + \frac{1}{2}z - \frac{1}{4}\sqrt{2} = 0.$$

The curve consists of one component and possesses one cusp (see Fig. 2).

7. THE BOUNDARY GENERATING CURVE OF MATRICES OF DIMENSION 2 OR 3

If the dimension n of the matrix A is two, then also the boundary generating curve of A is a curve of class two. If A is normal, then the boundary generating curve consists of the two points which belong to the eigenvalues. If A does not have a unitary decomposition, then the boundary generating curve is a second order curve, and in fact an ellipse, whose two foci coincide with the eigenvalues of A . No other curve of second order occurs as a boundary generating curve, since the boundary generating curve must always be finite.

If the dimension n of the matrix A is three, then also the boundary generating curve of A is a curve of class three. It shall be investigated which types of curves appear in this case.

First of all let A be normal. Then the boundary generating curve of A consists of the points belonging to the three eigenvalues.

If the matrix A has a unitary decomposition, but it is not normal, then by a unitary transformation it can be brought into the form

$$\begin{pmatrix} \alpha_1 & 0 \\ 0 & A_1 \end{pmatrix},$$

in which A_1 is a matrix of dimension 2 without a unitary decomposition. The boundary generating curve then consists of the point a_1 and the boundary generating ellipse of the matrix A_1 .

Now suppose A does not have a unitary decomposition. Then it is possible that even though A does not have a unitary decomposition, the boundary generating curve may be decomposed into a point and an ellipse (see § 6, Example 1 for $n = 3$). Thereby the point itself is always an eigenvalue of the matrix, since it corresponds to a linear factor in the left-hand side of the equation $f_A(u, v, w) = 0$. As was shown in the proof of Theorem 13, this factor corresponds once again to an eigenvalue. From this it now follows that the point must lie in the interior of the ellipse. Namely if its position were exterior, then according to Theorem 13, A would have a unitary decomposition, contrary to the assumption. If instead the point would lie on the ellipse, then an eigenvalue of the matrix A would lie on the boundary of the numerical range, which on account of Theorem 12 contradicts the fact that A does not have a unitary decomposition.

Thus the only case that remains is that where the boundary generating curve is itself an irreducible curve. The number of real cusps of an (irreducible) class three curve is 1 or 3. According to Theorem 17 the order of the boundary generating curve is either 4 or 6.

Theorem 25. *The matrix A of dimension 3 can be modified through an affine transformation so that the equation of its boundary generating curve in nonhomogeneous line coordinates takes the form*

$$(7) \quad (v-1)^2(u-1) = \alpha u^3 + \beta u^2 + \gamma u + \delta$$

with real $\alpha, \beta, \gamma, \delta$.

Proof. To every class three curve that is not decomposable a triangle of projective coordinates K_1 can be designated with the point coordinates x_1, x_2, x_3 and the line coordinates x'_1, x'_2, x'_3 so that in these the curve satisfies the equation¹⁰

$$(8) \quad x_1'^2 x_2' = \alpha x_2'^3 + \beta x_2'^2 x_3' + \gamma x_2' x_3'^2 + \delta x_3'^3.$$

If the curve is given in the complex plane, then the triangle K_1 may be acted on by an affine transformation, in such a way that that the points which in K_1 have coordinates $(1, 0, 0), (0, 1, 0), (0, 0, 1)$, may be mapped into the points corresponding to $1, i, 0$ in the Gaussian plane. These three latter points form a new coordinate triangle K_2 with the point coordinates y_1, y_2, y_3 and the line coordinates y'_1, y'_2, y'_3 . Now if the curve of class three is, in particular, the boundary generating curve of a matrix A , and if A is acted on by an affine transformation so that K_1 is carried over to K_2 , then a new matrix is obtained, whose boundary generating curve has the equation in K_1 given by (7) through the substitution $x'_j \rightarrow y'_j$. Between the homogeneous line coordinates u, v, w used further above and the line coordinates y'_1, y'_2, y'_3 there exists a relation of the form

$$\varrho_1 y'_1 = v - w, \quad \varrho_2 y'_2 = u - w, \quad \varrho_3 y'_3 = w$$

Thereby the proof of Theorem 25 is completed. \square

The discussion of equation (7) now gives all possible cases, as we will now show. Thereto we write (7) in the homogeneous form

$$(9) \quad y_1'^2 y_2' = \alpha (y_2' - a_1 y_3') (y_2' - a_2 y_3') (y_2' - a_3 y_3').$$

The particular differences become clearer when y'_1, y'_2, y'_3 are interpreted as homogeneous point coordinates. Then in each case (9) represents a curve of order three, and now out of all curves of order three, we only have to single out the types of curves whose dual curves correspond to boundary generating curves.

We first take a_1, a_2, a_3 as real and distinct¹¹:

1. $a_1 < a_2 < a_3$. The dual curve has two components. They consist of an oval and an infinite branch with three real points of inflection. Therefore the proper curve consists of an oval and a tricuspid component in its interior. The curve is of order six. Example 2 in § 6 showed that such a curve can indeed appear as a boundary generating curve.

2. *Exactly two of the a_i ($i = 1, 2, 3$) are equal.* In this case there are two possible forms of curves:

a) The dual curve has an isolated point; the proper curve contains a line. Therefore it is not finite and hence cannot be the boundary generating curve of a matrix.

¹⁰See the normal form corresponding to curves of order 3; for instance in WIELEITNER [11], p. 245.

¹¹We thereby connect closely to Newton's classification of curves of order 3; see for instance WIELEITNER [11], p. 245.

b) The dual curve has one component, and it has one real point of inflection and one node. The proper curve has one real cusp and one double tangent. It is of order four. Example 3 in § 6 showed that such a curve can appear as a boundary generating curve of a matrix.

3. *All a_i are equal.* The dual curve has one real cusp; the proper curve is not the boundary generating curve of a matrix, since it has one real turning point.

Finally, there still remains the case:

4. *Two of the a_i are complex conjugates.* Then the dual curve has one component. In this case there is not any real point in the plane such that every line through it three real points of intersection with the dual curve. But then the proper curve cannot have three real generating elements in each direction and therefore cannot be the boundary generating curve of a matrix A of dimension 3.

Thereby we have obtained:

Theorem 26. *In the case $n = 3$ a matrix A can only possess the following types of curves as its boundary generating curve:*

1. *three points,*
2. *a point and an ellipse,*
3. *a curve of order 4 with a double tangent and a cusp,*
4. *a proper^[III] curve of order 6, consisting of an oval and a curve with three cusps lying in its interior.*

8. THE MINIMAL EQUATION OF THE BOUNDARY GENERATING CURVE

The equation

$$f_A(u, v, w) \equiv |H_1u + H_2v + I_nw| \equiv w^n + C_1(u, v)w^{n-1} + \dots + C_n(u, v) = 0$$

for the boundary generating curve of a matrix $A = H_1 + iH_2$ may be understood as the characteristic equation of the polynomial matrix $H = -(uH_1 + vH_2)$. The coefficients $C_\nu(u, v)$ are thereby in each case equal to the sum of the determinants of the principal submatrices of H of order ν , therefore they are homogeneous polynomials in u, v of order ν . Exactly as in the case of constant matrices we have:

$$f_A(u, v, H) \equiv H^n + C_1(u, v)H^{n-1} + \dots + C_n(u, v) = 0.$$

For the proof¹² we consider both sides of the equation¹³

$$(H - wI_n)^{(n-1)}(H - wI_n) = |H - wI_n|I_n$$

after expansion of powers of w . Then the coefficients are polynomials in u, v . Since the corresponding coefficients on both sides of the equation are identical, the equation remains true when we replace w by H on both sides. But then the left-hand side is zero. Therefore it is true that $f_A(u, v, H) = 0$.

Now let $m(u, v, w)$ be a polynomial with complex coefficients:

$$m(u, v, w) = m_0(u, v) + m_1(u, v)w + \dots + m_k(u, v)w^k,$$

of smallest possible degree k , for which $m(u, v, H) = 0$ holds. Moreover, the polynomials $m_\chi(u, v)$ ($\chi = 1, \dots, k$) may be taken to be relatively prime. We call such a polynomial a *minimal polynomial of H* or also a *minimal polynomial of the boundary generating curve of A* .

¹²cf. MACDUFFEE [3]

¹³The matrix $M^{(n-1)}$ is thereby produced from the matrix M of dimension n , so that all of its elements are replaced by their algebraic complements.

Then if any other polynomial $p(u, v, w)$ is a polynomial of the same type, for which also $p(u, v, H) = 0$ holds, then $m(u, v, w)$ is a factor of $p(u, v, w)$. In proof of this we seek three polynomials $\varphi(u, v)$, $q(u, v, w)$, $r(u, v, w)$, such that

$$\varphi(u, v)p(u, v, w) = q(u, v, w)m(u, v, w) + r(u, v, w),$$

where $r(u, v, w)$ is smaller in the degree of w than $m(u, v, w)$ is. If we then set $w = H$, then it follows from the minimal property of $m(u, v, w)$ that $r(u, v, w)$ must vanish identically. Therefore $m(u, v, w)$ is a factor of $\varphi(u, v)p(u, v, w)$. Since $m(u, v, w)$ does not contain a factor depending only on u, v , it follows that $m(u, v, w) \mid p(u, v, w)$. From this it follows immediately that the minimal polynomial of H is uniquely determined.

In particular this gives us

$$m(u, v, w) \mid f_A(u, v, w).$$

If we now decompose $f_A(u, v, w)$ into irreducible factors:

$$f_A(u, v, w) \equiv f_1^{\gamma_1}(u, v, w)f_2^{\gamma_2}(u, v, w) \cdots f_s^{\gamma_s}(u, v, w),$$

then we can assume that the leading coefficient of each irreducible factor is 1, since in $f_A(u, v, w)$ the coefficient of w^n is equal to 1. Likewise it follows for $m(u, v, w)$, that $m_k(u, v)$ is a constant and may be assumed equal to 1.

In the decomposition of $m(u, v, w)$ into irreducible factors, only such irreducible factors which also appear in the factorization of $f_A(u, v, w)$ can appear. However, we have, as in the case of constant matrices¹⁴:

Theorem 27. *In the decomposition of $m(u, v, w)$ into irreducible factors each irreducible factor of $f_A(u, v, w)$ appears exactly once.*

The proof proceeds word for word exactly as in the case of constant matrices.

If $d(u, v, w)$ is the greatest common factor of all elements of $H^{(n-1)}$, then as in the case of constant matrices we have the equation $m(u, v, w) = f_A(u, v, w)/d(u, v, w)$, so that therefore the minimal polynomial of the boundary generating curve can always be written as a rational curve.

In relation to this we notice:

Theorem 28. *If $A = H_1 + iH_2$ is a matrix of dimension n without a unitary decomposition and if the degree k of the minimal polynomial of the boundary generating curve of A is such that $k \leq 2$, then it also is true that $n \leq 2$.*

Proof. For $k = 1$ it follows that H_1 and H_2 differ from the identity matrix by only a constant factor, so that A itself is normal.

For $k = 2$ we have an equation of the form:

$$(10) \quad (H_1u + H_2v)^2 + a_1(u, v)(H_1u + H_2v) + a_2(u, v)I_n = 0.$$

Now we consider the ring $\mathfrak{R}(H_1, H_2)$ generated by the matrices H_1, H_2 . On account of (10) it is possible for the terms

$$H_1^2, H_2^2, H_1H_2 + H_2H_1$$

to be expressed as a linear combination of the matrices I_n, H_1, H_2 with complex constant coefficients. From this it follows immediately that it is possible for all elements of $\mathfrak{R}(H_1, H_2)$ to be represented as linear combinations of the four elements

$$I_n, H_1, H_2, H_1H_2$$

¹⁴See MACDUFFEE [3]

with complex coefficients. Therefore $\Re(H_1, H_2)$ is an algebra of rank four over the field of complex numbers. On the other hand, it is the group generated by two Hermitian matrices, therefore the ring $\Re(H_1, H_2)$ is fully reducible¹⁵. Therefore when $n > 2$, then H_1, H_2 must be decomposed through the same unitary transformation into this form, because otherwise on account of the theorem of Burnside¹⁶ the rank of $\Re(H_1, H_2)$ should be $n^2 > 4$. But then A would have a unitary decomposition, contrary to hypothesis. \square

In another form we also have:

Theorem 28a. *If A is a matrix of dimension n and if $k \leq 2$, then A has a unitary decomposition.*

Geometrically interpreted, this theorem contains the remark that a matrix A of dimension n always has a unitary decomposition when its boundary generating curve consists of multiple copies of an ellipse. That is close to the hypothesis that a matrix of dimension n always has a unitary decomposition when its boundary generating curve contains irreducible components of multiplicity > 1 , or, equivalently, when the degree k of the minimal polynomial of $-(H_1u + H_2v)$ is smaller than n .^[V] However, the method of proof, with which we derived Theorem 28, fails in the general case.

9. THE LENGTH OF THE BOUNDARY GENERATING CURVE AND THE AREA OF THE NUMERICAL RANGE

Let

$$(11) \quad f_A(u, v, w) \equiv |H_1u + H_2v + I_n w| \equiv w^n + C_1(u, v)w^{n-1} + \dots + C_n(u, v) = 0$$

be the equation of the boundary generating curve of the numerical range $W(A)$ of the matrix A . For the following considerations we first assume that the trace $\sigma(A)$ of the matrix A is equal to zero: $C_1(u, v) = 0$. This restriction is of no consequence in this context, since it can always be accomplished through a parallel translation. Moreover, let the common factor of the homogeneous coordinates u, v, w be selected in such a way that

$$u^2 + v^2 = 1; \quad u = \cos \varphi, \quad v = \sin \varphi.$$

Now let $W_0 = W_0(\varphi)$ and $w_0 = w_0(\varphi)$ be the largest and respectively smallest (real) roots of equation (11). Then

$$d_0 = W_0 - w_0$$

represents the distance between the two support lines of the convex region $W(A)$ parallel to the φ -direction. The function $d_0 = d_0(\varphi)$ can now be estimated with the aid of a theorem from I. SCHUR¹⁷:

If the polynomial $f(w)$ of degree n has only real zeros, of which W_0 is the largest, and if W_1, \dots, W_{n-1} are the largest zeros of the derivatives $f^{(1)}, \dots, f^{(n-1)}$ of f , then we have

$$W_0 - W_1 \leq W_1 - W_2 \leq \dots \leq W_{n-2} - W_{n-1}.$$

¹⁵SPECHT [7]

¹⁶WEYL [10]

¹⁷I. SCHUR [6]

Correspondingly there are inequalities for the smallest zeros w_0, w_1, \dots, w_{n-1} ($W_{n-1} = w_{n-1}$) of f and its derivatives $f^{(1)}, \dots, f^{(n-1)}$

$$-w_0 + w_1 \leq -w_1 + w_2 \leq \dots \leq -w_{n-2} + w_{n-1},$$

from which the inequalities for the differences $d_i = W_i - w_i$ follow:

$$d_0 - d_1 \leq d_1 - d_2 \leq \dots \leq d_{n-3} - d_{n-2} \leq d_{n-2} \quad (d_{n-1} = 0).$$

But from this is produced – as one can prove by induction –

$$d_0 \leq (\chi + 1)d_\chi - (\chi)d_{\chi+1} \quad (\chi = 0, 1, \dots, n-1).$$

For $\chi = n-2$, from the fact that $d_{n-1} = 0$, this yields the inequality

$$d_0 \leq (n-1)d_{n-2}.$$

If we apply this consequence of the theorem of Schur to the polynomial $f(w) = f_A(u, v, w)$ of (11), then we have

$$f^{(n-2)}(w) = \frac{n!}{2}w^2 + (n-2)!C_2 \quad (C_2 = C_2(u, v)),$$

therefore

$$d_{n-2} = 2\sqrt{\frac{-2C_2}{n(n-1)}}.$$

Thereby we obtain

$$(12) \quad d_0 \leq 2(n-1)\sqrt{\frac{-2C_2}{n(n-1)}}.$$

Now C_2 is equal to the sum of the determinants of the principal submatrices of dimension 2 of the Hermitian matrix $H = -(H_1u + H_2v)$. If

$$h_1(\varphi) \leq h_2(\varphi) \leq \dots \leq h_n(\varphi)$$

are the eigenvalues of $H = -(H_1u + H_2v)$, then it follows

$$C_2 = \sum_{\lambda < \mu=2}^n h_\lambda h_\mu,$$

therefore

$$C_2 = \frac{[\sigma(H)]^2 - \sigma(H^2)}{2}$$

– a relation, that holds in general because of the unitary invariance of the quantities in it, even if H is not given in diagonal form. Since it was assumed that $C_1(u, v) = 0$, we have

$$C_2 = -\frac{\sigma(H^2)}{2};$$

thereby

$$\sigma(H^2) = \sigma(H_1^2)u^2 + 2\sigma(H_1H_2)uv + \sigma(H_2^2)v^2,$$

from which our estimate becomes

$$d_0 \leq 2\sqrt{n-1}\sqrt{\frac{\sigma(H^2)}{n}}.$$

It is also possible to bound the distance d_0 from below. Thereto we notice that we have the equation

$$\begin{aligned} \sum_{\chi < \lambda = 2} (h_\chi - h_\lambda)^2 &= \frac{1}{2} \sum_{\chi \neq \lambda} (h_\chi - h_\lambda)^2 = \frac{1}{2} \left\{ \sum_{\chi} h_\chi^2 - 2 \sum_{\chi, \lambda} h_\chi h_\lambda + \sum_{\lambda} h_\lambda^2 \right\} \\ &= \frac{1}{2} \{2n\sigma(H^2) - 2[\sigma(H)]^2\}, \end{aligned}$$

so

$$(13) \quad \sum_{\chi < \lambda = 2}^n (h_\chi - h_\lambda)^2 = n\sigma(H^2).$$

Occurring in this expression are the differences:

$$\begin{array}{ccccccc} h_2 - h_1, & & & & & & \\ h_3 - h_1, & h_3 - h_2, & & & & & \\ h_4 - h_1, & h_4 - h_2, & h_4 - h_3, & & & & \\ \dots & \dots & \dots & \dots & & & \\ h_n - h_1, & h_n - h_2, & h_n - h_3, & \dots, & h_n - h_{n-1}. & & \end{array}$$

For the elements on the main diagonal this gives

$$(h_2 - h_1) + (h_3 - h_2) + (h_4 - h_3) + \dots + (h_n - h_{n-1}) = d_0,$$

therefore, when we notice that the expressions in parentheses are nonnegative,

$$(h_2 - h_1)^2 + (h_3 - h_2)^2 + \dots + (h_n - h_{n-1})^2 \leq d_0^2.$$

Out of the differences in the first diagonal parallel to the main diagonal we form

$$\begin{aligned} (h_3 - h_1) + (h_5 - h_3) + (h_7 - h_5) + \dots &\leq d_0, \\ (h_4 - h_2) + (h_6 - h_4) + (h_8 - h_6) + \dots &\leq d_0 \end{aligned}$$

and thereby we obtain

$$(h_3 - h_1)^2 + (h_4 - h_2)^2 + \dots + (h_n - h_{n-2})^2 \leq 2d_0^2.$$

Continuing in this manner we finally reach the second-last diagonal:

$$(h_{n-1} - h_1) \leq d_0, \quad (h_n - h_2) \leq d_0,$$

and this yields

$$(h_{n-1} - h_1)^2 + (h_n - h_2)^2 \leq 2d_0^2,$$

whereas the last, a diagonal consisting of only a single element, yields

$$(h_n - h_1)^2 = d_0^2.$$

Thereby all of the sums appearing in (13) are estimated and hence it now becomes

$$n\sigma(H^2) = \sum_{\chi < \lambda = 2}^n (h_\chi - h_\lambda)^2 \leq (1 + 2 + 3 + \dots + 3 + 2 + 1)d_0^2 = \frac{s(n)}{4}d_0^2$$

with

$$s(n) = \begin{cases} n^2 - 1 & (n \text{ odd}), \\ n^2 & (n \text{ even}), \end{cases}$$

therefore

$$d_0^2 \geq 4n \frac{\sigma(H^2)}{s(n)} \quad (n > 1).$$

Therefore, we have for d_0 the bounds

$$2\frac{\sqrt{n}}{\sqrt{s(n)}}\sqrt{\sigma(H^2)} \leq d_0 \leq 2\sqrt{\frac{n-1}{n}}\sqrt{\sigma(H^2)},$$

or, when we replace $s(n)$ by n^2 ,

Theorem 29.

$$2\sqrt{\frac{\sigma(H^2)}{n}} \leq d_0 \leq 2\sqrt{n-1}\sqrt{\frac{\sigma(H^2)}{n}} \quad (n > 1).$$

For the diameter D and the width Δ of the region $W(A)$ we then have¹⁸

$$\begin{aligned} \frac{2}{\sqrt{n}}\sqrt{\max(\sigma(H^2))} &\leq D \leq 2\sqrt{\frac{n-1}{n}}\sqrt{\max(\sigma(H^2))}, \\ \frac{2}{\sqrt{n}}\sqrt{\min(\sigma(H^2))} &\leq \Delta \leq 2\sqrt{\frac{n-1}{n}}\sqrt{\min(\sigma(H^2))}. \end{aligned}$$

Now

$$\sigma(H^2) = \sigma(H_1^2)u^2 + 2\sigma(H_1H_2)uv + \sigma(H_2^2)v^2;$$

the extreme values of this form under the auxiliary condition $u^2 + v^2 = 1$ are the eigenvalues $0 \leq e_1 \leq e_2$ of the matrix

$$M = \begin{pmatrix} \sigma(H_1^2) & \sigma(H_1H_2) \\ \sigma(H_2H_1) & \sigma(H_2^2) \end{pmatrix}.$$

Thereby we have

$$\begin{aligned} \frac{2}{\sqrt{n}}\sqrt{e_2} &\leq D \leq 2\sqrt{\frac{n-1}{n}}\sqrt{e_2}, \\ \frac{2}{\sqrt{n}}\sqrt{e_1} &\leq \Delta \leq 2\sqrt{\frac{n-1}{n}}\sqrt{e_1}. \end{aligned}$$

From this it follows for the area F of $W(A)$

$$2F \geq \Delta D \geq \frac{4}{n}\sqrt{\det(M)}.$$

Since additionally

$$F \leq \Delta D$$

for a convex region, this yields

$$F \leq 4\frac{n-1}{n}\sqrt{e_1e_2} = 4\frac{n-1}{n}\sqrt{\det(M)}.$$

Consequently we have the estimate for F

$$\frac{2}{n}\sqrt{\sigma(H_1^2)\sigma(H_2^2) - [\sigma(H_1H_2)]^2} \leq F \leq 4\frac{n-1}{n}\sqrt{\sigma(H_1^2)\sigma(H_2^2) - [\sigma(H_1H_2)]^2}.$$

If we also observe that

$$\begin{aligned} &\sigma(A^2)\sigma(A^{*2}) - [\sigma(AA^*)]^2 \\ &= [\sigma(H_1^2) - \sigma(H_2^2) + 2i\sigma(H_1H_2)][\sigma(H_1^2) - \sigma(H_2^2) - 2i\sigma(H_1H_2)] - [\sigma(H_1^2) + \sigma(H_2^2)]^2 \\ &= [\sigma(H_1^2) - \sigma(H_2^2)]^2 + 4[\sigma(H_1H_2)]^2 - [\sigma(H_1^2) + \sigma(H_2^2)]^2 \\ &= -4\{\sigma(H_1^2)\sigma(H_2^2) - [\sigma(H_1H_2)]^2\}, \end{aligned}$$

¹⁸For the concepts and formulas from the theory of convex regions see BONNESEN-FENCHEL [1].

then the inequalities become:

Theorem 30.

$$\frac{1}{n} \sqrt{[\sigma(AA^*)]^2 - \sigma(A^2)\sigma(A^{*2})} \leq F \leq 2 \frac{n-1}{n} \sqrt{[\sigma(AA^*)]^2 - \sigma(A^2)\sigma(A^{*2})}.$$

The length

$$L = \frac{1}{2} \int_0^{2\pi} d_0 d\varphi$$

of the boundary generating curve of $W(A)$ may be estimated because of Theorem 29 by means of

$$\frac{1}{\sqrt{n}} J \leq L \leq \sqrt{\frac{n-1}{n}} J$$

with

$$J = \int_0^{2\pi} \sqrt{\sigma(H^2)} d\varphi = \int_0^{2\pi} \sqrt{\sigma(H_1^2) \cos^2 \varphi + 2\sigma(H_1 H_2) \cos \varphi \sin \varphi + \sigma(H_2^2) \sin^2 \varphi} d\varphi.$$

The integral may be brought through an orthogonal transformation of the unit vector $(\cos \varphi, \sin \varphi)$ to the form

$$J = \int_0^{2\pi} \sqrt{e_1 \cos^2 \varphi + e_2 \sin^2 \varphi} d\varphi.$$

However, this represents the length of an ellipse with the semiaxes $\sqrt{e_1}, \sqrt{e_2}$. Then this yields for the length of the boundary of this ellipse

$$J \geq 4\sqrt{e_1 + e_2} = 4\sqrt{\sigma(H_1^2) + \sigma(H_2^2)} = 4\sqrt{\sigma(AA^*)}.$$

On the other hand, we have

$$J \leq \int_0^{2\pi} \sqrt{e_1 + e_2} d\varphi = 2\pi \sqrt{\sigma(AA^*)}.$$

It follows that

Theorem 31.

$$4\sqrt{\frac{\sigma(AA^*)}{n}} \leq L \leq 2\pi\sqrt{n-1} \sqrt{\frac{\sigma(AA^*)}{n}}.$$

Part 2. Quaternion Matrices

10. QUATERNION MATRICES

The theory of matrices with complex elements may be carried over to quaternion matrices almost word for word. We now summarize the most important facts about them¹⁹:

A quaternion

$$\mathbf{a} = \alpha_1 + \alpha_2 i + \beta_1 \mathbf{j} + \beta_2 i\mathbf{j} \quad (\alpha_1, \alpha_2, \beta_1, \beta_2 \text{ real})$$

may be written in the form

$$\mathbf{a} = a + b\mathbf{j}$$

where $a = \alpha_1 + i\alpha_2$, $b = \beta_1 + i\beta_2$ are complex numbers, while \mathbf{j} is a quaternion which satisfies the equation

$$\mathbf{j}^2 = -1$$

¹⁹LEE [5].

and for which

$$ij = -ji$$

when i is the imaginary unit of the complex numbers. Since \mathbf{j} commutes with real numbers it follows that so does any quaternion.

α_1 is said to be the *real part* $\Re(\mathbf{a})$ of the quaternion \mathbf{a} .

The *conjugate quaternion* \mathbf{a}^* of \mathbf{a} is defined as

$$\mathbf{a}^* = \bar{a} - b\mathbf{j}.$$

Then $\mathbf{a}\mathbf{a}^* = \mathbf{a}^*\mathbf{a}$ is a positive number:

$$\mathbf{a}\mathbf{a}^* = \mathbf{a}^*\mathbf{a} = \alpha_1^2 + \alpha_2^2 + \beta_1^2 + \beta_2^2.$$

Its positive square root is called the *absolute value* $|\mathbf{a}|$ of the quaternion \mathbf{a} . A quaternion \mathbf{u} of absolute value 1 is called *unitary*. Then we have $\mathbf{u}^{-1} = \mathbf{u}^*$.

The reciprocal \mathbf{a}^{-1} and the conjugate \mathbf{a}^* of a quaternion \mathbf{a} satisfy the relation

$$\mathbf{a}^* = \mathbf{a}^{-1}\mathbf{a}\mathbf{a}^* = \mathbf{a}^{-1}|\mathbf{a}|^2.$$

Two quaternions \mathbf{a}, \mathbf{b} are called *similar* when a quaternion \mathbf{x} exists such that

$$\mathbf{x}^{-1}\mathbf{a}\mathbf{x} = \mathbf{b}.$$

If two quaternions are similar, then they are unitarily similar, i.e.,

$$\mathbf{u}^*\mathbf{a}\mathbf{u} = \mathbf{b}, \quad \mathbf{u}^*\mathbf{u} = 1$$

holds. Namely if $\mathbf{x}^{-1}\mathbf{a}\mathbf{x} = \mathbf{b}$, then we set $\mathbf{u} = \frac{\mathbf{x}}{|\mathbf{x}|}$. Then $\mathbf{u}^*\mathbf{a}\mathbf{u} = \mathbf{x}^{-1}\mathbf{a}\mathbf{x} = \mathbf{b}$.

In particular every quaternion $\mathbf{a} = a + b\mathbf{j}$ is similar to the quaternion $\bar{\mathbf{a}} = \bar{a} + \bar{b}\mathbf{j}$, for we have

$$\mathbf{j}^{-1}\mathbf{a}\mathbf{j} = \mathbf{j}^{-1}(a + b\mathbf{j})\mathbf{j} = -\mathbf{j}(a\mathbf{j} - b) = \bar{a} + \bar{b}\mathbf{j} = \bar{\mathbf{a}}.$$

If we associate to every quaternion $\mathbf{a} = a + b\mathbf{j}$ the complex *image matrix*

$$(14) \quad M(\mathbf{a}) = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix},$$

then we get a true representation of the algebra of quaternions. The matrices A of the form (14), can be characterized by the fact that for

$$M(\mathbf{j}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad M(\mathbf{j}^{-1}) = M(-\mathbf{j}) = -M(\mathbf{j})$$

the equation

$$(15) \quad M(\mathbf{j}^{-1})AM(\mathbf{j}) = \bar{A}$$

holds. Each matrix A of dimension 2 which satisfies the equation (15) represents a quaternion.

Sums and products of two quaternions correspond to sums and products of image matrices of the form (14), and vice versa.

The real numbers correspond to real (Hermitian) diagonal matrices, the complex numbers to (normal) diagonal matrices. The image matrices of unitary quaternions are unitary. Furthermore,

$$M(\bar{\mathbf{a}}) = \overline{M(\mathbf{a})}, \quad M(\mathbf{a}^*) = (M(\mathbf{a}))^*$$

With the help of image matrices one can directly prove the following theorem:

Theorem 32. *Two quaternions \mathbf{a}, \mathbf{b} are similar if and only if their real parts and their absolute values agree.*

Proof. The eigenvalues of the image matrix of a quaternion are equal to

$$\Re(\mathbf{a}) + \sqrt{\Re(\mathbf{a})^2 - |\mathbf{a}|^2}, \quad \Re(\mathbf{a}) - \sqrt{\Re(\mathbf{a})^2 - |\mathbf{a}|^2},$$

therefore they are equal for two image matrices precisely when both the real parts and the absolute values of the quaternions agree. Now in the case of non-real quaternions the elementary divisors of an image matrix are always linear, since for non-real quaternions the eigenvalues are distinct. Consequently the elementary divisors of two image matrices are equal precisely when both the real parts and the absolute values of their quaternions agree. Thus this happens exactly when the two image matrices are similar. Hence there exists a matrix P such that

$$P^{-1}M(\mathbf{a})P = M(\mathbf{b}).$$

But then we also have

$$Q^{-1}M(\mathbf{a})Q = M(\mathbf{b}) \quad \text{with} \quad Q = M(\mathbf{j}^{-1})\bar{P}M(\mathbf{j})$$

and from this in general

$$[cP + \bar{c}Q]^{-1}M(\mathbf{a})[cP + \bar{c}Q] = M(\mathbf{b}),$$

where the complex number c is chosen such that $cP + \bar{c}Q$ is regular. Since P and Q are regular, this is always possible. But now

$$M(\mathbf{j}^{-1})[cP + \bar{c}Q]^{-1}M(\mathbf{j}) = c\bar{Q} + \bar{c}P = [cP + \bar{c}Q],$$

so that $cP + \bar{c}Q = M(\mathbf{x})$ is the image matrix of a quaternion \mathbf{x} . Consequently

$$\mathbf{x}^{-1}\mathbf{a}\mathbf{x} = \mathbf{b}.$$

□

Now let $\mathbf{A} = (\mathbf{a}_{\mu\nu})$ ($\mu = 1, \dots, m; \nu = 1, \dots, n$) be an $m \times n$ matrix of quaternions. Sums and products are defined in the usual way as in the case of complex matrices and

$$\mathbf{A}^* = (\mathbf{a}_{\nu\mu}^*).$$

Then we have the well known formulas

$$(\mathbf{A} \pm \mathbf{B})^* = \mathbf{A}^* \pm \mathbf{B}^*, \quad (\mathbf{AB})^* = \mathbf{B}^* \mathbf{A}^*.$$

We call a square quaternion matrix \mathbf{A} of dimension n ($m = n$) *Hermitian* when $\mathbf{A} = \mathbf{A}^*$, *normal* when $\mathbf{AA}^* = \mathbf{A}^*\mathbf{A}$, and *unitary* when $\mathbf{AA}^* = \mathbf{A}^*\mathbf{A} = I_n$, where I_n denotes the unit matrix ($\delta_{\mu\nu}$) ($\delta_{\mu\nu}$ is the Kronecker symbol). A quaternion matrix is called *regular* when a matrix \mathbf{A}^{-1} exists, such that $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = I_n$. In particular every unitary matrix is regular.

Just as we are able to split every matrix with complex components into Hermitian parts, we are also able to decompose every quaternion matrix into a complex and a quaternion part:

$$\mathbf{A} = \mathbf{H}_1 + \mathbf{H}_2\mathbf{j}; \quad \mathbf{H}_1 = \frac{1}{2}(\mathbf{A} + \mathbf{A}^*), \quad \mathbf{H}_2 = \frac{1}{2}(\mathbf{A} - \mathbf{A}^*)\mathbf{j}^{-1}$$

Here $\mathbf{H}_1, \mathbf{H}_2$ are Hermitian matrices^[V]. The decomposition is unique.

We now pair each quaternion matrix $\mathbf{A} = (\mathbf{a}_{\mu\nu})$ ($\mu = 1, \dots, m; \nu = 1, \dots, n$) with a complex image matrix $M(\mathbf{A})$ of $2m$ rows and $2n$ columns through the relation

$$M(\mathbf{A}) = (M(\mathbf{a}_{\mu\nu})).$$

Since the $M(\mathbf{a}_{\mu\nu})$ form a representation of the quaternions, we have the rules

$$M(\mathbf{A} \pm \mathbf{B}) = M(\mathbf{A}) \pm M(\mathbf{B}), \quad M(\mathbf{AB}) = M(\mathbf{A})M(\mathbf{B}), \quad M(\mathbf{A}^*) = (M(\mathbf{A}))^*,$$

and from this it follows directly that Hermitian, normal, and unitary quaternion matrices have Hermitian, normal, and unitary image matrices, respectively.

A complex matrix A is an image matrix of a quaternion matrix of m rows and n columns exactly when the equation

$$[I_m \otimes M(\mathbf{j}^{-1})] A [I_n \otimes M(\mathbf{j})] = \bar{A}$$

is satisfied, in which I_m, I_n are the identity matrices of m respectively n .

The following theorems for quaternion matrices hold:

Theorem 33. *If \mathbf{A} is a quaternion matrix of dimension n , then there exists a unitary quaternion matrix \mathbf{U} such that $\mathbf{U}^* \mathbf{A} \mathbf{U}$ has triangular form:*

$$\mathbf{U}^* \mathbf{A} \mathbf{U} = \begin{pmatrix} a_1 & 0 & 0 & \dots & 0 \\ \mathbf{b}_{21} & a_2 & 0 & \dots & 0 \\ \mathbf{b}_{31} & \mathbf{b}_{32} & a_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ \mathbf{b}_{n1} & \dots & \dots & \dots & a_n \end{pmatrix}.$$

Thereby the diagonal elements are determined up to similarity. In particular we can arrange so that they are complex.

The similarity classes fixed by the diagonal elements can be called *eigenvalue classes of the matrix \mathbf{A}* . From Theorem 33 it follows:

Theorem 34. *If the matrix \mathbf{A} is normal, then it is possible bring it into diagonal form through a unitary transformation. Thereby it is possible to arrange that the diagonal elements are complex. If \mathbf{A} is Hermitian, then it is possible to bring it into real diagonal form.*

11. THE NUMERICAL RANGE OF A QUATERNION MATRIX

Let $\mathbf{A} = (\mathbf{a}_{\mu\nu})$ be a square quaternion matrix of dimension n . The numerical range $W(\mathbf{A})$ of \mathbf{A} may be defined as the set of all quaternions which can be assumed by the form

$$\Phi(\mathbf{A}, \mathbf{X}) = \mathbf{X}^* \mathbf{A} \mathbf{X}, \quad \mathbf{X}^* \mathbf{X} = 1.$$

Here \mathbf{X} is an n -dimensional vector whose elements are quaternions. If we interpret each quaternion as a point in a four-dimensional space R_4 with cartesian coordinates, then the numerical range of a quaternion matrix represents a set of points in R_4 . The numerical range of a quaternion matrix includes with each quaternion also all of its similar quaternions; for when the quaternion \mathbf{a} can be represented in the form

$$\mathbf{X}^* \mathbf{A} \mathbf{X} = \mathbf{a},$$

then it is possible to represent all of its similar quaternions in the form

$$\mathbf{u}^{-1} \mathbf{a} \mathbf{u} = \mathbf{u}^* \mathbf{X}^* \mathbf{A} \mathbf{X} \mathbf{u} = \mathbf{Y}^* \mathbf{A} \mathbf{Y}, \quad \mathbf{Y}^* \mathbf{Y} = 1 \quad (\mathbf{u}^* = \mathbf{u}^{-1}, \mathbf{Y} = \mathbf{X} \mathbf{u}),$$

thus they also belong to the numerical range $W(\mathbf{A})$. Since all similar quaternions have equal real parts and equal absolute values, all of their corresponding points lie in the hyperplane determined by the common real part, in which they form a (two-dimensional) sphere determined by their common absolute value. The center of each such sphere lies on the real axis of the quaternion space. The numerical

range of a quaternion matrix is therefore just composed of spheres whose centers lie on the real axis. But this means that the numerical range lies rotationally symmetric along the real axis in R_4 . Because of the position of the center of these special solids, the quaternion numerical ranges do not admit general affine transformations, in particular the property of the numerical range of a quaternion matrix being invariant under any rotation may be lost.

However, as in the complex case, the numerical range of a quaternion matrix is invariant under a unitary transformation:

$$W(\mathbf{A}) = W(\mathbf{U}^* \mathbf{A} \mathbf{U}), \quad \mathbf{U}^* \mathbf{U} = 1.$$

For since \mathbf{U} is a regular matrix, $\mathbf{U}\mathbf{X}$ runs with \mathbf{X} through all vectors of norm one with n components. Now we slice the numerical range $W(\mathbf{A})$ with the complex plane. If \mathbf{a} is a non-real quaternion in $W(\mathbf{A})$:

$$\mathbf{a} = \alpha + \beta i + \gamma \mathbf{j} + \delta i\mathbf{j}, \quad (\beta^2 + \gamma^2 + \delta^2 \neq 0),$$

then exactly two complex numbers lie in the similarity class of \mathbf{a} ,

$$\alpha + i\sqrt{\beta^2 + \gamma^2 + \delta^2}, \quad \alpha - i\sqrt{\beta^2 + \gamma^2 + \delta^2},$$

the eigenvalues of $M(\mathbf{a})$; the sphere on which all of the quaternions similar to \mathbf{a} lie, therefore, intersects the complex plane in two complex conjugate points. This pair of points is called the *bild of the class of \mathbf{a}* or the *bild of \mathbf{a}* . Hence each class of $W(\mathbf{A})$ is thus represented by a pair of complex conjugate points at the intersection of the region $W(\mathbf{A})$ with the Gaussian plane. Therefore all classes of $W(\mathbf{A})$ are mapped to a set of points in the Gaussian plane lying symmetric to the real axis. This set of points is called the *bild $B(\mathbf{A})$ of $W(\mathbf{A})$* .^[VI] Since the pairs of complex conjugate points of $B(\mathbf{A})$ and the classes of $W(\mathbf{A})$ determine each other, since $W(\mathbf{A})$ contains only entire classes of quaternions, and since the set of real numbers of $B(\mathbf{A})$ is the same as the set of real numbers of $W(\mathbf{A})$, $B(\mathbf{A})$ and $W(\mathbf{A})$ determine each other uniquely. Therefore to determine the properties of $W(\mathbf{A})$ it suffices to investigate the properties of $B(\mathbf{A})$.

Theorem 35. *If \mathbf{A} is a square quaternion matrix and $B(\mathbf{A})$ is the bild of the numerical range $W(\mathbf{A})$, then $B(\mathbf{A})$ is the (complex) numerical range of the complex matrix $M(\mathbf{A})$, thus*

$$B(\mathbf{A}) = W(M(\mathbf{A})).^[VII]$$

Proof. If $\mathbf{a} \in W(\mathbf{A})$, then

$$\mathbf{a} = \mathbf{X}_0 \mathbf{A} \mathbf{X}_0 \quad \text{with} \quad \mathbf{X}_0 = \begin{pmatrix} \mathbf{x}_{01} \\ \vdots \\ \mathbf{x}_{0n} \end{pmatrix},$$

then we also have

$$M(\mathbf{a}) = M(\mathbf{X}_0)^* M(\mathbf{A}) M(\mathbf{X}_0) = [M(\mathbf{x}_{01})^*, \dots, M(\mathbf{x}_{0n})^*] [M(\mathbf{a}_{\mu\nu})] \begin{pmatrix} M(\mathbf{x}_{01}) \\ \vdots \\ M(\mathbf{x}_{0n}) \end{pmatrix},$$

therefore

$$M(\mathbf{a}) = \sum_{\mu, \nu} M(\mathbf{x}_{0\mu})^* M(\mathbf{a}_{\mu\nu}) M(\mathbf{x}_{0\nu}).$$

The eigenvalues of $M(\mathbf{a})$:

$$\left. \begin{array}{l} c \\ \bar{c} \end{array} \right\} = \frac{a + \bar{a}}{2} \pm \sqrt{\left(\frac{a + \bar{a}}{2}\right)^2 - a\bar{a} - b\bar{b}},$$

however, are exactly the points of intersection of the sphere belonging to the class of \mathbf{a} and the Gaussian plane.

Since the quaternion \mathbf{a} is similar to the complex number c , there exists a unitary quaternion \mathbf{u} such that

$$\mathbf{u}^* \mathbf{a} \mathbf{u} = c$$

Now let

$$\mathbf{U} = M(\mathbf{u}).$$

Then this becomes

$$\begin{aligned} M(\mathbf{x}_{0\nu})\mathbf{U} &= M(\mathbf{x}_{0\nu}\mathbf{u}), & \mathbf{U}^* M(\mathbf{a})\mathbf{U} &= \begin{pmatrix} c & 0 \\ 0 & \bar{c} \end{pmatrix}, \\ \mathbf{U}^* M(\mathbf{a})\mathbf{U} &= \sum_{\mu,\nu} (M(\mathbf{x}_{0\mu}\mathbf{u}))^* M(\mathbf{a}_{\mu\nu}) M(\mathbf{x}_{0\nu}\mathbf{u}). \end{aligned}$$

If we decompose the matrix

$$M(\mathbf{X}_0\mathbf{u}) = \begin{pmatrix} M(\mathbf{x}_{01}\mathbf{u}) \\ \vdots \\ M(\mathbf{x}_{0n}\mathbf{u}) \end{pmatrix}$$

into two columns and if we call the resulting vectors x_0 and y_0 , each of norm one and with $2n$ components, then we get

$$x_0^* M(\mathbf{A}) x_0 = c, \quad y_0^* M(\mathbf{A}) y_0 = \bar{c}.$$

However, from this it follows that $W(\mathbf{A})$ contains both of the complex representations of each class of $W(\mathbf{A})$, so that $B(\mathbf{A}) \subseteq W(M(\mathbf{A}))$. Now for the converse let

$$x_0^* M(\mathbf{A}) x_0 = c, \quad x_0^* x_0 = 1.$$

The matrix $M(\mathbf{A})$ has the property that²⁰

$$M(\mathbf{j}^{-1})M(\mathbf{A})M(\mathbf{j}) = \overline{M(\mathbf{A})}.$$

With

$$y_0 = M(\mathbf{j})\bar{x}_0 \quad (y_0^* y_0 = 1, \quad x_0^* y_0 = y_0^* x_0 = 0)$$

hence this becomes

$$y_0^* M(\mathbf{A}) y_0 = \bar{x}_0^* M(\mathbf{j}^{-1})M(\mathbf{A})M(\mathbf{j})\bar{x}_0 = \bar{x}_0^* \overline{M(\mathbf{A})} \bar{x}_0 = \bar{c}.$$

Therefore for each complex number c contained in the region $W(M(\mathbf{A}))$, the complex conjugate number \bar{c} is also contained, so just like $B(\mathbf{A})$ it lies symmetric with respect to the real axis (see the discussion after Theorem 7). Now if we concatenate x_0 and y_0 into a matrix

$$Z_0 = (x_0, y_0) \quad \text{with} \quad Z_0^* M(\mathbf{A}) Z_0 = \begin{pmatrix} c & k \\ -\bar{k} & \bar{c} \end{pmatrix},$$

then through the equation

$$Z_0 = M(\mathbf{X}_0)$$

²⁰In what follows instead of $M(\mathbf{j}I_n)$ and $M(\mathbf{j}^{-1}I_n)$ we simply write $M(\mathbf{j})$ and $M(\mathbf{j}^{-1})$.

this defines a quaternion eigenvector \mathbf{X}_0 of n components and this becomes

$$\mathbf{X}_0^* \mathbf{A} \mathbf{X}_0 = c + k\mathbf{j}, \quad \mathbf{X}_0^* \mathbf{X}_0 = 1.$$

Now from this it follows: if the complex number c lies in $W(M(\mathbf{A}))$, then there exists a quaternion $c + k\mathbf{j}$ in $W(\mathbf{A})$, whose real part is equal to the real part of c , and whose absolute value is $\geq |c|$. The bild of $c + k\mathbf{j}$ is therefore a complex number, whose real part agrees with the real part of c , but it has an absolute value $\geq |c|$.

To prove from this that

$$W(M(\mathbf{A})) = B(\mathbf{A}),$$

we prove the following:

Lemma. *If the complex numbers a, \bar{a} lie in $B(\mathbf{A})$, then all numbers on the line segment connecting a, \bar{a} lie in $B(\mathbf{A})$.*^[IX]

Proof of Lemma. Let $\mathbf{X}_0^* \mathbf{A} \mathbf{X}_0 = a = \alpha_1 + i\alpha_2$. Then $M(\mathbf{X}_0) = (x_0, y_0)$, where $y_0 = M(\mathbf{j})\bar{x}_0$. For the complex vectors x_0, y_0 , we have

$$x_0^* M(\mathbf{A}) x_0 = a, \quad y_0^* M(\mathbf{A}) y_0 = \bar{a}, \quad x_0^* M(\mathbf{A}) y_0 = 0, \quad y_0^* M(\mathbf{A}) x_0 = 0.$$

If we now form

$$r_0 = \sqrt{\vartheta} x_0 + \sqrt{1 - \vartheta} y_0 \quad (0 \leq \vartheta \leq 1)$$

and

$$s_0 = M(\mathbf{j})\bar{r}_0 = \sqrt{\vartheta} M(\mathbf{j})\bar{x}_0 + \sqrt{1 - \vartheta} M(\mathbf{j})\bar{y}_0 = \sqrt{\vartheta} y_0 - \sqrt{1 - \vartheta} x_0,$$

then r_0 and s_0 define a quaternion vector \mathbf{Z}_0 of norm one through

$$(r_0, s_0) = M(\mathbf{Z}_0)$$

for which we have:

$$\begin{aligned} M(\mathbf{Z}_0^* \mathbf{A} \mathbf{Z}_0) &= M(\mathbf{Z}_0^*) M(\mathbf{A}) M(\mathbf{Z}_0) = \begin{pmatrix} r_0^* \\ s_0^* \end{pmatrix} M(\mathbf{A}) (r_0, s_0) \\ &= \begin{pmatrix} r_0^* M(\mathbf{A}) r_0 & r_0^* M(\mathbf{A}) s_0 \\ s_0^* M(\mathbf{A}) r_0 & s_0^* M(\mathbf{A}) s_0 \end{pmatrix}. \end{aligned}$$

Now we have

$$\begin{aligned} r_0^* M(\mathbf{A}) r_0 &= \vartheta a + (1 - \vartheta) \bar{a}, \\ s_0^* M(\mathbf{A}) s_0 &= \vartheta \bar{a} + (1 - \vartheta) a, \\ r_0^* M(\mathbf{A}) s_0 &= -2\sqrt{\vartheta} \sqrt{1 - \vartheta} \alpha_2 i, \\ s_0^* M(\mathbf{A}) r_0 &= -2\sqrt{\vartheta} \sqrt{1 - \vartheta} \alpha_2 i, \end{aligned}$$

and thereby

$$M(\mathbf{Z}_0^* \mathbf{A} \mathbf{Z}_0) = \begin{pmatrix} \vartheta a + (1 - \vartheta) \bar{a} & -2\sqrt{\vartheta} \sqrt{1 - \vartheta} \alpha_2 i \\ -2\sqrt{\vartheta} \sqrt{1 - \vartheta} \alpha_2 i & \vartheta \bar{a} + (1 - \vartheta) a \end{pmatrix}.$$

Now if we let ϑ run along the interval $0 \leq \vartheta \leq 1$, then the bild of the quaternions

$$\mathbf{Z}_0^* \mathbf{A} \mathbf{Z}_0 = \vartheta a + (1 - \vartheta) \bar{a} - 2\sqrt{\vartheta} \sqrt{1 - \vartheta} \alpha_2 i \mathbf{j}$$

describes a piece of the line parallel to the imaginary axis, determined by the real part α_1 of a , on which the points a, \bar{a} also lie.^[X] For $\vartheta = 0$ the bild coincides with \bar{a} , for $\vartheta = 1$ the bild coincides with a .^[XI] For reasons of continuity the bild must

also assume every intermediate value. \square

From all of the preceding arguments follows Theorem 35: in any case

$$B(\mathbf{A}) \subseteq W(M(\mathbf{A})).$$

If now $w \in W(M(\mathbf{A}))$, then there exists a complex number $b \in B(\mathbf{A})$ which has the same real part that w has, but also with imaginary part at least as large in absolute value as that of the imaginary part of w . Applying the lemma to b , each point of the line segment connecting b and \bar{b} lies in $B(\mathbf{A})$. But since w is a point on this line segment, this gives us $w \in B(\mathbf{A})$ and therefore $B(\mathbf{A}) = W(M(\mathbf{A}))$. \square

Since all properties of the numerical range $W(\mathbf{A})$ of a quaternion matrix \mathbf{A} are mirrored in its bild, the theorem just proven reduces the theory of the numerical ranges of quaternion matrices of dimension n to the theory of the numerical ranges of complex matrices of dimension $2n$. In particular now we can directly prove:

Theorem 36. *The numerical range of a quaternion matrix is represented as a closed convex set of points in the space of quaternions.^[XII]*

Proof. That the set of points is closed is trivial. We prove the convexity and show that with any two of its points $W(U)$ always contains the line segment between them. Let

$$\mathbf{a} = \alpha_1 + \alpha_2i + \alpha_3\mathbf{j} + \alpha_4i\mathbf{j}, \quad \mathbf{b} = \beta_1 + \beta_2i + \beta_3\mathbf{j} + \beta_4i\mathbf{j}$$

be two quaternions in $W(\mathbf{A})$. The points on the connection line between them are then described by the quaternions

$$(1 - \lambda)\mathbf{a} + \lambda\mathbf{b} \quad (0 \leq \lambda \leq 1).$$

The classes of \mathbf{a} and \mathbf{b} may be represented by the complex numbers a, \bar{a} and b, \bar{b} . We examine the complex representations

$$(1 - \lambda)a + \lambda b$$

of the points a and b . The real part of the bild of these points is $(1 - \lambda)\alpha_1 + \lambda\beta_1$. The imaginary part η of both points certainly has absolute value $\leq |\eta_0|$ with

$$|\eta_0| = (1 - \lambda)\sqrt{\alpha_2^2 + \alpha_3^2 + \alpha_4^2} + \lambda\sqrt{\beta_2^2 + \beta_3^2 + \beta_4^2}.$$

Then this yields

$$\begin{aligned} \eta_0^2 - \eta^2 &= (1 - \lambda)^2(\alpha_2^2 + \alpha_3^2 + \alpha_4^2) + \lambda^2(\beta_2^2 + \beta_3^2 + \beta_4^2) \\ &\quad + 2\lambda(1 - \lambda)\sqrt{\alpha_2^2 + \alpha_3^2 + \alpha_4^2}\sqrt{\beta_2^2 + \beta_3^2 + \beta_4^2} \\ &\quad - \{[(1 - \lambda)\alpha_2 + \lambda\beta_2]^2 + [(1 - \lambda)\alpha_3 + \lambda\beta_3]^2 + [(1 - \lambda)\alpha_4 + \lambda\beta_4]^2\} \\ &= 2\lambda(1 - \lambda)\left\{\sqrt{\alpha_2^2 + \alpha_3^2 + \alpha_4^2}\sqrt{\beta_2^2 + \beta_3^2 + \beta_4^2} - (\alpha_2\beta_2 + \alpha_3\beta_3 + \alpha_4\beta_4)\right\}, \end{aligned}$$

thus $\eta_0^2 - \eta^2 \geq 0$. Hence from this it follows that (see Fig. 3) the bild of a point on the connection line cannot lie outside the trapezoid formed by the points a, \bar{a}, b, \bar{b} and the lines connecting them. The bild of the connection line consists therefore of two (continuous) curves symmetric to the real axis, which cannot have a point outside the trapezoid a, \bar{a}, b, \bar{b} . Since $B(\mathbf{A})$ is the numerical range of a complex matrix^[XIII] it is convex, so all points of the trapezoid a, \bar{a}, b, \bar{b} belong to it and from this also the bilds of all points on the connection line from \mathbf{a} to \mathbf{b} are in $B(\mathbf{A})$,

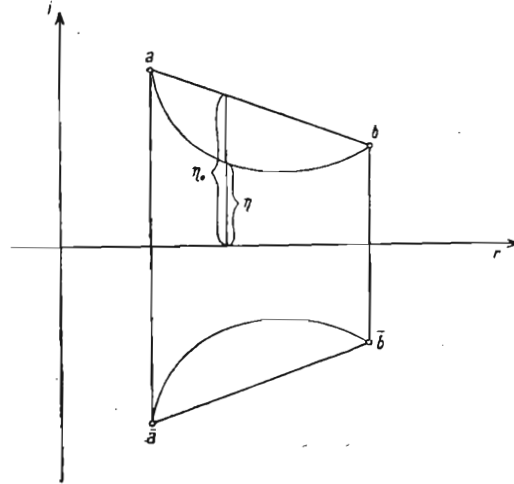


Fig. 3.

therefore also all points on the connection lines between the quaternions \mathbf{a} , \mathbf{b} are in $W(\mathbf{A})$. \square

12. THE BOUNDARY GENERATING HYPERSURFACE

The equation of the boundary generating curve of the image matrix $M(\mathbf{A})$ of a quaternion matrix $\mathbf{A} = \mathbf{H}_1 + \mathbf{H}_2\mathbf{j}$ of dimension n is obtained from the decomposition of $M(\mathbf{A})$ into real and imaginary parts:

$$M(\mathbf{A}) = M(\mathbf{H}_1) + \left(\frac{M(\mathbf{H}_2)M(\mathbf{j})}{i} \right) i.$$

From this follows the equation for the boundary generating curve of $M(\mathbf{A})$

$$\left| M(\mathbf{H}_1)u + \frac{M(\mathbf{H}_2)M(\mathbf{j})}{i}v + I_{2n}w \right| = 0.$$

The boundary generating hypersurface of the matrix \mathbf{A} in R_4 is then produced from the boundary generating curve $M(\mathbf{A})$ by means of rotation about the real axis. But now we have: if $w_1 = u$, $w_2 = v$, $w_3 = w_4 = 0$, $w_5 = w$ are the hyperplane coordinates of a hyperplane ε of R_4 in the homogenous point coordinates x_1, x_2, x_3, x_4, x_5 , then the hyperplane coordinates of all hyperplanes which are produced from ε by rotation about the axis $x_2 = x_3 = x_4 = 0$, satisfy the equations

$$w_1 = u, \quad w_2 + w_3 + w_4 = v^2, \quad w_5 = w.$$

From this we get the equation for the boundary generating hypersurface of the matrix \mathbf{A} :

$$\left| M(\mathbf{H}_1)w_1 + \frac{M(\mathbf{H}_2)M(\mathbf{j})}{i} \sqrt{w_2^2 + w_3^2 + w_4^2} + I_{2n}w_5 \right| = 0.$$

The real foci of the boundary generating curve of $M(\mathbf{A})$ are the eigenvalues of $M(\mathbf{A})$. Now the eigenvalues of $M(\mathbf{A})$ are also representatives of the eigenvalue classes of \mathbf{A} . With the rotation of the boundary generating curve of $M(\mathbf{A})$ about

the real axis, the foci describe precisely the spheres belonging to the eigenvalue classes of \mathbf{A} . From this it follows immediately:

Theorem 37. *The real focal spheres of the boundary generating hypersurface of a quaternion matrix \mathbf{A} are the spheres which represent the eigenvalue classes of \mathbf{A} .*

Now let $\mathbf{H} = -(\mathbf{H}_1 u + \mathbf{H}_2 v)$, where u, v are real variables. Then we have

$$H = M(\mathbf{H}) = -(M(\mathbf{H}_1)u + M(\mathbf{H}_2)v) = -(H_1 u + H_2 v),$$

and since $\mathbf{H}_1, \mathbf{H}_2$ are Hermitian, also their image matrices are Hermitian.^[XIV] The boundary generating curve of the complex matrix $A = M(\mathbf{H}_1) + iM(\mathbf{H}_2)$, which in general is not the image matrix of a quaternion matrix, is then obtained through the equation

$$\begin{aligned} f_A(u, v, w) &\equiv |M(\mathbf{H}_1)u + M(\mathbf{H}_2)v + I_{2n}w| \\ &\equiv w^{2n} + C_1(u, v)w^{2n-1} + \cdots + C_{2n}(u, v) = 0. \end{aligned}$$

For real values of u, v the coefficients $C_\nu(u, v)$ are thereby always real. But now $H = -(M(\mathbf{H}_1)u + M(\mathbf{H}_2)v)$ satisfies the equation

$$f_A(u, v, H) = 0.$$

Since the coefficients in $f_A(u, v, w)$ are real, it follows

$$f_A(u, v, \mathbf{H}) = 0.$$

So to each arbitrary square quaternion matrix $\mathbf{A} = \mathbf{H}_1 + \mathbf{H}_2 \mathbf{j}$ of dimension n can be specified an algebraic equation of order $2n$, where for the Hermitian matrix $\mathbf{H} = -(\mathbf{H}_1 u + \mathbf{H}_2 v)$ corresponding to \mathbf{A} it suffices to consider u, v as real variables. Now let $m(u, v, w)$ be the minimal polynomial of H . Then $m(u, v, w)$ is a factor of $f_A(u, v, w)$. Just as $f_A(u, v, w)$ has only real coefficients, the same is true for $m(u, v, w)$, for otherwise $\overline{m(u, v, w)}$ should also be a minimal polynomial, which contradicts the uniqueness of the minimal polynomial, unless all of the coefficients are real. But then we also have

$$m(u, v, \mathbf{H}) = 0.$$

Certainly \mathbf{H} does not satisfy any algebraic equation with real coefficients lower than the order of $m(u, v, w)$; for then $M(\mathbf{H})$ would also do the same, contradicting the minimality of $m(u, v, w)$.

If we define as the *minimal polynomial of $\mathbf{H} = -(\mathbf{H}_1 u + \mathbf{H}_2 v)$* the polynomial of smallest order with real coefficients depending on u and v which vanishes for \mathbf{H} , then we get:

Theorem 38. *The minimal polynomial of the quaternion matrix $\mathbf{H} = -(\mathbf{H}_1 u + \mathbf{H}_2 v)$ agrees with the minimal polynomial of the image matrix $M(\mathbf{H}) = -(M(\mathbf{H}_1)u + M(\mathbf{H}_2)v)$.*

In contrast to the properties for complex matrices, the minimal polynomial of $\mathbf{H} = -(\mathbf{H}_1 u + \mathbf{H}_2 v)$ is not simultaneously the minimal polynomial of the boundary generating hypersurface.

13. BILINEAR FORMS OF QUATERNION MATRICES

Finally we carry over some results for complex bilinear forms to the realm of quaternions.

If A is a square complex matrix, then the totality of the points which can be assumed by the bilinear form x^*Ay under the auxiliary condition $x^*x = y^*y = 1$ is a closed circular disc $R(A)$ about the origin which contains the numerical range $W(A)$. If we set

$$\max(|x^*Ax|) = \omega(A) \quad (x^*x = 1),$$

then this gives us for the radius $\varrho(A)$ of the disc $R(A)$:

$$(16) \quad 0 \leq \omega(A) \leq \varrho(A) \leq 2\omega(A).^{21}$$

All of this may be carried over to quaternion matrices. We define the *numerical range* $R(\mathbf{A})$ of the bilinear form of U to be the set of points in quaternion space that can be assumed by the bilinear form

$$\mathbf{X}^*\mathbf{A}\mathbf{Y} \quad \text{under the auxiliary condition} \quad \mathbf{X}^*\mathbf{X} = \mathbf{Y}^*\mathbf{Y} = 1.$$

The set of points $R(\mathbf{A})$ is a 4-dimensional sphere centered at the origin; for if $\mathbf{a} \in R(\mathbf{A})$, so

$$\mathbf{X}_0^*\mathbf{A}\mathbf{Y}_0 = \mathbf{a} \quad (\mathbf{X}_0^*\mathbf{X}_0 = \mathbf{Y}_0^*\mathbf{Y}_0 = 1),$$

then for any unitary quaternion \mathbf{u} it is true that $\mathbf{u}\mathbf{a} \in R(\mathbf{A})$, since

$$\mathbf{u}\mathbf{X}_0^*\mathbf{A}\mathbf{Y}_0 = \mathbf{Z}_0^*\mathbf{A}\mathbf{Y}_0 = \mathbf{u}\mathbf{a} \quad (\mathbf{Z}_0 = \mathbf{X}_0\mathbf{u}^*, \mathbf{Z}_0^*\mathbf{Z}_0 = \mathbf{Y}_0^*\mathbf{Y}_0 = 1).$$

As for the set of points $W(\mathbf{A})$ we have for $R(\mathbf{A})$:

Theorem 39. *If $R(\mathbf{A})$ is the numerical range of the bilinear form of \mathbf{A} , then $R(\mathbf{A})$ intersects the complex plane in the set of points $Q(\mathbf{A})$ for which it holds that:*

$$Q(\mathbf{A}) = R(M(\mathbf{A})),$$

where $R(M(\mathbf{A}))$ is the (complex) numerical range of the bilinear form of the matrix $M(\mathbf{A})$.

Proof. Suppose $R(\mathbf{A})$ contains the quaternion \mathbf{a} :

$$\mathbf{X}_0^*\mathbf{A}\mathbf{Y}_0 = \mathbf{a} \quad (\mathbf{X}_0^*\mathbf{X}_0 = \mathbf{Y}_0^*\mathbf{Y}_0 = 1).$$

Then there is at least one unitary quaternion \mathbf{u} such that $\mathbf{u}^*\mathbf{a}\mathbf{u} = c$ is complex; thus

$$\mathbf{u}^*\mathbf{X}_0^*\mathbf{A}\mathbf{Y}_0\mathbf{u} = \mathbf{X}_1^*\mathbf{A}\mathbf{Y}_1 = \mathbf{u}^*\mathbf{a}\mathbf{u} = c \quad (\mathbf{X}_1^*\mathbf{X}_1 = \mathbf{Y}_1^*\mathbf{Y}_1 = 1),$$

consequently c is a complex representative of the class of \mathbf{a} . If we set

$$M(\mathbf{X}_1) = (x_1, M(\mathbf{j})\bar{x}_1), \quad M(\mathbf{Y}_1) = (y_1, M(\mathbf{j})\bar{y}_1),$$

then

$$x_1^*M(\mathbf{A})y_1 = \mathbf{u}^*\mathbf{a}\mathbf{u} = c,$$

thus $Q(\mathbf{A}) \subseteq R(M(\mathbf{A}))$.

Conversely if $\mathbf{a} \in R(M(\mathbf{A}))$:

$$x_0^*M(\mathbf{A})y_0 = \mathbf{a} \quad (x_0^*x_0 = y_0^*y_0 = 1),$$

then also

$$\bar{a} = \bar{x}_0^*\overline{M(\mathbf{A})}\bar{y}_0 = \bar{x}_0^*M(\mathbf{j}^{-1})M(\mathbf{A})M(\mathbf{j})\bar{y}_0 = x_1^*M(\mathbf{A})y_1$$

²¹WINTNER [12]

with

$$x_1^* = \bar{x}_0^* M(\mathbf{j}^{-1}), \quad y_1 = M(\mathbf{j}) \bar{y}_0; \quad x_1^* x_1 = y_1^* y_1 = 1.$$

The matrices $(x_0, x_1), (y_0, y_1)$ are image matrices of two quaternion vectors $\mathbf{X}_0, \mathbf{Y}_0$, for which because of

$$\begin{pmatrix} x_0^* \\ x_1^* \end{pmatrix} M(\mathbf{A})(y_0, y_1) = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}$$

it follows

$$\mathbf{X}_0^* \mathbf{A} \mathbf{Y}_0 = a + b\mathbf{j}.$$

From this it follows if a is a complex number in $R(M(\mathbf{A}))$, then there exists a complex number in $Q(\mathbf{A})$, whose real part agrees with the real part of a , but whose absolute value is no smaller than the absolute value of a . Since $R(M(\mathbf{A}))$ and $Q(\mathbf{A})$ are discs and since $Q(\mathbf{A}) \subseteq R(M(\mathbf{A}))$, now it follows that $Q(\mathbf{A}) = R(M(\mathbf{A}))$. \square

From the proof of this theorem it now follows that the inequality (16) may be carried over. Thus if $\omega(\mathbf{A})$ is the maximum absolute value of all points of $W(\mathbf{A})$ and if $\varrho(\mathbf{A})$ is the radius of $R(\mathbf{A})$, then we obtain

Theorem 40.

$$0 \leq \omega(\mathbf{A}) \leq \varrho(\mathbf{A}) \leq 2\omega(\mathbf{A}).$$

REFERENCES

1. T. BONNESEN AND W. FENCHEL, *Theorie der konvexen Körper*. *Ergebn. Math. Grenzgeb.* **3**, Nr. 1 (1934).
2. H. BRUNN, *Über Kurven ohne Wendepunkte*. Habilitationsschrift München (1889).
3. C.C. MACDUFFEE, *The Theory of Matrices*, *Ergebn. Math. Grenzgeb.* **2**, Nr.5 (1933).
4. F. HAUSDORFF, *Der Wertevorrat einer Bilinearform*, *Math Z.*, Berlin **3**, 314–316 (1919).
5. H.C. LEE, *Eigenvalues and canonical forms of matrices with quaternion coefficients*. *Proc. Irish Acad. A* **52**, 253–260 (1949).
6. I. SCHUR, *Zwei Sätze über algebraische Gleichungen mit lauter reellen Wurzeln*. *J. reine angew. Math.* **144**, 75–88 (1914).
7. W. SPECHT, *Zur Theorie der Gruppen linearer Substitution*. *II. Jber. Deutsche Math.-Vereinig.* **49**, 207–215 (1939).
8. ———, *Zur Theorie der Matrizen*. *II. Jber. Deutsche Math.-Vereinig.* **50**, 19–23 (1940).
9. O. TOEPLITZ, *Das algebraische Analogon zu einem Satz von Fejér*. *Math Z.*, Berlin **2**, 187–197 (1918).
10. H. WEYL, *Gruppentheorie und Quantenmechanik*. Leipzig (1931).
11. H. WIELEITNER, *Theorie der ebenen algebraischen Kurven höherer Ordnung*. Leipzig (1905).
12. A. WINTNER, *Spektraltheorie der unendlichen Matrizen*. Leipzig (1929).

NOTES FROM THE TRANSLATORS

- [I] – AU-YEUNG’s counterexample to Theorem 36 is the matrix $\mathbf{A} = \begin{pmatrix} i & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.
cf. Y.H.AU-YEUNG, *On the Convexity of Numerical Range in Quaternionic Hilbert Spaces*, LINEAR AND MULTILINEAR ALGEBRA, 16:93-100, (1984).
- [II] – In this section (Section 3) the words *complex plane* refer to \mathbb{C}^2 , while the words *Gaussian plane* refer to \mathbb{C} . The *projective plane* here is $\mathbb{C}\mathbb{P}^2$.
- [III] – A literal translation would be “finite”, but the translators felt that “proper” makes the discussion preceding this theorem more lucid.
- [IV] – This hypothesis was the subject of later research, see, for example, H. SHAPIRO, *On a Conjecture of Kippenhahn about the Characteristic Polynomial of a Pencil Generated by Two Hermitian Matrices. I*, LINEAR ALGEBRA AND ITS APPLICATIONS, 43:201-221, (1982).
- [V] – While \mathbf{H}_1 is indeed Hermitian here, \mathbf{H}_2 is not. It is true that $\mathbf{H}_2\mathbf{j}$ is skew-Hermitian, i.e., $(\mathbf{H}_2\mathbf{j})^* = -(\mathbf{H}_2\mathbf{j})$, and it follows that $\frac{M(\mathbf{H}_2\mathbf{j})}{i}$ is Hermitian.
- [VI] – A direct translation of the German word “bild” is “image” or “picture”. The translators chose to keep the German word, as was done in the English paper F. ZHANG, *Quaternions and Matrices of Quaternions*, LINEAR ALGEBRA AND ITS APPLICATIONS, 251:21-57, (1997).
- [VII] – Actually, Theorem 35 is false. It is true that $B(\mathbf{A}) \subseteq W(M(\mathbf{A}))$.
- [VIII] – Actually, here and throughout the remainder of the paper, $\mathbf{y}_0 = M(\mathbf{j}^{-1})\bar{x}_0$. Since we also have $M(\mathbf{j})M(\mathbf{A})M(\mathbf{j}^{-1}) = \overline{M(\mathbf{A})}$, we can exchange the roles of $M(\mathbf{j})$ and $M(\mathbf{j}^{-1})$ without affecting the proofs.
- [IX] – This lemma is false, which is why Theorems 35, 36 do not hold.
- [X] – This line was the flaw in the argument. Check that for any value of ϑ the bild of the quaternion $\mathbf{Z}_0^*\mathbf{A}\mathbf{Z}_0$ consists of the two complex numbers a, \bar{a} . As ϑ varies the bild does not trace out a segment, but it remains constant.
- [XI] – Notice the problem with this statement, for the bild of a quaternion always consists of a complex number and its complex conjugate, in this case a and \bar{a} .
- [XII] – For a nice discussion of the question of convexity of the quaternionic numerical range, see F. ZHANG’s work mentioned in [VI] above.
- [XIII] – Here Theorem 35 is used, but this theorem is not valid.
- [XIV] – \mathbf{H}_2 is not Hermitian, nor is $M(\mathbf{H}_2)$ (see [V] above).

The translators wish to thank PROFESSOR RUDOLF KIPPENHAHN and also WILEY-VCH VERLAG GMBH & CO for giving permission to publish this translation. They also wish to thank DAVID SINGER of CASE WESTERN RESERVE UNIVERSITY for all of his assistance with this project.