Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Linear Algebra and its Applications 390 (2004) 121–136
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

Numerical range of a normal compression II ☆

Hwa-Long Gau [a,*], Pei Yuan Wu [b]

[a] *Department of Mathematics, National Central University, Chung-Li 320, Taiwan*
[b] *Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan*

Received 9 September 2003; accepted 14 April 2004

Submitted by C.-K. Li

Abstract

As in the predecessor [Numerical range of a normal compression, Linear and Multilinear Algebra, in press] of this paper, we consider properties of matrices of the form V^*NV, where $N = \mathrm{diag}(a_1, \ldots, a_{n+1})$ is a diagonal matrix with distinct eigenvalues a_js such that each of them is a corner of the convex hull they generate, and V is an $(n+1)$-by-n matrix with $V^*V = I_n$ such that any nonzero vector orthogonal to the range space of V has all its components nonzero. We obtain that such a matrix A is determined by its eigenvalues up to unitary equivalence, is irreducible and cyclic, and the boundary of its numerical range is a differentiable curve which contains no line segment. We also consider the condition for the existence of another matrix of the above type which dilates to A such that their numerical ranges share some common points with the boundary of the $(n+1)$-gon $a_1 \cdots a_{n+1}$.

AMS classification: 15A60; 47A20

Keywords: Numerical range; Normal compression; Irreducible matrix; Cyclic matrix

1. Introduction

Following our work in [9], we continue the study of properties of matrices obtained in the following way. Let

$$N = \mathrm{diag}(a_1, \ldots, a_{n+1}) \tag{1}$$

☆ Research supported by the National Science Council of the Republic China.
* Corresponding author.
E-mail addresses: hlgau@math.ncu.edu.tw (H.-L. Gau), pywu@math.nctu.edu.tw (P.Y. Wu).

0024-3795/$ - see front matter
doi:10.1016/j.laa.2004.04.013

be a diagonal matrix with distinct eigenvalues a_js such that each of them is a corner of the convex hull they generate. We assume that the a_js are arranged in the counterclockwise orientation. Let V be an $(n+1)$-by-n matrix with $V^*V = I_n$, the n-by-n identity matrix, such that its n-dimensional range space is orthogonal to a unit vector $x = [x_1 \ \cdots \ x_{n+1}]^{\mathrm{T}}$ in $\mathbb{C}^{n+1}$ with $x_j \neq 0$ for all j. The type of matrices which we will study is of the form

$$A = V^*NV. \tag{2}$$

Recall that an n-by-n matrix B $(n < m)$ is said to *dilate* to the m-by-m matrix C or B is a *compression* of C if $B = V^*CV$ for some m-by-n matrix V with $V^*V = I_n$ or, equivalently, C is unitarily equivalent to a matrix of the form $\left[\begin{smallmatrix} B & * \\ * & * \end{smallmatrix}\right]$. Hence the matrix A in (2) dilates to (i.e., it is a compression of) N with N and V having some special properties. For convenience, we denote by $\mathscr{N}_n$ the class of n-by-n matrices which are unitarily equivalent to one of the form in (2). When the above diagonal matrix N is unitary, the corresponding A is exactly the $\mathscr{S}_n$-matrix or the UB-matrix studied in [5–7,10,12–14] (cf. [10, Lemma 2.2]). The numerical ranges of latter matrices are known to have the $(n+1)$-Poncelet property. For an account of the developments on this subject, the reader can consult the survey paper [16] and the more recent one [8]. It turns out that many of the properties for the $\mathscr{S}_n$-matrices can be generalized to ones for the more general $\mathscr{N}_n$-matrices. Adam and Maroulas [1] and Mirman and Wu [15] are the first attempts along this line. A more systematic investigation was carried out in [9]. The purpose of this paper is to further develop the ideas from [9].

Recall that the *numerical range* of an n-by-n matrix A is by definition the set $W(A) = \{\langle Ax, x\rangle : x \in \mathbb{C}^n, \|x\| = 1\}$, where $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ denote, respectively, the standard inner product and Euclidean norm in $\mathbb{C}^n$. The numerical range is a nonempty compact convex subset of the complex plane. It is invariant under unitary equivalence and contains the eigenvalues. When the matrix is normal, its numerical range coincides with the convex hull of its eigenvalues. Moreover, the numerical range of a compression of the matrix A is contained in the numerical range of A. For other properties of the numerical range, the reader can consult [11, Chapter 1].

We have to emphasize that the generalizations from $\mathscr{S}_n$-matrices to $\mathscr{N}_n$-matrices are not straightforward. As opposed to the unitary compression case in which the unit circle in the complex plane plays the role of the parametrizing curve for the $(n+1)$-gons $a_1 \cdots a_{n+1}$ circumscribing the numerical range of A, the corresponding curve for the normal case is still unknown. As a consequence, we can only consider one $(n+1)$-gon at a time and we are not able to reach those results which depend on the existence of infinitely many such circumscribing $(n+1)$-gons. Another difficulty is that the matrices in class $\mathscr{N}_n$ are not necessarily contractions, which renders the rich structure theory of contractions not applicable. To overcome these obstacles, we need to devise the arguments more cleverly, which in turn gives more insight even to the unitary case.

In Section 2, we first prove that an $\mathcal{N}_n$-matrix A is determined by its eigenvalues up to unitary equivalence. This is comparable to [9, Theorem 4], in which we have the determination of A by its numerical range. The former is even true for an arbitrary n-dimensional compression of any $(n+1)$-by-$(n+1)$ normal matrix, while the same cannot be said for the latter. We also supplement [9, Theorem 3] by characterizing the $\mathcal{N}_n$-matrices A whose numerical range has boundary tangent to every edge of the $(n+1)$-gon $a_1 \cdots a_{n+1}$ at its midpoints as those whose eigenvalues are the zeros of the derivative of the characteristic polynomial of N. In Section 3, we prove that every $\mathcal{N}_n$-matrix is irreducible and cyclic, and that the boundary of its numerical range is differentiable and contains no line segment. In Section 4, we consider the problem of compressing an $\mathcal{N}_n$-matrix to another $\mathcal{N}_m$-matrix so that the boundaries of their numerical ranges share m tangent points with the edges of the $(n+1)$-gon $a_1 \cdots a_{n+1}$. We show that this is the case exactly when the m tangent points are on successive edges of the polygon. This gives a clear illustration in a most natural context why [3, Theorem 6] should be true.

2. Spectrum

In this section, we present some results which supplement those in [9], the predecessor of this paper. The first one says that the matrices in $\mathcal{N}_n$ associated with the same diagonal matrix are determined up to unitary equivalence by their eigenvalues. This is comparable to [9, Theorem 4], which says that such matrices are determined by their numerical ranges. In fact, more is true.

Theorem 2.1. *Let*

$$N = \operatorname{diag}(\underbrace{a_1, \ldots, a_1}_{k_1}, \ldots, \underbrace{a_m, \ldots, a_m}_{k_m})$$

with a_js the successive vertices of the convex m-gon $a_1 a_2 \cdots a_m$ and $k_j \geqslant 1$ for all j, let $A = V_1^* N V_1$ and $B = V_2^* N V_2$, where V_1 and V_2 are $(n+1)$-by-n matrices with $V_1^* V_1 = V_2^* V_2 = I_n$ $(n \equiv (\sum_{j=1}^m k_j) - 1)$, and let $x = [x_1 \ \cdots \ x_{n+1}]^{\mathrm{T}}$ and $y = [y_1 \ \cdots \ y_{n+1}]^{\mathrm{T}}$ be associated unit vectors orthogonal to the range spaces of V_1 and V_2, respectively. Then the following statements are equivalent:

(a) *A is unitarily equivalent to B*;
(b) *the eigenvalues of A and B coincide, counting (algebraic) multiplicities*;
(c) $\sum_{l=l_j+1}^{l_{j+1}} |x_l|^2 = \sum_{l=l_j+1}^{l_{j+1}} |y_l|^2$ *for $j = 1, 2, \ldots, m$, where $l_1 = 0$, $l_j = \sum_{p=1}^{j-1} k_p$ for all j, $2 \leqslant j \leqslant m$, and $l_{m+1} = n+1$.*

The proof can be facilitated by the following lemma.

Lemma 2.2. *Let N and A be as in Theorem* 2.1. *Then A is unitarily equivalent to the direct sum $N_0 \oplus A'$, where*

$$N_0 = \operatorname{diag}(\underbrace{a_1, \ldots, a_1}_{k_1-1}, \ldots, \underbrace{a_m, \ldots, a_m}_{k_m-1})$$

and $A' = V'^* N' V'$ with $N' = \operatorname{diag}(a_1, \ldots, a_m)$ and V' an m-by-$(m-1)$ matrix satisfying $V'^* V' = I_{m-1}$ with range space orthogonal to $x' = [x'_1 \ \cdots \ x'_m]^{\mathrm{T}}$, $x'_j = \left(\sum_{l=l_j+1}^{l_{j+1}} |x_l|^2\right)^{1/2}$, $l_1 = 0$, $l_j = \sum_{p=1}^{j-1} k_p$, $2 \leqslant j \leqslant m$, and $l_{m+1} = n+1$.

Proof. Let K be the range space of V_1. Since

$$\begin{aligned}
&\dim(K \cap \ker(N - a_j I_{n+1})) \\
&\quad = \dim K + \dim \ker(N - a_j I_{n+1}) - \dim(K + \ker(N - a_j I_{n+1})) \\
&\quad \geqslant n + k_j - (n+1) \\
&\quad = k_j - 1
\end{aligned}$$

for every j, there are orthonormal vectors $y_1, \ldots, y_{n-m+1}$ with $y_{l_j-j+2}, \ldots, y_{l_{j+1}-j}$ in $K \cap \ker(N - a_j I_{n+1})$, $1 \leqslant j \leqslant m$. For each j, we add one more vector $y_{n-m+1+j}$ to the jth group of such vectors to form an orthonormal basis of $\ker(N - a_j I_{n+1})$. Then $\{y_1, \ldots, y_{n+1}\}$ is an orthonormal basis of $\mathbb{C}^{n+1}$. Let U be the $(n+1)$-by-$(n+1)$ unitary matrix such that $U y_l = e_l$, the vector with its lth component equal to 1 and the rest of its components equal to 0, $1 \leqslant l \leqslant n+1$. Since x is orthogonal to $y_1, \ldots, y_{n-m+1}$, we have

$$Ux = \big[\underbrace{0 \quad \cdots \quad 0}_{n-m+1} \quad \mathrm{e}^{\mathrm{i}\theta_1} x'_1 \quad \cdots \quad \mathrm{e}^{\mathrm{i}\theta_m} x'_m\big]^{\mathrm{T}}$$

for some real $\theta_1, \ldots, \theta_m$. By applying another diagonal unitary matrix, we may assume that U is such that $Ux = [0 \ \cdots \ 0 \ x'_1 \ \cdots \ x'_m]^{\mathrm{T}} \equiv y$. As a linear transformation, U maps the orthogonal complement of M_x to that of M_y, where M_x and M_y denote the subspaces generated by x and y, respectively. The former coincides with the range space K of V_1 and the latter coincides with $\mathbb{C}^{n-m+1} \oplus K'$, where K' is the orthogonal complement in $\mathbb{C}^m$ of the subspace generated by $[x'_1 \ \cdots \ x'_m]^{\mathrm{T}}$. Thus U can be decomposed as

$$U = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} : K \oplus M_x \to (\mathbb{C}^{n-m+1} \oplus K') \oplus M_y.$$

Since U commutes with N and N is decomposed as

$$\begin{bmatrix} A & * \\ * & * \end{bmatrix} \quad \text{on } K \oplus M_x$$

and

$$\left[\begin{array}{cc|c} N_0 & 0 & 0 \\ 0 & A' & * \\ \hline 0 & * & * \end{array}\right] \quad \text{on } (\mathbb{C}^{n-m+1} \oplus K') \oplus M_y,$$

we infer that $U_1 A = (N_0 \oplus A')U_1$, which implies that A is unitarily equivalent to $N_0 \oplus A'$ of the asserted form. □

Proof of Theorem 2.1. To prove (b) ⇒ (c), let the (common) eigenvalues of A and B be $b_1, \ldots, b_n$. By [9, Theorem 1], these are related to the a_js by

$$\begin{aligned} \prod_{k=1}^{n}(z-b_k) &= \sum_{j=1}^{n+1} |x_j|^2 (z-a'_1)\cdots\widehat{(z-a'_j)}\cdots(z-a'_{n+1}) \\ &= \sum_{j=1}^{n+1} |y_j|^2 (z-a'_1)\cdots\widehat{(z-a'_j)}\cdots(z-a'_{n+1}), \end{aligned}$$

where the hat "∧" over $z - a'_j$ indicates that the term is missing from the products, and $a'_j = a_p$ if $l_p + 1 \leqslant j \leqslant l_{p+1}$, $1 \leqslant p \leqslant m$. Taking out the common factor $\prod_{j=1}^{m}(z-a_j)^{k_j-1}$ from the latter two polynomials, we have

$$\begin{aligned} &\sum_{j=1}^{m} |x'_j|^2 (z-a_1)\cdots\widehat{(z-a_j)}\cdots(z-a_m) \\ &\quad = \sum_{j=1}^{m} |y'_j|^2 (z-a_1)\cdots\widehat{(z-a_j)}\cdots(z-a_m), \end{aligned}$$

where $x'_j = \left(\sum_{l=l_j+1}^{l_{j+1}} |x_l|^2\right)^{1/2}$ and $y'_j = \left(\sum_{l=l_j+1}^{l_{j+1}} |y_l|^2\right)^{1/2}$, $1 \leqslant j \leqslant m$. Plugging $z = a_j$ into the above equation yields

$$\begin{aligned} &|x'_j|^2 (a_j-a_1)\cdots\widehat{(a_j-a_j)}\cdots(a_j-a_m) \\ &\quad = |y'_j|^2 (a_j-a_1)\cdots\widehat{(a_j-a_j)}\cdots(a_j-a_m). \end{aligned}$$

Since the a_js are distinct, we obtain $x'_j = y'_j$ for all j, that is, (c) holds.

The implication (c) ⇒ (a) follows by Lemma 2.2. □

Note that when N has multiple eigenvalues, the analogue of Theorem 2.1 for numerical ranges is in general false. For example, if $N = \mathrm{diag}(0, 0, 1, 2)$, $A = \mathrm{diag}(0, 1, 2)$ and $B = \mathrm{diag}(0, 1/2, 2)$, then A and B both dilate to N, have numerical ranges equal to $[0, 2]$, and are not unitarily equivalent. Another example is provided by $N = \mathrm{diag}(1, \omega, \omega^2, \omega^3, \omega^4)$, $A = \mathrm{diag}(1, \omega^2, \omega^3, \mathrm{Re}\,\omega)$ and $B = \mathrm{diag}(1, \omega^2, \omega^3, (\mathrm{Re}\,\omega) + \varepsilon\mathrm{i})$, where ω is the fifth primitive root of unity and $\varepsilon > 0$ is sufficiently small.

Part of the next result is mentioned in [9]. Here we give the complete version.

Theorem 2.3. *Let $N = \operatorname{diag}(a_1, \ldots, a_{n+1})$ be a diagonal matrix as in* (1) *with the characteristic polynomial $p(z) = \prod_{j=1}^{n+1}(z - a_j)$, and let $A = V^*NV$ be an $\mathscr{N}_n$-matrix as in* (2). *Then the following statements are equivalent*:

(a) $|x_j| = 1/\sqrt{n+1}$ *for all* j, $1 \leqslant j \leqslant n+1$;
(b) *the eigenvalues of A coincide with the zeros of the derivative p' of p, counting multiplicities*;
(c) *the edge $[a_j, a_{j+1}]$ of the $(n+1)$-gon $a_1 \cdots a_{n+1}$ is tangent to $\partial W(A)$, the boundary of the numerical range of A, at its midpoint $(a_j + a_{j+1})/2$ for every j, $1 \leqslant j \leqslant n+1$ $(a_{n+2} \equiv a_1)$.*

The implication (b) $\Rightarrow$ (c) is essentially proved in [15].

Proof of Theorem 2.3. If (a) holds, then, by [9, Theorem 1], the characteristic polynomial of A is

$$\frac{1}{n+1}\sum_{j=1}^{n+1}(z-a_1)\cdots\widehat{(z-a_j)}\cdots(z-a_{n+1}),$$

which is a scalar multiple of p'. This proves (b).

Assume next that (b) is true, that is, the eigenvalues of A coincide with the zeros (counting multiplicities) of

$$p'(z) = \sum_{j=1}^{n+1}(z-a_1)\cdots\widehat{(z-a_j)}\cdots(z-a_{n+1}).$$

Since these eigenvalues are exactly the zeros (counting multiplicities) of

$$\sum_{j=1}^{n+1}|x_j|^2(z-a_1)\cdots\widehat{(z-a_j)}\cdots(z-a_{n+1})$$

by [9, Theorem 1], we infer from the equality of $(1/(n+1))p'$ and this latter polynomial that

$$\frac{1}{n+1}\prod_{l\neq j}(a_j-a_l) = |x_j|^2\prod_{l\neq j}(a_j-a_l), \quad 1 \leqslant j \leqslant n+1.$$

As the a_js are distinct, we have $|x_j| = 1/\sqrt{n+1}$ for all j. Since the tangent point of $[a_j, a_{j+1}]$ with $\partial W(A)$ is given by

$$\frac{1}{|x_j|^2+|x_{j+1}|^2}\left(a_j|x_{j+1}|^2 + a_{j+1}|x_j|^2\right)$$

in general (see [9, Theorem 3] and its proof), condition (c) follows.

Finally, if (c) holds, then

$$\frac{1}{2}(a_j+a_{j+1}) = \frac{1}{|x_j|^2+|x_{j+1}|^2}\left(a_j|x_{j+1}|^2 + a_{j+1}|x_j|^2\right)$$

for all j, $1 \leqslant j \leqslant n+1$, where $a_{n+2} \equiv a_1$ and $x_{n+2} \equiv x_1$. These imply that $|x_1|^2 = \cdots = |x_{n+1}|^2$. As their sum is equal to one, we have $|x_j| = 1/\sqrt{n+1}$ for all j, that is, (a) holds. □

3. Numerical range

Let $A = V^*NV$ be a matrix in the class $\mathcal{N}_n$ as in (2). Our first theorem says, among other things, that the boundary of its numerical range $\partial W(A)$ contains no line segment and is a differentiable curve.

Theorem 3.1. *If A is an $\mathcal{N}_n$-matrix, then*

(a) $\partial W(A)$ *contains no line segment*;
(b) *for any* λ *in* $\partial W(A)$, *the set* $\{y \in \mathbb{C}^n : \langle Ay, y\rangle = \lambda\|y\|^2\}$ *is a vector space of dimension one*;
(c) A *is irreducible*; *and*
(d) $\partial W(A)$ *is a differentiable curve.*

A matrix A is said to be *irreducible* if it is not unitarily equivalent to the direct sum of two other matrices; otherwise, A is called *reducible*. These definitions of reducible and irreducible matrices are commonly used in operator theory and are different from the classical ones for entrywise nonnegative matrices (in the Perron–Frobenius theory).

Proof of Theorem 3.1. (a) Assume that $[a, b]$ is a line segment in $\partial W(A)$ with maximum length. As proved in [9, Theorem 3], for each j, $1 \leqslant j \leqslant n+1$, the intersection $W(A) \cap [a_j, a_{j+1}]$ is a singleton $\{c_j\}$ given by $c_j = \langle Ay_j, y_j\rangle$ with

$$Vy_j = \frac{1}{(|x_j|^2 + |x_{j+1}|^2)^{1/2}}(\overline{x_{j+1}}e_j - \overline{x_j}e_{j+1}) \tag{3}$$

($x_1, \ldots, x_{n+1} \neq 0$, $x_{n+2} \equiv x_1$ and $e_{n+2} \equiv e_1$). We may assume that $[a, b]$ lies between c_j and c_{j+1} of the boundary $\partial W(A)$. Let L denote the subspace of $\mathbb{C}^{n+1}$ generated by the vectors Vy with $y \in \mathbb{C}^n$ satisfying $\langle Ay, y\rangle = c\|y\|^2$ for some c in $[a, b]$, and let M be the subspace generated by the linearly independent $Vy_1, \ldots, Vy_{j-1}, Vy_{j+2}, \ldots, Vy_{n+1}$. Then

$$\dim(L \cap M) = \dim L + \dim M - \dim(L + M) \geqslant 2 + (n-1) - n = 1.$$

Hence there is a unit vector Vy_0 in $L \cap M$. Since L is equal to the union

$$\bigcup_{c\in[a,b]} \{Vy \in \mathbb{C}^{n+1} : \langle Ay, y\rangle = c\|y\|^2\}$$

(cf. [4, Theorem 1]), the point $\langle Ay_0, y_0\rangle$ lies in $[a, b]$. On the other hand, since

$$Vy_0 = \sum_{\substack{k=1 \\ k\neq j,j+1}}^{n+1} \lambda_k V y_k$$

for some λ_ks, we have

$$\langle Ay_0, y_0\rangle = \langle NVy_0, Vy_0\rangle = \left\langle \sum_{k\neq j,j+1} \lambda_k NVy_k, \sum_{l\neq j,j+1} \lambda_l Vy_l \right\rangle,$$

which, in light of (3), is easily seen to be a convex combination of $a_1, \ldots, a_j, a_{j+2}, \ldots, a_{n+1}$. This leads to a contradiction since $[a, b]$, lying between c_j and c_{j+1}, is disjoint from the convex hull of $a_1, \ldots, a_j, a_{j+2}, \ldots, a_{n+1}$. We conclude that $\partial W(A)$ contains no line segment.

(b) By (a), the boundary of $W(A)$ consists of extreme points of $W(A)$. Hence $L \equiv \{Vy \in \mathbb{C}^{n+1} : \langle Ay, y\rangle = \lambda\|y\|^2\}$ is a vector space for every λ in $\partial W(A)$ (cf. [4, Theorem 1]). Let the c_js be as in (a) and assume that λ lies between c_j and c_{j+1} of the boundary $\partial W(A)$. Let M be the subspace of $\mathbb{C}^{n+1}$ generated by $Vy_1, \ldots, Vy_{j-1}, Vy_{j+2}, \ldots, Vy_{n+1}$. If $\dim L \geqslant 2$, then, as in (a), there is a unit vector Vy_0 in M with $\langle Ay_0, y_0\rangle = \lambda$, which shows that λ is in the convex hull of $a_1, \ldots, a_j, a_{j+2}, \ldots, a_{n+1}$. This contradicts our assumption of λ lying between c_j and c_{j+1}. Hence L can only have dimension one.

(c) Assume that A is reducible, that is, A is unitarily equivalent to a direct sum $A_1 \oplus A_2$. Let U be a unitary matrix such that $U^*AU = A_1 \oplus A_2$. We first prove that each c_j, $1 \leqslant j \leqslant n+1$, lies in one of the numerical ranges $W(A_1)$ and $W(A_2)$. Indeed, since c_j is in $W(A) = \mathrm{conv}\{W(A_1) \cup W(A_2)\}$, the convex hull of $W(A_1) \cup W(A_2)$, it can be expressed as $ta + (1-t)b$ for some $0 \leqslant t \leqslant 1$, a in $W(A_1)$ and b in $W(A_2)$. Since c_j is an extreme point of $W(A)$ by (a), one of the following must hold: $t = 0$, $t = 1$ and $a = b$. This shows that c_j is indeed in either $W(A_1)$ or $W(A_2)$. Assume that c_1 is in $W(A_1)$. We next show that c_2 must also be in $W(A_1)$. For, otherwise, if c_2 is in $W(A_2)$, then there are unit vectors u and v such that $c_1 = \langle A_1 u, u\rangle$ and $c_2 = \langle A_2 v, v\rangle$. Thus $c_1 = \langle AU(u \oplus 0), U(u \oplus 0)\rangle$ and $c_2 = \langle AU(0 \oplus v), U(0 \oplus v)\rangle$. Part (b) implies that $U(u \oplus 0)$ and $U(0 \oplus v)$ are scalar multiples of y_1 and y_2, respectively. Hence y_1 and y_2, and thus Vy_1 and Vy_2, are orthogonal. From the expressions of Vy_1 and Vy_2 in (3), this leads to a contradiction. Thus c_2 is in $W(A_1)$ as asserted. In a similar way, we prove successively that all the other c_js are in $W(A_1)$. It is easily seen from (3) that the Vy_js span the range space of V. Hence the y_js span $\mathbb{C}^n$. We conclude from above that A_1 is of size n, which shows that A is irreducible.

(d) The differentiability of $\partial W(A)$ follows easily from (c) since any nondifferentiable point λ of $\partial W(A)$ is a reducing eigenvalue of A (i.e., $Ay = \lambda y$ and $A^*y = \overline{\lambda} y$ for some nonzero vector y) (cf. [11, Theorems 1.6.3 and 1.6.6]). □

An n-by-n matrix A is called *cyclic* if there is a vector y in $\mathbb{C}^n$ such that $y, Ay, \ldots, A^{n-1}y$ span $\mathbb{C}^n$; in this case, y is called a *cyclic vector* of A. It is easy to see that a

normal matrix is cyclic if and only if its eigenvalues are distinct. The next theorem proves the cyclicity of $\mathcal{N}_n$-matrices.

Theorem 3.2. *Every $\mathcal{N}_n$-matrix A is cyclic. In fact, if y is any unit vector such that $\langle Ay, y\rangle$ is the tangent point of some edge of the $(n+1)$-gon $a_1\cdots a_{n+1}$ with $\partial W(A)$, then y is a cyclic vector of both A and A^*.*

In the unitary compression case, more is true: every unit vector y for which $\langle Ay, y\rangle$ lies on the boundary of $W(A)$ is a cyclic vector of both A and A^*. This is proved in [6, Lemma 3.2] and also in [2, Proposition 2]. We have only had this restricted version in the normal case because of our inability to prove the existence of a circumscribing $(n+1)$-gon of $W(A)$ passing through every given point of $\partial W(A)$. Our proof of Theorem 3.2 is similar to the one for [6, Lemma 3.2].

Proof of Theorem 3.2. We verify that the vector y in $\mathbb{C}^n$ with

$$Vy = \frac{1}{(|x_1|^2 + |x_2|^2)^{1/2}}(\overline{x_2}e_1 - \overline{x_1}e_2)$$

is cyclic for A. Indeed, we first prove that for each j, $1 \leqslant j \leqslant n-1$, the vector VA^jy can be expressed as $N^jVy - \sum_{k=1}^{j}\lambda_k N^{j-k}x$, where $\lambda_k = \langle NVA^{k-1}y, x\rangle$ for each k. This is done by induction on j. If $j=1$, then $VAy = (VV^*)(NVy)$. Since VV^* is the orthogonal projection from $\mathbb{C}^{n+1}$ onto the range space of V, this implies

$$VAy = NVy - \langle NVy, x\rangle x = NVy - \lambda_1 x$$

as asserted. On the other hand, if we assume that $VA^jy = N^jVy - \sum_{k=1}^{j}\lambda_k N^{j-k}x$ holds for some j, then

$$\begin{aligned}
VA^{j+1}y &= V(V^*NV)A^jy \\
&= VV^*\left(N^{j+1}Vy - \sum_{k=1}^{j}\lambda_k N^{j-k+1}x\right) \\
&= \left(N^{j+1}Vy - \sum_{k=1}^{j}\lambda_k N^{j-k+1}x\right) - \langle NVA^jy, x\rangle x \\
&= N^{j+1}Vy - \sum_{k=1}^{j+1}\lambda_k N^{(j+1)-k}x,
\end{aligned}$$

which is the desired expression for $VA^{j+1}y$.

To complete the proof, we show that $y, Ay, \ldots, A^{n-1}y$ are linearly independent. Suppose that $\sum_{j=0}^{n-1}\alpha_j A^jy = 0$ for some scalars α_js. Since

$$
\begin{aligned}
0 = V\left(\sum_{j=0}^{n-1}\alpha_j A^j y\right) &= \sum_{j=0}^{n-1}\alpha_j V A^j y \\
&= \sum_{j=0}^{n-1}\alpha_j\left(N^j V y - \sum_{k=1}^{j}\lambda_k N^{j-k}x\right) \\
&= p_1(N)Vy - p_2(N)x,
\end{aligned}
$$

where p_1 and p_2 are the polynomials $p_1(z) = \sum_{j=0}^{n-1}\alpha_j z^j$ and $p_2(z) = \sum_{j=1}^{n-1}\sum_{k=1}^{j}\alpha_j\lambda_k z^{j-k}$, respectively, we have

$$p_1(a_1)\frac{\overline{x_2}}{(|x_1|^2+|x_2|^2)^{1/2}} = p_2(a_1)x_1, \tag{4}$$

$$p_1(a_2)\frac{-\overline{x_1}}{(|x_1|^2+|x_2|^2)^{1/2}} = p_2(a_2)x_2 \tag{5}$$

and

$$p_2(a_l)x_l = 0, \quad 3 \leqslant l \leqslant n+1. \tag{6}$$

Since $x_l \neq 0$ for all l and p_2 is a polynomial of degree at most $n-2$, the equalities in (6) yield

$$p_2(z) = \sum_{j=0}^{n-2}\left(\sum_{k=j+1}^{n-1}\alpha_k\lambda_{k-j}\right)z^j \equiv 0.$$

Hence $\sum_{k=j+1}^{n-1}\alpha_k\lambda_{k-j} = 0$ for all $j, 0 \leqslant j \leqslant n-2$. For $j = n-2$, this gives $\alpha_{n-1}\lambda_1 = 0$. Note also that

$$\lambda_1 = \langle NVy, x\rangle = \frac{a_1\overline{x_2x_1} - a_2\overline{x_1x_2}}{(|x_1|^2+|x_2|^2)^{1/2}} \neq 0$$

since the a_js are distinct and the x_js are nonzero. Therefore, we have $\alpha_{n-1} = 0$. Proceeding successively from $j = n-3$ to $j = 0$, we obtain $\alpha_{n-2} = \cdots = \alpha_1 = 0$. Thus $p_1(z) = \alpha_0$. From (4) or (5), we derive that $\alpha_0 = 0$. Hence y is a cyclic vector of A. Notice that $A^* = V^*N^*V$ is also an $\mathcal{N}_n$-matrix with the same vector x. The above arguments show that y is also cyclic for A^*. This completes the proof. □

The following is a slight generalization of Theorem 3.2.

Theorem 3.3. *If $N = \mathrm{diag}(a_1, \ldots, a_{n+1})$ is a diagonal matrix as in* (1)*, then any n-by-n compression of N is cyclic.*

Proof. Let $A = V^*NV$, where V is an $(n+1)$-by-n matrix with $V^*V = I_n$, and let x be a nonzero vector in $\mathbb{C}^{n+1}$ which is orthogonal to the range space of V. For

convenience, we may assume that x is of the form $[x_1 \ \cdots \ x_m \ 0 \ \cdots \ 0]^{\mathrm{T}}$ with $x_j \neq 0$ for $1 \leqslant j \leqslant m$. Then A is unitarily equivalent to the direct sum of an $\mathcal{N}_m$-matrix A' and the diagonal matrix $N' = \mathrm{diag}(a_{m+1}, \ldots, a_{n+1})$. Since A' and N' are both cyclic (the former by Theorem 3.2) and have no common eigenvalue, we infer that $A' \oplus N'$, and hence A, is also cyclic. □

Note that the preceding theorem is false for a general normal matrix N. For example, if $N = \mathrm{diag}(0, 1, 2)$ and $A = I_2$, then A is a compression of N but is not cyclic.

4. Dilation

The main result of this section is the following theorem relating general compressions of a normal matrix to $\mathcal{N}_n$-matrices via the tangency property of their numerical ranges. This line of investigation is taken in attempt to understand what [3, Theorem 6] really means.

Theorem 4.1. *Let $N = \mathrm{diag}(a_1, \ldots, a_{n+1})$ be a diagonal matrix as in* (1)*, and let A be an m-by-m compression of N $(1 \leqslant m \leqslant n)$.*

(a) *Assume that $\partial W(A)$ is tangent to l edges of the $(n+1)$-gon $a_1 \cdots a_{n+1}$, each of them at exactly one point different from the vertices $a_1, \ldots, a_{n+1}$. Then $l \leqslant m$, if $m < n$, and $l = n + 1$ if $m = n$.*
(b) *Consider k such tangent points $(1 \leqslant k \leqslant \min\{l, n-1\})$. If these k points are on* (*resp.*, *not on*) *successive edges of the $(n+1)$-gon $a_1 \cdots a_{n+1}$, then there is a k-by-k matrix B which dilates to A and is of class $\mathcal{N}_k$* (*resp.*, *is reducible*) *such that $\partial W(B)$ is tangent to the edges of $a_1 \cdots a_{n+1}$ at exactly these k points. In the latter case, $\partial W(B)$ contains a line segment.*

A special case of (b) above, first shown to the second author by Choi, is that if A is a 2-by-2 matrix which dilates to $N = \mathrm{diag}(1, \mathrm{i}, -1, -\mathrm{i})$ with $\partial W(A)$ tangent to the square formed by 1, i, -1, and $-\mathrm{i}$ at exactly two opposite edges, then A must be normal. Our theorem can be seen as an elaborate generalization of this.

Proof of Theorem 4.1. Assume that $A = V^*NV$, where V is an $(n+1)$-by-m matrix with $V^*V = I_m$. Let y be a unit vector in $\mathbb{C}^m$ such that $\langle Ay, y\rangle$ is the tangent point of some (open) edge (a_j, a_{j+1}) of $a_1 \cdots a_{n+1}$ with $\partial W(A)$. If $Vy = [u_1 \ \cdots \ u_{n+1}]^{\mathrm{T}}$, then

$$\langle Ay, y\rangle = \langle NVy, Vy\rangle = \sum_{k=1}^{n+1} |u_k|^2 a_k.$$

Since $\langle Ay, y\rangle$ belongs to the edge (a_j, a_{j+1}) of $a_1 \cdots a_{n+1}$, the above expression implies that $u_k = 0$ for all $k \neq j, j+1$, and $u_j, u_{j+1} \neq 0$. Hence if $q = \min\{l, n\}$

and $y_1, \ldots, y_l$ are unit vectors in $\mathbb{C}^m$ such that the $\langle Ay_j, y_j\rangle$s are the tangent points of the edges of $a_1 \cdots a_{n+1}$ with $\partial W(A)$, then any q vectors among the y_js are linearly independent.

(a) If $m < n$, then it follows from the above discussion that $l \leqslant m$. On the other hand, if $m = n$, then, letting $x = [x_1 \ \cdots \ x_{n+1}]^{\mathrm{T}}$ be any nonzero vector orthogonal to the range space K of V, we claim that all x_js are nonzero. Indeed, if some $x_j = 0$, then e_j, being orthogonal to x, is in K. Hence $e_j = Vu$ for some unit vector u in $\mathbb{C}^m$. We have

$$\langle Au, u\rangle = \langle NVu, Vu\rangle = \langle Ne_j, e_j\rangle = a_j,$$

which shows that a_j lies in $W(A)$, contradicting our assumption on $\partial W(A)$. Thus A is in class $\mathcal{N}_n$. In this case, all the $n+1$ edges of $a_1 \cdots a_{n+1}$ are tangent to $\partial W(A)$ by [9, Theorem 3]. Hence $l = n+1$ as asserted.

(b) If the k points are on successive edges, then, for simplicity, we may assume that these are $(a_1, a_2), \ldots, (a_k, a_{k+1})$ and that $\langle Ay_j, y_j\rangle$ is in (a_j, a_{j+1}), $1 \leqslant j \leqslant k$. As shown before, the y_js are linearly independent. Hence the subspace M of $\mathbb{C}^m$ generated by the y_js is of dimension k. Let V_1 be some m-by-k matrix of the inclusion map from M into $\mathbb{C}^m$, and let $B = V_1^* A V_1$. As each Vy_j is a vector in $\mathbb{C}^{n+1}$ with only its jth and $(j+1)$st components nonzero, VM is in fact contained in the subspace $\mathbb{C}^{k+1} \oplus \{0\}$ of $\mathbb{C}^{n+1}$. Thus

$$B = (VV_1)^* N (VV_1) = V_2^* N' V_2$$

for $N' = \mathrm{diag}(a_1, \ldots, a_{k+1})$ and some $(k+1)$-by-k matrix V_2 with $V_2^* V_2 = I_k$. To show that B is of class $\mathcal{N}_k$, we let $x = [x_1 \ \cdots \ x_{k+1}]^{\mathrm{T}}$ be any nonzero vector in $\mathbb{C}^{k+1}$ which is orthogonal to the range space of V_2. If any component x_j is zero, then, as shown in (a), a_j is in $W(B)$ and hence in $W(A)$, contradicting our assumption on $\partial W(A)$. Thus B is indeed an $\mathcal{N}_k$-matrix. The statement on $\partial W(B)$ is trivial.

On the other hand, if the k points are not on successive edges, then we may assume for simplicity that these are $(a_1, a_2), \ldots, (a_l, a_{l+1}), (a_{l+2}, a_{l+3}), \ldots, (a_{k+1}, a_{k+2})$ with $a_{k+2} \not\equiv a_1$, and that $\langle Ay_j, y_j\rangle$ is in (a_j, a_{j+1}) for $1 \leqslant j \leqslant l$, and in (a_{j+1}, a_{j+2}) for $l+1 \leqslant j \leqslant k$. Let M (resp., M_1 and M_2) denote the subspace of $\mathbb{C}^m$ generated by the y_js (resp., $y_1, \ldots, y_l$ and $y_{l+1}, \ldots, y_k$), and let V (resp., V_1 and V_2) be some m-by-k (resp., m-by-l and m-by-$(k-l)$) matrix of the inclusion map from M (resp., M_1 and M_2) into $\mathbb{C}^m$. Then, as proved before, $B = V^* A V$ (resp., $B_1 = V_1^* A V_1$ and $B_2 = V_2^* A V_2$) is a k-by-k matrix (resp., an $\mathcal{N}_l$-matrix and $\mathcal{N}_{k-l}$-matrix). Moreover, for any y_i $(1 \leqslant i \leqslant l)$ and y_j $(l+1 \leqslant j \leqslant k)$, we have

$$\langle y_i, y_j\rangle = \langle Vy_i, Vy_j\rangle = 0$$

and

$$\langle Ay_i, y_j\rangle = \langle NVy_i, Vy_j\rangle = 0$$

by the special structure of the components of Vy_i and Vy_j. We conclude that B is unitarily equivalent to the direct sum $B_1 \oplus B_2$ and hence is reducible. In addition, it is easy to see that $W(B_1)$ (resp., $W(B_2)$) is contained in the convex hull

of $\{a_1, \ldots, a_{l+1}\}$ (resp., $\{a_{l+2}, \ldots, a_{k+2}\}$). Thus $W(B)$, being the convex hull of $W(B_1) \cup W(B_2)$, must contain at least one line segment in its boundary. This completes the proof. □

The next corollary is an easy consequence. Its sufficiency gives a clear geometric illustration of [3, Theorem 6].

Corollary 4.2. *Let $A = V^*NV$ be an $\mathcal{N}_n$-matrix as given by* (1) *and* (2), *and let c_j, $1 \leqslant j \leqslant n+1$, be the tangent point of the edge $[a_j, a_{j+1}]$ of the $(n+1)$-gon $a_1 \cdots a_{n+1}$ with $\partial W(A)$ ($a_{n+2} \equiv a_1$). Then for any fixed k points ($1 \leqslant k < n$) among the c_js, there is an $\mathcal{N}_k$-matrix B which dilates to A such that the tangent points of the edges of $a_1 \cdots a_{n+1}$ with $\partial W(B)$ are exactly these points if and only if they are on successive edges of $a_1 \cdots a_{n+1}$. In this case, B is unique up to unitary equivalence.*

The uniqueness of B here follows from [9, Theorems 3 and 4].

We now give an illustrative example for Theorem 4.1 and Corollary 4.2.

Example 4.3. Let $N = \operatorname{diag}(1, \omega, \omega^2, -1, \omega^3, \omega^4)$, where ω is the fifth primitive root of unity, and let

$$A = \begin{bmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & 0 & 1 & & \\ & & & 0 & \sqrt{1/6} & \\ \sqrt{1/6} & & & & -5/6 \end{bmatrix}.$$

Then A dilates to

$$N' = \begin{bmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & 0 & 1 & & \\ & & & 0 & \sqrt{1/6} & -\sqrt{5/6} \\ \sqrt{1/6} & & & & -5/6 & -\sqrt{5}/6 \\ -\sqrt{5/6} & & & & -\sqrt{5}/6 & -1/6 \end{bmatrix}.$$

It is easily seen that N' is unitary and has characteristic polynomial $p_{N'}(z) = (z^5 - 1)(z + 1)$. Therefore N' is unitarily equivalent to N and thus A dilates to N. Since the characteristic polynomial p_A of A is $z^5 + (5/6)z^4 - (1/6)$, we have that A has no eigenvalue of modulus one. Together with $\operatorname{rank}(I_5 - A^*A) = 1$, this implies that A is of class $\mathcal{N}_5$ (or even $\mathcal{S}_5$). Since p_A is a scalar multiple of the derivative of p_N, we obtain from Theorem 2.3 that the boundary of $W(A)$ is tangent to the edge $[a_j, a_{j+1}]$ of the 6-gon G formed by $a_1 = 1, a_2 = \omega, a_3 = \omega^2, a_4 = -1, a_5 = \omega^3$ and $a_6 = \omega^4$ at its midpoint $c_j = (a_j + a_{j+1})/2$, $1 \leqslant j \leqslant 6$ ($a_7 \equiv a_1$). For the successive c_5, c_6, c_1 and c_2, the $\mathcal{N}_4$-matrix

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

dilates to A and has $\partial W(B)$ tangent to the edges of G at exactly these four points. On the other hand, for the nonsuccessive c_1, c_2, c_4 and c_5, consider the reducible matrix $C = C_1 \oplus C_2$, where

$$C_1 = \begin{bmatrix} \alpha_1 & \beta_1 \\ 0 & \overline{\alpha_1} \end{bmatrix}, \quad C_2 = \begin{bmatrix} \alpha_2 & \beta_2 \\ 0 & \overline{\alpha_2} \end{bmatrix},$$

$\alpha_1 = (\omega^2 + \omega + 1 + (\omega^4 - \omega^3 - \omega + 1)^{1/2})/3$, $\beta_1 = 1 - |\alpha_1|^2$, $\alpha_2 = (\omega^4 + \omega^3 - 1 + \omega(\omega - 2)^{1/2})/3$, and $\beta_2 = 1 - |\alpha_2|^2$. Since α_1 and $\overline{\alpha_1}$ (resp., α_2 and $\overline{\alpha_2}$) are zeroes of the derivative of the polynomial $(z-1)(z-\omega)(z-\omega^2)$ (resp., $(z+1)(z-\omega^3)(z-\omega^4)$), the boundary $\partial W(C_1)$ (resp., $\partial W(C_2)$) is tangent to the edges of G at c_1 and c_2 (resp., c_4 and c_5) (by Theorem 2.3 again). That C dilates to A follows as in the proof of Theorem 4.1(b).

This example is illustrated in Figs. 1 and 2.

Note that in general an $\mathcal{N}_m$-matrix B which dilates to another $\mathcal{N}_n$-matrix A may have its numerical range $W(B)$ disjoint from the boundary of $W(A)$. For example, this is the case with

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

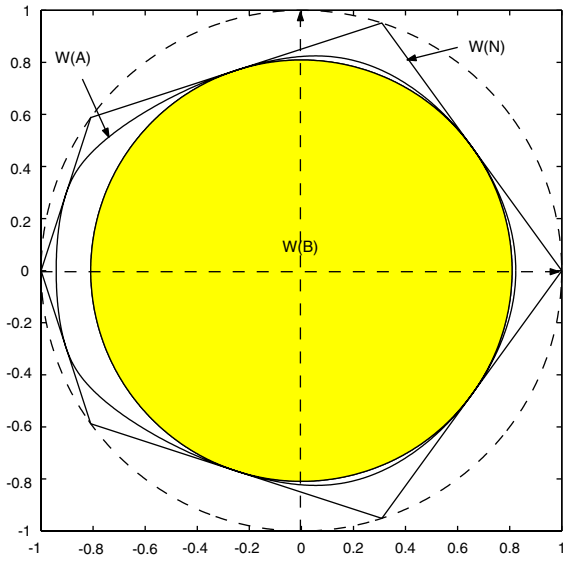

Fig. 1.

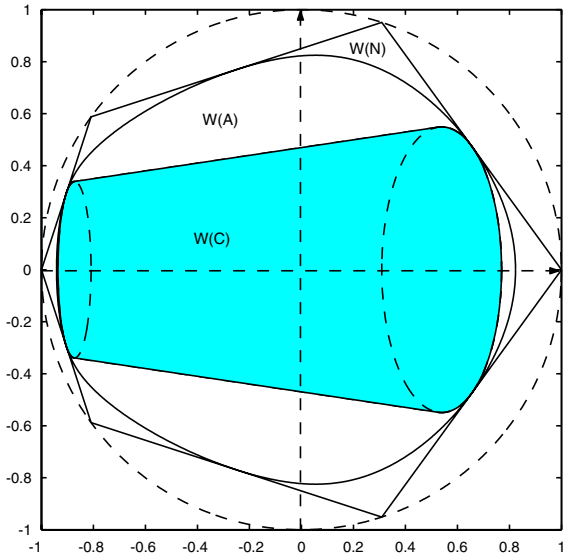

Fig. 2.

It is known more generally that if A and B are $\mathscr{S}_n$- and $\mathscr{S}_m$-matrices, respectively, such that A is unitarily equivalent to a matrix of the form $\left[\begin{smallmatrix} B & * \\ 0 & * \end{smallmatrix}\right]$, then we always have $W(B) \cap \partial W(A) = \emptyset$ (see [6, Corollary 3.4]). Whether the same can be said about $\mathscr{N}_n$- and $\mathscr{N}_m$-matrices seems to be unknown. We end this section with an example showing that even when A is not unitarily equivalent to $\left[\begin{smallmatrix} B & * \\ 0 & * \end{smallmatrix}\right]$ but only a dilation of B, the disjointness of $W(B)$ and $\partial W(A)$ can still happen.

Example 4.4. Let

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & a & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},$$

where $|a| < 1$. If $a = 0$, then $A = \left[\begin{smallmatrix} B & * \\ 0 & * \end{smallmatrix}\right]$, which is the case discussed above. It can be easily seen that A is a contraction ($\|A\| \leqslant 1$), has all its eigenvalues in the open unit disc and satisfies $\operatorname{rank}(I_3 - A^*A) = 1$. Hence A is an $\mathscr{S}_3$-matrix by [10, Lemma 2.2]. We now show that $W(B) \cap \partial W(A) = \emptyset$. Indeed, if c is a point in $W(B) \cap \partial W(A)$, then, since A, ωA and $\omega^2 A$ are mutually unitarily equivalent for $\omega = (-1 + \sqrt{3}\mathrm{i})/2$ and $W(B) = \{z \in \mathbb{C} : |z| \leqslant 1/2\}$, the two points ωc and $\omega^2 c$ also lies in $W(B) \cap \partial W(A)$. Note that the tangent line L_1 (resp., L_2 and L_3) to $\partial W(A)$ at c (resp., ωc and $\omega^2 c$) is also tangent to the circle $\partial W(B)$. Hence the L_js intersect, say, at the points a_1, a_2 and a_3 on the unit circle. Thus $W(A)$ is contained in the 3-gon $a_1a_2a_3$ contradicting [5, Corollary 2.5]. This proves that $W(B) \cap \partial W(A) = \emptyset$.

References

[1] M. Adam, J. Maroulas, On compressions of normal matrices, Linear Algebra Appl. 341 (2002) 403–418.
[2] H. Bercovici, Numerical ranges of operators of class C_0, Linear and Multilinear Algebra 50 (2002) 219–222.
[3] U. Daepp, P. Gorkin, R. Mortini, Ellipses and finite Blaschke products, Amer. Math. Monthly 109 (2002) 785–795.
[4] M.R. Embry, The numerical range of an operator, Pacific J. Math. 32 (1970) 647–650.
[5] H.-L. Gau, P.Y. Wu, Numerical range of $S(\phi)$, Linear and Multilinear Algebra 45 (1998) 49–73.
[6] H.-L. Gau, P.Y. Wu, Dilation to inflations of $S(\phi)$, Linear and Multilinear Algebra 45 (1998) 109–123.
[7] H.-L. Gau, P.Y. Wu, Lucas' theorem refined, Linear and Multilinear Algebra 45 (1998) 359–373.
[8] H.-L. Gau, P.Y. Wu, Numerical range and Poncelet property, Taiwanese J. Math. 7 (2003) 173–193.
[9] H.-L. Gau, P.Y. Wu, Numerical range of a normal compression, Linear and Multilinear Algebra 52 (2004) 195–201.
[10] H.-L. Gau, P.Y. Wu, Numerical range circumscribed by two polygons, Linear Algebra Appl. 382 (2004) 155–170.
[11] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge, 1991.
[12] B. Mirman, Numerical ranges and Poncelet curves, Linear Algebra Appl. 281 (1998) 59–85.
[13] B. Mirman, UB-matrices and conditions for Poncelet polygon to be closed, Linear Algebra Appl. 360 (2003) 123–150.
[14] B. Mirman, V. Borovikov, L. Ladyzhensky, R. Vinograd, Numerical ranges, Poncelet curves, invariant measures, Linear Algebra Appl. 329 (2001) 61–75.
[15] B. Mirman, P.Y. Wu, Matrix interpretation of Marden's proof of Siebeck's theorem, preprint.
[16] P.Y. Wu, Polygons and numerical ranges, Amer. Math. Monthly 107 (2000) 528–540.