Linear Algebra and its Applications 350 (2002) 1–23

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

The numerical range of a nonnegative matrix

Chi-Kwong Li [a,1], Bit-Shun Tam [b,*,2], Pei Yuan Wu [c,2]

[a]Department of Mathematics, College of William and Mary, Williamsburg, VA 23187, USA
[b]Department of Mathematics, Tamkang University, Tamsui, 25137 Taiwan, ROC
[c]Department of Applied Mathematics, National Chiao Tung University, Hsinchu, 30050 Taiwan, ROC

Received 30 March 2001; accepted 10 November 2001

Submitted by H. Schneider

Abstract

We offer an almost self-contained development of Perron–Frobenius type results for the numerical range of an (irreducible) nonnegative matrix, rederiving and completing the previous work of Issos, Nylen and Tam, and Tam and Yang on this topic. We solve the open problem of characterizing nonnegative matrices whose numerical ranges are regular convex polygons with center at the origin. Some related results are obtained and some open problems are also posed. © 2002 Elsevier Science Inc. All rights reserved.

AMS classification: 15A60; 15A48

Keywords: Numerical range; Nonnegative matrix; Irreducible real part; Numerical radius; Perron–Frobenius theory; Regular polygon; Sharp point

1. Introduction

By the classical Perron–Frobenius theory, if A is a (square, entrywise) nonnegative matrix, then its spectral radius $\rho(A)$ is an eigenvalue of A and there is a corresponding nonnegative eigenvector. If, in addition, A is irreducible, then $\rho(A)$ is a simple eigenvalue and the corresponding eigenvector can be chosen to be positive. Moreover, for an irreducible nonnegative matrix with index of imprimitivity $m > 1$ (i.e.,

* Corresponding author. Tel.: +886-2-26215656; fax: +886-2-26209916.
E-mail address: bsm01@mail.tku.edu.tw (B.-S. Tam).

[1] Research partially supported by a USA NSF grant.
[2] Supported by the National Science Council of the Republic of China.

PII: S0024-3795(02)00291-4

one having exactly m eigenvalues with modulus $\rho(A)$), Frobenius has also obtained a deeper structure theorem: The set of eigenvalues of A with modulus $\rho(A)$ consists precisely of $\rho(A)$ times all the mth roots of unity, the spectrum $\sigma(A)$ of A is invariant under a rotation about the origin of the complex plane through an angle of $2\pi/m$, and A is an m-cyclic matrix, i.e., there is a permutation matrix P such that $P^t AP$ is a matrix of the form

$$\begin{bmatrix} \mathbf{0} & A_{12} & & & \\ & \mathbf{0} & A_{23} & & \\ & & \ddots & \ddots & \\ & & & \mathbf{0} & A_{m-1,m} \\ A_{m1} & & & & \mathbf{0} \end{bmatrix}, \tag{1.1}$$

where the zero blocks along the diagonal are all square. To establish the above structure theorem, the most popular proof is to use Wielandt's lemma. (Some relevant definitions and a full statement of Wielandt's lemma will be given later on. For the logical relations between the various conditions that appear in the above-mentioned Frobenius's result, in the setting of a complex matrix, see the recent paper [15] by the second author.)

By now, the Perron–Frobenius theory of a nonnegative matrix is very well known. Almost every textbook of matrix theory contains a chapter on the subject, and there are also monographs specially devoted to nonnegative matrices and their applications. On the other hand, there is not much literature on the numerical range of a nonnegative matrix, although it is well known that the numerical range of a matrix and its spectral properties are related. As a matter of fact, as early as 1966, Issos [10] in his unpublished Ph.D. thesis has obtained some Perron–Frobenius type results on the numerical range of an irreducible nonnegative matrix. However, for many years, except for a reference by Fiedler [5], it appears that Issos's work was almost unnoticed. Recall that the numerical range of an n-by-n complex matrix A is denoted and defined by

$$W(A) = \left\{x^*Ax : x \in \mathbb{C}^n, x^*x = 1\right\}.$$

Here is the main result obtained by Issos [10, Theorem 7]:

Theorem 1.1. *Let A be an irreducible nonnegative matrix with index of imprimitivity m. Denote the numerical radius of A by $w(A)$. Then*

$$\left\{\lambda \in W(A) : |\lambda| = w(A)\right\} = \left\{w(A)e^{2\pi ti/m} : t = 0, 1, \ldots, m-1\right\}.$$

Issos's proof depends on a number of auxiliary results and is rather tedious. Recently, the second author and Yang [16] also obtained Issos's main result as a side-product of their treatment. The proof given in [16, Corollary 2] for Issos's result may not be easily accessible to the general readers. This is because the proof is indirect, graph-theoretic, and depends on results from the previous paper [15] of the second author, on the less well-known concepts of the signed length of a cycle (which is different from that of a circuit) and matrix cycle products, and also on a characterization of diagonal similarity between matrices in terms of matrix cycle products due to Saunders and Schneider. (Indeed, it is a purpose of the papers [15,16] to demonstrate the usefulness of these less well-known concepts and the characterization of Saunders and Schneider.) This research was initiated by our attempt to find a direct, self-contained proof of Issos's main result. In a graph-free manner, we are able to do this and also obtain an extension of the result in the setting of a nonnegative matrix with irreducible real part. Then, in terms of certain graph-theoretic concepts, we put the latter result in a more concrete usable form, depending our proof on some results of [14,15].

Since the literature on numerical range analogs of the Perron–Frobenius theory is scanty, we also think it is worthwhile to offer a complete and, as far as possible, self-contained development here.

The results we obtain also enable us to solve the open problem of characterizing nonnegative matrices whose numerical ranges are regular polygons with center at the origin. We treat this problem and related problems in the second half of the paper.

2. Preliminaries

We assume knowledge of the Perron–Frobenius theory of nonnegative matrices, which is available in many standard textbooks such as [1,8], or [11], as well as familiarity with numerical ranges (see, for instance, [6] or [9]).

Below we give a list of notations which we will follow. We always use A to denote an n-by-n complex matrix for some fixed positive integer n.

M_n	the set of all n-by-n complex matrices
$\mathbb{R}^n_+$	the nonnegative orthant of $\mathbb{R}^n$
$W(A)$	the (classical) numerical range of A
$w(A)$	the numerical radius of A
$\sigma(A)$	the spectrum of A
$\rho(A)$	the spectral radius of A
A^{t}	the transpose of A
A^*	the conjugate transpose of A
$\operatorname{Re} A$	the real part of A, i.e., $(A + A^*)/2$
$\|A\|$	the matrix $(\|a_{rs}\|)$ (where $A = (a_{rs})$)
$\operatorname{Re} z$	the real part of z (where z is a complex number)
$\|x\|$	the vector $(\|\xi_1\|, \ldots, \|\xi_n\|)^{\mathrm{t}}$ (where $x = (\xi_1, \ldots, \xi_n)^{\mathrm{t}}$)
$\lambda_{\max}(H)$	the largest eigenvalue of H (where H is hermitian)
i	the imaginary unit $\sqrt{-1}$
$\langle n \rangle$	the set $\{1, 2, \ldots, n\}$.

For a vector $x \in \mathbb{C}^n$, we use $\|x\|$ to denote the Euclidean norm of x, i.e., $\|x\| = (x^*x)^{1/2}$. For a matrix A, we use $\|A\|$ to denote the operator norm of A, i.e., $\|A\| = \max_{\|x\|=1} \|Ax\|$.

For real matrices A, B of the same size, we use $A \geqslant B$ (respectively, $A > B$) to mean $a_{rs} \geqslant b_{rs}$ (respectively, $a_{rs} > b_{rs}$) for all indices r, s. The notation will also apply to vectors.

We call a matrix $A \in M_n$ *irreducible* if $n = 1$ or $n \geqslant 2$ and there does not exist a permutation matrix P such that

$$P^{\mathrm{t}}AP = \begin{bmatrix} B & C \\ O & D \end{bmatrix},$$

where B, D are nonempty square matrices.

Given $A, B \in M_n$, A is said to be *diagonally similar* to B if there exists a nonsingular diagonal matrix D such that $A = D^{-1}BD$; if, in addition, D can be chosen to be unitary, then we say A is *unitarily diagonally similar* to B.

It is known [16, Remarks 2 and 5] and not difficult to show the following:

Remark 2.1. For any $A \in M_n$ and any unit complex number ξ, we have:

(i) A is unitarily diagonally similar to ξA if and only if A is diagonally similar to ξA;

(ii) $\operatorname{Re} A$ is unitarily diagonally similar to $\operatorname{Re}(\xi A)$ if and only if $\operatorname{Re} A$ is diagonally similar to $\operatorname{Re}(\xi A)$.

For graph-theoretic definitions, we follow those of [15,16]. We need, in particular, the concepts of cyclic index of a matrix, a cycle in a digraph, and the signed length of a cycle, which we are going to explain.

For any $A \in M_n$, as usual, by the *digraph of A*, denoted by $G(A)$, we mean the directed graph with vertex set $\langle n \rangle$ such that (r, s) is an arc if and only if $a_{rs} \neq 0$. By the *undirected graph of A* we mean the undirected graph obtained from $G(A)$ by removing the direction of its arcs. We call an undirected graph *connected* if either it has exactly one vertex or it has more than one vertex and every pair of distinct vertices are joined by a path.

It is well known (see, for instance, [8, Theorem 6.2.24]) that a matrix $A \in M_n$ is irreducible if and only if its digraph $G(A)$ is strongly connected (in the sense that given any two vertices r, s of $G(A)$, there is a directed path in $G(A)$ from r to s and vice versa). It is not difficult to show the following:

Remark 2.2. For any $A \in M_n$, if $\operatorname{Re} A$ is irreducible, then the undirected graph of A is connected. The converse also holds if A is nonnegative.

We call a matrix $A \in M_n$ *m-cyclic* if there exists a permutation matrix P such that $P^{\mathrm{t}}AP$ is of the form (1.1) where the zero blocks along the diagonal are all square. The largest positive integer m for which a matrix A is m-cyclic is called the *cyclic index* of A. We call A a *block-shift matrix* if for some integer $m \geqslant 2$, A is of the form (1.1) and with $A_{m1} = 0$. An m-cyclic matrix (respectively, a matrix which is permutationally similar to a block-shift matrix) can be characterized as one whose digraph is cyclically m-partite (respectively, linearly partite) (see [15] for definitions).

We reserve the term "circuit" (in a digraph) for its usual meaning, i.e., a simple closed directed path. For instance, a sequence of arcs, like (1,2), (2,3), (3,4), and (4,1), forms a circuit of length 4. The term "cycle" in our usage means something different. For example, a sequence of arcs, like

$$1 \longrightarrow 2 \longrightarrow 3 \longleftarrow 4 \longrightarrow 5 \longleftarrow 1,$$

forms a cycle of length 5 and signed length 1. Here we use $r \longrightarrow s$ to denote the arc (r, s) traversed from r to s and referred to it as a positive link, and use $s \longleftarrow r$ to denote the arc (r, s) traversed from s to r and referred to it as a negative link. The number of positive links minus the number of negative links in a cycle gives the signed length of the cycle. (For formal definitions, see [15, Section 2].)

3. Numerical range analogs of the Perron–Frobenius theory

Let A be an n-by-n nonnegative matrix. In parallel to the Perron–Frobenius theory, it is natural to assert that $w(A) \in W(A)$ and there is a unit nonnegative vector x such that $x^*Ax = w(A)$. The assertion is, indeed, true and is also pretty obvious. The reason is, for any unit vector $z \in \mathbb{C}^n$, we have $|z^*Az| \leqslant |z|^*A|z|$; hence

$$w(A) = \sup\left\{|z^*Az| : \|z\| = 1\right\} = \sup\left\{y^{\mathrm{t}}Ay : \|y\| = 1,\ y \in \mathbb{R}^n_+\right\}.$$

Clearly, the continuous real-valued map $y \mapsto y^{\mathrm{t}}Ay$ attains its maximum on the intersection of the unit sphere with the nonnegative orthant $\mathbb{R}^n_+$, which is a compact set. Hence our assertion follows. (The foregoing discussion has essentially appeared in [10, Theorem 1 and its proof], where it is assumed that the matrix A is irreducible nonnegative.)

The next thing one may try to prove is that, if A is irreducible nonnegative, then there is a positive unit vector x such that $x^*Ax = w(A)$, and furthermore x is unique. It is desirable that we can somehow apply the Perron–Frobenius theory. The following general result enables us to do this.

Lemma 3.1. *Let $A \in M_n$ and let ξ be a unit complex number such that $\xi w(A) \in W(A)$. Then*:

(i) $\lambda_{\max}(\mathrm{Re}(\bar{\xi}A)) = \rho(\mathrm{Re}(\bar{\xi}A)) = w(A)$;

(ii) *the set* $V = \{x \in \mathbb{C}^n : x^*Ax = \xi w(A)\|x\|^2\}$ *is equal to the eigenspace of* $\mathrm{Re}(\bar{\xi}A)$ *corresponding to* $\lambda_{\max}(\mathrm{Re}(\bar{\xi}A))$.

Proof. (i) Since $\xi w(A) \in W(A)$, we can find a nonzero vector u that satisfies $u^*Au = \xi w(A)\|u\|^2$. Then $u^*(\bar{\xi}A)u = w(A)\|u\|^2$, and so $u^*(\xi A^*)u = w(A)\|u\|^2$. Adding the two equations, we obtain $u^*\mathrm{Re}(\bar{\xi}A)u = w(A)\|u\|^2$ or $u^*(w(A)I_n - \mathrm{Re}(\bar{\xi}A))u = 0$. Note that the matrix $w(A)I_n - \mathrm{Re}(\bar{\xi}A)$ is positive semidefinite, as we have

$$w(A) = w(\bar{\xi} A) \geqslant w(\mathrm{Re}(\bar{\xi} A)) = \rho(\mathrm{Re}(\bar{\xi} A)) \geqslant \lambda_{\max}(\mathrm{Re}(\bar{\xi} A)).$$

Hence, u is an eigenvector of $\mathrm{Re}(\bar{\xi} A)$ corresponding to $w(A)$, and also it follows that we have $w(A) = \rho(\mathrm{Re}(\bar{\xi} A)) = \lambda_{\max}(\mathrm{Re}(\bar{\xi} A))$.

(ii) If x is any nonzero vector in V, then by what we have done in the proof of part (i) (with x in place of u), we see that x is an eigenvector of $\mathrm{Re}(\bar{\xi} A)$ corresponding to $\lambda_{\max}(\mathrm{Re}(\bar{\xi} A))$.

Conversely, if x is an eigenvector of $\mathrm{Re}(\bar{\xi} A)$ corresponding to $\lambda_{\max}(\mathrm{Re}(\bar{\xi} A))$ ($= w(A)$), then we have

$$w(A)\|x\|^2 = x^*\mathrm{Re}(\bar{\xi} A)x = \mathrm{Re}(x^*(\bar{\xi} A)x) \leqslant |x^*(\bar{\xi} A)x| \leqslant w(A)\|x\|^2,$$

and hence $x^*(\bar{\xi} A)x = w(A)\|x\|^2$, i.e., $x \in V$. □

Lemma 3.1(ii) is well known to researchers of numerical range. For example, it is essentially contained in [3, Corollary 1.4], and is also partly a consequence of the following result in [4] (see also [6, Theorems 1.5–1 and 1.5–2]:

A point $\alpha \in W(A)$ is an extreme point if and only if the associated subset $\{x \in \mathbb{C}^n : x^*Ax = \alpha\|x\|^2\}$ is a linear subspace.

We take a digression here. By examining the above proof of Lemma 3.1 (or the proof of [3, Corollary 1.4]) carefully, one can see that our argument also shows the following:

Remark 3.2. Let $A \in M_n$. For any unit complex number ξ, the set $\{x \in \mathbb{C}^n : x^*Ax = \xi w(A)\|x\|^2\}$ is equal to the nullspace of $w(A)I_n - \mathrm{Re}(\bar{\xi} A)$. Consequently, $\xi w(A) \in W(A)$ if and only if $\det(w(A)I_n - \mathrm{Re}(\bar{\xi} A)) = 0$.

The last part of the above remark (the "only if" part of which is implicit in the proof of [16, Lemma 6]) enables us to check whether a given nonnegative matrix A with irreducible real part has a circular disk centered at origin as its numerical range, or whether it satisfies $e^{2\pi i/m}W(A) = W(A)$ for a given positive integer m. This is because, by [16, Theorems 1 and 2], for such a matrix A, $W(A)$ is a circular disk centered at the origin if and only if for some real number θ which is an irrational multiple of π or is a rational multiple of the form $2\pi p/q$, where $p,\ q$ are relatively prime integers with $q > n$, we have $e^{i\theta}w(A) \in W(A)$; $e^{2\pi i/m}W(A) = W(A)$ if and only if $e^{2\pi i/m}w(A) \in W(A)$.

Now back to numerical range analogs of the Perron–Frobenius theory. If A is a nonnegative matrix, we already know that $w(A) \in W(A)$. So in this case we can apply Lemma 3.1 to A by taking $\xi = 1$. Then we see that we have

$$\lambda_{\max}(\mathrm{Re}\,A) = \rho(\mathrm{Re}\,A) = w(A),$$

and the set $\{x \in \mathbb{C}^n : x^*Ax = w(A)\|x\|^2\}$ is equal to the eigenspace of the nonnegative matrix $\mathrm{Re}\,A$ corresponding to its spectral radius $\rho(\mathrm{Re}\,A)$. If, in addition, $\mathrm{Re}\,A$ is irreducible (which is the case if A is irreducible), then by the Perron–Frobenius theory, $\rho(\mathrm{Re}\,A)$ is a simple eigenvalue of $\mathrm{Re}\,A$ and the said subspace is of dimension 1, spanned by a positive vector.

Summarizing, we have obtained the following:

Proposition 3.3. *Let $A \in M_n$ be nonnegative. Then $w(A) \in W(A)$ and each of the following numbers is equal to $w(A)$:*

$$\max\{y^{\mathrm{t}}Ay : y \in \mathbb{R}^n_+, \|y\| = 1\},\ \lambda_{\max}(\mathrm{Re}\,A),\ \textit{and}\ \rho(\mathrm{Re}\,A).$$

Moreover, there is a unit nonnegative vector x such that $x^*Ax = w(A)$. If, in addition, Re *A is irreducible, then the vector x is unique and is positive.*

The relation $w(A) = \rho(\mathrm{Re}\,A)$ for a nonnegative matrix A was shown in [7]. It also appeared implicitly in the proof of [10, Theorem 1].

We make another digression and take note of the following interesting consequence of the fact that $w(A) \in W(A)$ for a nonnegative matrix A:

Corollary 3.4. *Let $A = (a_{ij})$ be an n-by-n nonnegative matrix. If $W(A)$ is a (possibly degenerate) elliptic disk or a regular polygon, then the center p of $W(A)$ must be a nonnegative real number such that $p \geqslant \min_{1\leqslant j\leqslant n} a_{jj}$.*

Proof. Since A is a real matrix, $W(A)$ must be symmetric about the real axis (see, for instance, [12, Lemma 3.1]). So p must lie on the real axis. Let α denote $\min_{1\leqslant j\leqslant n} a_{jj}$, and assume to the contrary that $p < \alpha$. Then $A - \alpha I_n$ is still a nonnegative matrix and its numerical range is $W(A) - \alpha$, with center at $p - \alpha$, which is a negative number. If $W(A)$ is an elliptic disk, then, clearly, the left vertical supporting line for $W(A - \alpha I_n)$ is farther away from the origin than the right vertical supporting line. Hence, $w(A - \alpha I_n) \notin W(A - \alpha I_n)$, which contradicts the result of Proposition 3.3. On the other hand, if $W(A)$ is a regular polygon, then the distance from the origin to the vertices of $W(A - \alpha I_n)$ other than $w(A) - \alpha$ is greater than $w(A) - \alpha$, which again contradicts Proposition 3.3. □

The above corollary may suggest that, in general, if $A = (a_{ij})$ is an n-by-n nonnegative matrix, then the centroid p of $W(A)$ satisfies $p \geqslant \min_{1\leqslant j\leqslant n} a_{jj}$. Our next example will show that this is not true.

Example 3.5. Consider the nonnegative matrix

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Here $W(A)$ is the convex hull of the equilateral triangle with vertices 1, $\mathrm{e}^{2\pi\mathrm{i}/3}$ and $\mathrm{e}^{4\pi\mathrm{i}/3}$ and the line segment with endpoints 1 and -1. So we have $\min_{1\leqslant j\leqslant n} a_{jj} = 0 > p$, where p denotes the centroid of $W(A)$. By perturbing the above A a little bit, one can also give an irreducible nonnegative matrix as an example.

Next, we turn to a comparison of $w(A)$ and $w(B)$ for nonnegative matrices A and B with $A \geqslant B$. For spectral radius, the following is well known (see [11, p. 38, Corollary 2.2]):

Let $A, B \in M_n$ be nonnegative, and suppose that $B \leqslant A$. Then $\rho(B) \leqslant \rho(A)$. If, in addition, A is irreducible and $A \neq B$, then $\rho(B) < \rho(A)$.

Using the relation $w(A) = \rho(\operatorname{Re} A)$ for a nonnegative matrix A, we immediately obtain the following corresponding result for numerical radius. Below we also give an alternative short proof of the result.

Corollary 3.6. *Let $A, B \in M_n$ be nonnegative, and suppose that $B \leqslant A$. Then $w(B) \leqslant w(A)$. If, in addition,* $\operatorname{Re} A$ *is irreducible and $A \neq B$, then $w(B) < w(A)$.*

Proof. Since $0 \leqslant B \leqslant A$, we have

$$\begin{aligned} w(B) &= \max\left\{x^{\mathrm{t}}Bx : x \in \mathbb{R}^n_+,\ \|x\| = 1\right\} \\ &\leqslant \max\left\{x^{\mathrm{t}}Ax : x \in \mathbb{R}^n_+,\ \|x\| = 1\right\} \\ &= w(A). \end{aligned}$$

Now assume that $\operatorname{Re} A$ is irreducible, and suppose that $w(B) = w(A)$. Choose a nonnegative unit vector x such that $x^*Bx = w(B)$. Then we have

$$w(A) = w(B) = x^*Bx \leqslant x^*Ax \leqslant w(A).$$

Thus, the two inequalities become equalities. Since $\operatorname{Re} A$ is irreducible, by the last part of Proposition 3.3, the vector x is positive. So we have $x^*(A - B)x = 0$, and together with the assumption $A \geqslant B$, it follows that $A = B$, which is a contradiction. □

We want to emphasize that in the last part of Corollary 3.6 (also Proposition 3.3) we are assuming that $\operatorname{Re} A$ is irreducible instead of A being irreducible. And this is the right setting for results on numerical radius. For the corresponding results on spectral radius, we do need the irreducibility assumption. As an example, consider

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

Then $\operatorname{Re} A$ is irreducible, but not A, and we have $\rho(A) = \rho(B) = 0$, and $w(A) > w(B)$.

In the above we have treated the rudimentary part of the numerical range analogs of the Perron–Frobenius theory. To proceed further, we need the following Wielandt's lemma [17]:

Wielandt's lemma. *Let $A, B \in M_n$, and assume that A is nonnegative. If $|B| \leqslant A$, then $\rho(B) \leqslant \rho(A)$. Assume, in addition, that A is irreducible. If $\rho(A) = \rho(B)$ and*

ξ is a unit complex number such that $\xi\rho(B) \in \sigma(B)$, then $B = \xi DAD^{-1}$ for some (unitary) diagonal matrix D.

In the above formulation of Wielandt's lemma, in order to emphasize its nontrivial part, we have deliberately omitted the obvious converse part for its second half (cf. [11, p. 36, Theorem 2.1]).

Now consider an irreducible nonnegative matrix A with index of imprimitivity m. By the Perron–Frobenius theory, we have $e^{2\pi i/m}\rho(A) \in \sigma(A)$, and so by the second half of Wielandt's lemma (with $B = A$ and $\xi = e^{2\pi i/m}$), $e^{2\pi i/m}A$ is (unitarily) diagonally similar to A. But the numerical range of a matrix is unchanged if we apply a unitary similarity to the matrix, hence we have $e^{2\pi i/m}W(A) = W(e^{2\pi i/m}A) = W(A)$, i.e., $W(A)$ is invariant under a rotation about the origin of the complex plane through an angle of $2\pi/m$. But we also have $w(A) \in W(A)$, hence $w(A)e^{2\pi ti/m} \in W(A)$ for $t = 0, 1, \ldots, m-1$. This proves the easy half of Theorem A. (An argument almost the same as the preceding one can be found in [10, Theorems 6 and 7], except that Issos used the m-cyclicity of A to deduce the diagonal similarity between A and $e^{2\pi i/m}A$ instead of applying Wielandt's lemma.) To prove the reverse inclusion, we need the following:

Proposition 3.7. *Let $A \in M_n$ be nonnegative, and suppose* Re A *is irreducible. If ξ is a unit complex number such that $\xi w(A) \in W(A)$, then $DAD^{-1} = \xi A$ for some unitary diagonal matrix D.*

To see how Proposition 3.7 can be used to establish the remaining inclusion for Theorem A, consider any unit complex number ξ for which $\xi w(A) \in W(A)$. By the proposition, ξA is similar to A. But $\rho(A)$ is an eigenvalue of A, hence so is $\xi\rho(A)$. By the Frobenius theorem for an irreducible nonnegative matrix, it follows that ξ must be an mth root of unity, where m is the index of imprimitivity of A. This completes the proof of Theorem A.

Note that when A is an irreducible nonnegative matrix with index of imprimitivity 1 (or, equivalently, when it is a primitive matrix, i.e., a nonnegative matrix, one of whose powers is positive), Theorem A tells us that the numerical range of A contains exactly one point with modulus $w(A)$, namely, $w(A)$ itself. Our above proof also covers this special case.

Theorem A first appeared in [10, Theorem 7] and then in [16, Corollary 2]; whereas Proposition 3.7 is contained in [16, Lemma 1], but not in [10]. The proof given in [16] for Theorem A is longer than necessary; it makes use of [16, Lemma 1], but not in the best way. Graph-theoretic arguments as well as results from [15] are needed in [16] to establish its Lemma 1. We shall give two proofs for Proposition 3.7, which are self-contained and graph-free.

First proof of Proposition 3.7. Since A is nonnegative, we have $\rho(\mathrm{Re}\,A) = w(A)$. By Lemma 3.1(i), we also have $w(A) = \rho(\mathrm{Re}(\bar{\xi}A)) = \lambda_{\max}(\mathrm{Re}(\bar{\xi}A))$, and hence

$\rho(\mathrm{Re}(\bar{\xi}A)) = \rho(\mathrm{Re}\,A)$. From the above, it is also clear that $\rho(\mathrm{Re}(\bar{\xi}A))$ is an eigenvalue of $\mathrm{Re}(\bar{\xi}A)$. In view of $|\mathrm{Re}(\bar{\xi}A)| \leqslant \mathrm{Re}\,A$ and the irreducibility of $\mathrm{Re}\,A$, by the second half of Wielandt's lemma, it follows that there is a unitary diagonal matrix D, say $D = \mathrm{diag}(d_1, \ldots, d_n)$, such that $D(\mathrm{Re}(\bar{\xi}A))D^{-1} = \mathrm{Re}\,A$, i.e., $D(\bar{\xi}A + \xi A^{\mathrm{t}})D^{-1} = A + A^{\mathrm{t}}$. By equating the corresponding entries of both sides, we obtain $d_r(\bar{\xi}a_{rs} + \xi a_{sr})d_s^{-1} = a_{rs} + a_{sr}$ for all $r, s \in \langle n\rangle$. Since $|d_r\bar{\xi}a_{rs}d_s^{-1}| = a_{rs}$ and $|d_r\xi a_{sr}d_s^{-1}| = a_{sr}$ (as $|d_r| = |d_s| = |\xi| = 1$), it follows that we have $d_r\bar{\xi}a_{rs}d_s^{-1} = a_{rs}$ (and $d_r\xi a_{sr}d_s^{-1} = a_{sr}$) for all $r, s \in \langle n\rangle$. Hence, we have $D(\bar{\xi}A)D^{-1} = A$, or $DAD^{-1} = \xi A$. □

Our second proof of Proposition 3.7 will depend on the following numerical radius analog of Wielandt's lemma, which is of independent interest.

Lemma 3.8. *Let $A, B \in M_n$, and assume that A is nonnegative. If $|B| \leqslant A$, then $w(B) \leqslant w(A)$. Suppose, in addition, that* Re A *is irreducible. If $w(A) = w(B)$ and ξ is a unit complex number such that $\xi w(A) \in W(B)$, then $B = \xi DAD^{-1}$ for some unitary diagonal matrix D.*

Proof. The first half of this lemma can be readily proved by modifying the argument given in the proof for the first half of Corollary 3.6. Alternatively, apply the first part of Wielandt's lemma to the pair of matrices $\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}B)$, $\mathrm{Re}\,A$ (where $\theta \in \mathbb{R}$), and use the fact that for any matrix A, we have

$$w(A) = \max\left\{\lambda_{\max}(\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}A)) : \theta \in \mathbb{R}\right\} = \max\left\{\rho(\mathrm{Re}(\mathrm{e}^{\mathrm{i}\theta}A)) : \theta \in \mathbb{R}\right\}.$$

The proof for the second half of this lemma runs parallel to a known proof for the corresponding part of Wielandt's lemma (cf. [11, pp. 37–38]). Let y be a unit vector such that $y^*By = \xi w(A)$. Then

$$w(A) = y^*(\bar{\xi}B)y \leqslant |y|^{\mathrm{t}}|B||y| \leqslant |y|^{\mathrm{t}}A|y| \leqslant w(A).$$

Hence, the above inequalities all become equalities. Since $|y|^{\mathrm{t}}A|y| = w(A)$ and $\mathrm{Re}\,A$ is irreducible, by the last part of Proposition 3.3, we have $|y| > 0$. Now, in view of

$$|y|^{\mathrm{t}}(A - |B|)|y| = w(A) - w(A) = 0, \quad A - |B| \geqslant 0 \text{ and } |y| > 0,$$

we have $|B| = A$. Let D denote the unitary diagonal matrix $\mathrm{diag}(\eta_1/|\eta_1|, \ldots, \eta_n/|\eta_n|)$, where $y = (\eta_1, \ldots, \eta_n)^{\mathrm{t}}$. Then we have

$$|y|^{\mathrm{t}}D^*(\bar{\xi}B)D|y| = y^*(\bar{\xi}B)y = w(A),$$

where the second equality has already been established above. But we also already have $|y|^{\mathrm{t}}A|y| = w(A)$, so

$$|y|^{\mathrm{t}}D^*(\bar{\xi}B)D|y| = |y|^{\mathrm{t}}A|y|.$$

And since $|D^*(\bar{\xi}B)D| = |B| = A$ and $|y| > 0$, it follows that $\bar{\xi}D^{-1}BD = A$, or $B = \xi DAD^{-1}$. □

Second proof of Proposition 3.7. Apply Lemma 3.8 with $B = A$. □

It is easy to show that for any $A \in M_n$ and any unit complex number ξ, if A is (unitarily) diagonally similar to ξA, then $\operatorname{Re} A$ is also (unitarily) diagonally similar to $\operatorname{Re}(\xi A)$ (see [16, Remarks 2, 4 and 5]). The converse is not true in general. In [16, Lemma 2] it is shown that the converse is true if we assume, in addition, that the entries of A satisfy $a_{rs}a_{sr} = 0$ for all $r, s \in \langle n\rangle$. In view of the argument given in the last part of the first proof of Proposition 3.7, we now have another situation when the converse is true:

Remark 3.9. Let A be a nonnegative matrix. For any unit complex number ξ, ξA is diagonally similar to A if and only if $\operatorname{Re}(\xi A)$ is diagonally similar to $\operatorname{Re} A$.

The above remark is actually implicit in the work of [16]. This is because, by [16, Lemma 3], we readily obtain our remark under the additional assumption that $\operatorname{Re} A$ is irreducible (as $\lambda_{\max}(\operatorname{Re}(\xi A)) = \lambda_{\max}(\operatorname{Re} A)$ whenever $\operatorname{Re}(\xi A)$ is similar to $\operatorname{Re} A$), and then after a simple calculation we can drop the additional assumption. Certainly, our present proof is more direct and easier.

For more equivalent conditions for $\xi w(A) \in W(A)$ (when A is nonnegative and $|\xi| = 1$), see [16, Lemma 3].

We would like to make another observation.

Corollary 3.10. *Let A be a nonnegative matrix with irreducible real part. Let ξ be a unit complex number such that $\xi w(A) \in W(A)$. Then the subspace $\{x \in \mathbb{C}^n : x^*Ax = \xi w(A)\|x\|^2\}$ is of dimension* 1.

Proof. By Lemma 3.1, the set $\{x \in \mathbb{C}^n : x^*Ax = \xi w(A)\|x^2\|\}$ is equal to the eigenspace of $\operatorname{Re}(\bar{\xi} A)$ corresponding to $\lambda_{\max}(\operatorname{Re}(\bar{\xi} A))$ $(= w(A))$. By the first proof of Proposition 3.7 or by the proposition itself (and Remark 3.9), $\operatorname{Re}(\bar{\xi} A)$ is diagonally similar to $\operatorname{Re} A$. Since $\operatorname{Re} A$ is irreducible nonnegative, by the Perron–Frobenius theory, $\rho(\operatorname{Re} A)$ $(= w(A))$ is a simple eigenvalue of $\operatorname{Re} A$. Hence, $\lambda_{\max}(\operatorname{Re}(\bar{\xi} A))$ is also a simple eigenvalue of $\operatorname{Re}(\bar{\xi} A)$, and the said subspace is of dimension 1. □

We are going to extend Theorem A to the case when A is a nonnegative matrix with irreducible real part.

For any $A \in M_n$, it is easy to verify that the set

$$H = \left\{\xi \in \mathbb{C} : |\xi| = 1,\ \xi A \text{ is (unitarily) diagonally similar to } A\right\}$$

forms a subgroup of the group of all unit complex numbers, and moreover it is included in the set $\{\xi \in \mathbb{C} : |\xi| = 1,\ \xi W(A) = W(A)\}$. If A is nonnegative, then since $w(A) \in W(A)$, the latter set, in turn, is included in $\{\xi \in \mathbb{C} : |\xi| = 1,\ \xi w(A) \in W(A)\}$. Now assume, in addition, that $\operatorname{Re} A$ is irreducible. Then, in view of Proposition 3.7, the three sets are all equal. The group H may be infinite or finite. If H is infinite or has more than n elements, then the numerical range of A contains more than n points with modulus equal to $w(A)$. In this case, by a known result due to Anderson (see, for instance, [16, Lemma 6]), $W(A)$ is equal to the circular disk with center at the origin and radius $w(A)$. Hence, H is precisely the group of all unit complex numbers. On the other hand, if H is a finite group, say, with order m ($\leqslant n$), then by Lagrange's theorem in group theory, for any $\xi \in H$, we have $\xi^m = 1$, i.e., each element of H is an mth root of unity. But the cardinality of H is m, so it follows that H is precisely the group of all mth roots of unity. Summarizing, we have, in fact, established the following:

Proposition 3.11. *Let A be a nonnegative matrix with irreducible real part.*

(i) *For any unit complex number* ξ, *the following conditions are equivalent*:
   (a) ξA *is diagonally similar to A;*
   (b) $\xi W(A) = W(A)$;
   (c) $\xi w(A) \in W(A)$.

(ii) *The set* $\{\xi \in \mathbb{C} : |\xi| = 1,\ \xi w(A) \in W(A)\}$ *is a group under multiplication, and is either the group of all unit complex numbers or is a finite* (*necessarily cyclic*) *subgroup of it.*

(iii) *If* $W(A)$ *is not a circular disk with center at the origin, then*

$$\big\{z \in W(A) : |z| = w(A)\big\} = \big\{w(A)\mathrm{e}^{2\pi t\mathrm{i}/m} : t = 0, 1, \ldots, m-1\big\},$$

where m is the largest positive integer such that A is diagonally similar to $\mathrm{e}^{2\pi\mathrm{i}/m}A$.

So far we are graph-free. Next, in terms of certain graph-theoretic concepts, we are going to rewrite part (iii) of Proposition 3.11 in a readily usable form.

In [14, Theorem 1], the second author gave equivalent conditions on a complex matrix A with the property that the numerical range of any matrix with the same digraph as A is a circular disk centered at the origin. One equivalent condition is that A is permutationally similar to a block-shift matrix. Another equivalent condition is that all cycles of $G(A)$ have zero signed length. In [16, Theorem 1], a long list of further new equivalent conditions were added. In particular, rather unexpectedly, it was found that in the case when A is nonnegative and has a connected undirected graph (or equivalently, with irreducible real part), the condition that $W(A)$ is a circular disk centered at the origin is also an equivalent condition.

On the other hand, by [15, Theorem 4.1], for any $A \in M_n$ and any positive integer k, if A is k-cyclic, then A is diagonally similar to $\mathrm{e}^{2\pi\mathrm{i}/k}A$; if, in addition, the digraph $G(A)$ has at least one cycle with nonzero signed length, then the converse also holds.

In view of the above (and Remark 2.2), we can now rewrite Proposition 3.11(iii) as follows:

Theorem 3.12. *Let A be a nonnegative matrix with connected undirected graph. Suppose that the digraph $G(A)$ has at least one cycle with nonzero signed length. Then*

$$\left\{z \in W(A) : |z| = w(A)\right\} = \left\{w(A)\mathrm{e}^{2\pi t\mathrm{i}/m} : m = 0, 1, \ldots, m-1\right\},$$

where m is the cyclic index of A.

By [15, Corollary 4.2(i)], when the digraph $G(A)$ has at least one cycle with nonzero signed length (A not necessarily nonnegative), the cyclic index of A is equal to the greatest common divisor of the signed lengths of the cycles in $G(A)$. So, the cyclic index m considered in Theorem 3.12 can be determined.

We have already offered a self-contained proof (via Proposition 3.7) for Theorem A. Now let us show that Theorem A can be recovered also from Theorem 3.12: If A is irreducible, then, by part (iii) of the above-mentioned corollary of [15], the cyclic index of A is also equal to the greatest common divisor of the circuit lengths of $G(A)$. But it is well known (see, for instance, [1, p. 35, Theorem 2.30]) that the index of imprimitivity of an irreducible nonnegative matrix is equal to the greatest common divisor of the circuit lengths of its associated digraph. And, of course, the digraph of an irreducible matrix, being strongly connected, has at least one cycle with nonzero signed length (as every circuit can be regarded as a cycle with signed length equal to its length). Hence, we can recover Theorem A from Theorem 3.12.

More generally, we have the following:

Remark 3.13. Let A be a nonnegative matrix whose digraph has at least one cycle with nonzero signed length. Suppose A is permutationally similar to $A_1 \oplus \cdots \oplus A_k$, where $A_1, \ldots, A_k$ are nonnegative matrices each with connected undirected graph. Then:

(i) The cyclic index of A equals the greatest common divisor of the cyclic indices of those A_j whose digraphs have cycles with nonzero signed lengths.

(ii) The set $\{\xi \in \mathbb{C} : |\xi| = 1,\ \xi w(A) \in W(A)\}$ is equal to $\bigcup_j \{\xi \in \mathbb{C} : |\xi| = 1,\ \xi w(A_j) \in W(A_j)\}$, where the union is taken over all j for which $w(A_j) = w(A)$. If there is at least one j for which $w(A_j) = w(A)$ and the digraph $G(A_j)$ has no cycles with nonzero signed length, then $W(A)$ is a circular disk and, consequently, the above set is precisely the group of all unit complex numbers. Otherwise, the set is a union of certain Z_p's, where Z_p denotes the group of all complex pth roots of unity, and moreover it always includes the set $\{w(A)\mathrm{e}^{2\pi t\mathrm{i}/m} : t = 0, 1, \ldots, m-1\}$, where m is the cyclic index of A.

4. Nonnegative matrices whose numerical ranges are regular polygons

In his thesis [10, p. 24], Issos asked the question of when the numerical range of an irreducible nonnegative matrix is a regular (convex) polygon (with center not necessarily at the origin). In [16, Problem 2], Tam and Yang also posed the problem of characterizing nonnegative matrices whose numerical ranges are regular polygons with center at the origin. In this section, we are going to treat these problems.

A point α lying on the boundary of $W(A)$ is called a *sharp point* of $W(A)$ if $W(A)$ is included in an angular sector with apex at α and angle less than π. For a nonnegative matrix A, if $W(A)$ is a polygon, then $w(A)$ (being an extreme point) is necessarily one of the vertices and hence is a sharp point of $W(A)$. The problem of characterizing when $w(A)$ is a sharp point of $W(A)$ for a nonnegative matrix A has been solved by Tam and Yang [16, Theorem 4]. But we are going to rederive the result in a different way, relying ourselves on a general result about radial matrices. As the reader will see, our present approach has the merit that it gives us a better understanding, throws light on a known result in [12] and in addition yields more new results.

A matrix $A \in M_n$ is called *spectral* if $\rho(A) = w(A)$; A is *radial* if $\|A\| = \rho(A)$ or, equivalently, $w(A) = \|A\|$. (For equivalent conditions on a radial matrix or a spectral matrix, see [9, p. 45, Problem 27; pp. 61–62, Problem 37].)

Proposition 4.1. *Let $A \in M_n$ be a radial matrix. Then*:

(i) *There exists a unitary matrix $U \in M_n$ such that $U^*AU = D \oplus B$, where D is a diagonal matrix each of whose diagonal entries is of modulus $w(A)$ and B is a* (*possibly empty*) *matrix that satisfies $w(B) < w(A)$.*

(ii) *$W(A)$ is the convex hull of the polygon whose vertices are all the points in $W(A)$ with modulus $w(A)$ and a* (*possible empty*) *compact convex set*, *included in the open circular disk centered at the origin with radius $w(A)$.*

(iii) *Every point z in $W(A)$ with modulus $w(A)$ is a sharp point.*

Proof. Part (i) was proved in [13]. Part (ii) follows from the fact that $W(A) = \text{conv}(W(D) \cup W(B))$, where $W(B)$ is in the open disk centered at the origin with radius $w(A)$. Part (ii) follows readily from (iii). □

Concerning the problem of characterizing when $w(A)$ is a sharp point, we treat the case of a nonnegative matrix with irreducible real part first. The following result is a strengthening of [16, Remark 16]. We give an independent proof.

Theorem 4.2. *Consider the following conditions for a nonnegative matrix A*:

(a) *A is radial;*

(b) *$w(A)$ is a sharp point of $W(A)$;*

(c) *A is spectral;*

(d) $\rho(A) = \rho(\text{Re}\, A)$*;*

(e) *A and A^{t} have a common nonnegative eigenvector corresponding to $\rho(A)$.*

(i) *We always have the implications* (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Leftrightarrow$ (d) $\Rightarrow$ (e).

(ii) *When* Re A *is irreducible*, *conditions* (a)–(e) *are all equivalent.*

(iii) *If* Re A *is irreducible and conditions* (a)–(e) *are all satisfied*, *then A is necessarily irreducible.*

Proof. (i) (a) $\Rightarrow$ (b). This follows from Proposition 4.1(ii), as we always have $w(A) \in W(A)$ for a nonnegative matrix.

(b) $\Rightarrow$ (c). This follows from a result of Kippenhahn [9, Theorem 1.6.3], which says that if α is a sharp point of $W(A)$, where A is any complex matrix, then α must be an eigenvalue of A.

The equivalence of (c) and (d) follows from the relation $w(A) = \rho(\operatorname{Re} A)$ for a nonnegative matrix A.

The implication (c) $\Rightarrow$ (e). can be deduced from known results about normal eigenvalues; see [9, Theorem 1.6.6 and Section 1.6, Problem 11]. To make this proof self-contained, we offer an argument: Let x be a unit nonnegative eigenvector of A corresponding to $\rho(A)$. Then

$$x^{\mathrm{t}}(\operatorname{Re} A)x = x^{\mathrm{t}}Ax = \rho(A) = w(A).$$

Since A is nonnegative, $w(A)$ is also equal to $\lambda_{\max}(\operatorname{Re} A)$. This, together with our choice of x, implies that x is the desired common nonnegative eigenvector of A and A^{t} corresponding to $\rho(A)$ $(= \lambda_{\max}(\operatorname{Re} A))$.

(ii) In view of part (i), it suffices to show that when $\operatorname{Re} A$ is irreducible, we have the implication (e) $\Rightarrow$ (a).. Let x be a common nonnegative eigenvector of A and A^{t} corresponding to $\rho(A)$. Then clearly x is also a nonnegative eigenvector of the irreducible nonnegative matrix $\operatorname{Re} A$ (corresponding to $\rho(A)$). As such, x must be a positive vector (see, for instance, [11, p. 7, Theorem 2.2]). But we also have $A^{\mathrm{t}}Ax = \rho(A)^2 x$, and it is well known that a positive eigenvector of a nonnegative matrix must correspond to its spectral radius (see, for instance, [8, Corollary 8.1.30]), so $\rho(A^{\mathrm{t}}A) = \rho(A)^2$. Hence we have $\|A\|^2 = \rho(A^{\mathrm{t}}A) = \rho(A)^2$, or $\|A\| = \rho(A)$, i.e., A is radial.

(iii) It suffices to show that if A is a nonnegative matrix with irreducible real part, and if A is spectral, then A is irreducible. We assume to the contrary that A is reducible. By applying a permutation similarity, we may assume that A is already in the Frobenius normal form, i.e., a (lower) triangular block form with, say, p irreducible blocks $A_{11}, \ldots, A_{pp}$ along the diagonal (see, for instance, [1, p. 39]). Let B denote the matrix $A_{11} \oplus \cdots \oplus A_{pp}$. Then we have $\rho(A) = \max_{1 \leqslant j \leqslant p} \rho(A_j) = \rho(B)$. Since A is reducible and the undirected graph of A is connected, clearly we have $p \geqslant 2$, $A \geqslant B$ and $A \neq B$. By Corollary 3.6, it follows that we have $w(A) > w(B) \geqslant \rho(B) = \rho(A)$, which is a contradiction. $\square$

Our next example will show that, for a general nonnegative matrix A, the missing implications in Theorem 4.2(i) do not hold in general.

Example 4.3. Consider the nonnegative matrix $A = A_1 \oplus A_2$, with

$$A_1 = [1] \quad \text{and} \quad A_2 = \begin{bmatrix} 0 & \alpha \\ 0 & 0 \end{bmatrix},$$

where α is a positive number to be chosen. Note that we have $\rho(A) = 1$, $w(A) = \max\{1, \alpha/2\}$ and $\|A\| = \max\{1, \alpha\}$. Moreover, $W(A_1) = \{1\}$, $W(A_2)$ is the circular disk centered at the origin with radius $\alpha/2$, and $W(A) = \text{conv}(W(A_1) \cup W(A_2))$. It is clear that A and A^t always have a common nonnegative eigenvector corresponding to $\rho(A)$ $(= 1)$, namely, the vector $(1, 0, 0)^t$. If $\alpha > 2$, then $w(A) = \alpha/2 > \rho(A)$, and so A is not spectral. This shows that for the conditions (a)–(e) of Theorem 4.2, (e) $\not\Rightarrow$ (c). If $\alpha = 2$, then $\rho(A) = w(A) = 1$ and so A is spectral. However, in this case, $W(A)$ is the circular disk centered at the origin with radius $w(A)$, and $w(A)$ is not a sharp point of $W(A)$. This shows that (c) $\not\Rightarrow$ (b). Finally, if $1 < \alpha < 2$, then clearly $w(A)$ is a sharp point of $W(A)$. Since $\|A\| = \alpha > 1 = \rho(A)$, A is not radial. This shows that (b) $\not\Rightarrow$ (a).

It is clear that we can use Theorem 4.2 (and an argument given in the first paragraph of the proof of [16, Lemma 5]) to recover [16, Lemma 5]. Now we use Theorem 4.2 to derive [16, Theorem 4]:

Corollary 4.4. *Let A be a nonnegative matrix, and suppose A is permutationally similar to $A_1 \oplus \cdots \oplus A_k$, where $A_1, \ldots, A_k$ are nonnegative matrices each with connected undirected graph. A necessary and sufficient condition for $w(A)$ to be a sharp point of $W(A)$ is that, for any j, $1 \leqslant j \leqslant k$, we have*
(a) *If $\rho(A_j) = \rho(A)$, then A_j is itself an irreducible matrix and $\rho(A_j) = w(A_j)$.*
(b) *If $\rho(A_j) < \rho(A)$, then $w(A_j) < \rho(A)$.*

Proof. “Necessity” Consider any $j \in \langle k\rangle$ for which $\rho(A_j) = \rho(A)$. Since $w(A)$ is a sharp point of $W(A)$, we have $\rho(A) = w(A) \geqslant w(A_j) \geqslant \rho(A_j) = \rho(A)$, hence $w(A_j) = \rho(A_j)$, i.e., A_j is spectral. Since Re A_j is irreducible, by Theorem 4.2(ii) and (iii), it follows that A_j is irreducible.

Now consider any $j \in \langle k\rangle$ for which $\rho(A_j) < \rho(A)$. If $w(A_j) = \rho(A)$, then we have $\rho(A_j) < w(A_j)$, and so $w(A)$ $(= w(A_j))$ is not a sharp point of $W(A_j)$, and hence also not a sharp point of $W(A)$, which is a contradiction.

“Sufficiency” When conditions (a) and (b) are satisfied, clearly we have $w(A) = \max_{1 \leqslant j \leqslant k} w(A_j) = \rho(A)$. For any j for which $\rho(A_j) = \rho(A)$, by condition (a) and Theorem 4.2, $w(A_j)$ is a sharp point of $W(A_j)$ and the matrix A_j is radial. Moreover, by Proposition 4.1(ii), for any such j, $W(A_j)$ is the convex hull of $w(A)$ $(= w(A_j))$ and some compact convex set C_j not containing $w(A)$. On the other hand, if j is such that $\rho(A_j) < \rho(A)$, then, by condition (b), $W(A_j)$ is a compact convex set not containing $w(A)$ $(= \rho(A) > w(A_j))$. It is clear that the convex hull of all C_j for which $\rho(A_j) = \rho(A)$ and all $W(A_j)$ for which $\rho(A_j) < \rho(A)$ is a compact convex set C that does not contain $w(A)$. But $W(A)$ is the convex hull of $w(A)$ and C, hence $w(A)$ is a sharp point of $W(A)$. $\square$

In [12, Theorem 1.2], Nylen and Tam proved that if A is a primitive doubly stochastic matrix, then $W(A)$ is symmetric about the real axis and is the convex hull of the point 1 and a compact convex set included in the open unit disk. Motivated by their result, we have the following for a nonnegative matrix:

Proposition 4.5. *Let A be a nonnegative matrix, and suppose A is permutationally similar to $A_1 \oplus \cdots \oplus A_k$, where $A_1, \ldots, A_k$ are nonnegative matrices each with irreducible real part. If $w(A)$ is a sharp point of $W(A)$, then we have $W(A) =$ $\operatorname{conv}(P \cup C)$, where P is the polygon with vertices consisting of all points in $W(A)$ with modulus $w(A)$, and C is some compact convex set included in the open circular disk centered at the origin with radius $w(A)$.*

Proof. Since $w(A)$ is a sharp point of $W(A)$, conditions (a) and (b) of Corollary 4.4 are fulfilled. Consider any $j \in \langle k \rangle$. If $\rho(A_j) < \rho(A)$, then $W(A_j)$ is a compact convex set included in the open circular disk centered at the origin with radius $w(A)$. If $\rho(A_j) = \rho(A)$, then $w(A_j)$ is a sharp point of $W(A_j)$ (see the "necessity part" of the proof of Corollary 4.4), and by Theorem 4.2(ii), the matrix A_j is radial. In this case, by Proposition 4.1(ii), $W(A_j)$ is the convex hull of a polygon with vertices all of modulus $w(A_j)$ $(= w(A))$ and some compact convex set included in the open circular disk centered at the origin with radius $w(A)$. In view of $W(A) = \operatorname{conv}(W(A_1) \cup \cdots \cup W(A_k))$, it is ready to see that our assertion follows. □

Corollary 4.6. *Let A be a primitive matrix. If A satisfies one of the conditions* (a)–(e) *in Theorem* 4.2, *then $W(A)$ is symmetric about the real axis and is the convex hull of the point $\rho(A)$ and a compact convex set included in the open circular disk centered at the origin with radius $\rho(A)$.*

In view of Theorem 4.2(iii) and the following result, in solving the problem of characterizing nonnegative matrices whose numerical ranges are regular polygons with center at the origin, we may focus our attention to irreducible nonnegative matrices.

Theorem 4.7. *Let A be a nonnegative matrix. Suppose A is permutationally similar to $A_1 \oplus \cdots \oplus A_k$, where $A_1, \ldots, A_k$ are nonnegative matrices each with irreducible real part. Then $W(A)$ is a regular polygon with center at the origin if and only if there exists $s \in \langle k \rangle$ such that $W(A_s)$ is a regular polygon with center at the origin, and for every $j \in \langle k \rangle$, $j \neq s$, we have $W(A_j) \subseteq W(A_s)$.*

Proof. The 'if' part is obvious. Since $w(A)$ is always an extreme point of $W(A)$ (as A is nonnegative), to prove the "only if" part, we may suppose that $W(A)$ is the regular polygon with center at the origin given by

$$W(A) = \operatorname{conv}\left\{ w(A) e^{2\pi t \mathrm{i}/m} : t = 0, 1, \ldots, m-1 \right\} \quad \text{for some } m \geqslant 2.$$

By our assumption on A, clearly, $W(A) = \operatorname{conv}(W(A_1) \cup \cdots \cup W(A_k))$. But $w(A)$ $e^{2\pi \mathrm{i}/m}$ is an extreme point of $W(A)$, so it must belong to one of the sets $W(A_1), \ldots,$

$W(A_k)$, say, $W(A_s)$. Then $w(A_s) = w(A)$ and, by Proposition 3.11(ii), $W(A_s)$ contains each of the points $w(A)\mathrm{e}^{2\pi t\mathrm{i}/m},\ t = 0, 1, \ldots, m-1$. Hence, by the convexity of the numerical range of a matrix, we have

$$W(A_s) \supseteq \mathrm{conv}\{w(A)\mathrm{e}^{2\pi t\mathrm{i}/m} : t = 0, 1, \ldots, m-1\} = W(A).$$

Certainly, we also have $W(A_s) \subseteq W(A)$. So our assertion follows. □

We would like to mention that a similar result also holds for the question of when a general complex matrix has a circular disk with center at the origin as its numerical range (see [16, Theorem 3]). We also want to emphasize that in Theorem 4.7 the nonnegativity assumption on A cannot be dropped. Counterexamples can be easily constructed.

An application of Theorem 3.12 yields the following related result:

Proposition 4.8. *Let A be a nonnegative matrix with connected undirected graph. Suppose that the digraph $G(A)$ has at least one cycle with nonzero signed length. Assume that the cyclic index of A is greater than* 1. *Then $W(A)$ cannot be a circular disk, and moreover if $W(A)$ is a regular polygon then its center must be at the origin.*

Proof. Let m (>1) be the cyclic index of A. Assume first that $W(A)$ is a circular disk. In view of Theorem 3.12, each of the m points $w(A)\mathrm{e}^{2\pi t\mathrm{i}/m},\ t = 0, 1, \ldots, m-1$, is an extreme point of $W(A)$. Certainly, all of them lie on the circumference of the circular disk $W(A)$, and the center of the disk must be equidistant from all of them. It follows that the center of the disk is the origin of the complex plane. In other words, $W(A)$ is the circular disk with center at the origin and radius $w(A)$, in contradiction with the result of Theorem 3.12.

The same argument also shows that if $W(A)$ is a regular polygon, then its center must be at the origin. □

Corollary 4.9. *If A is an irreducible nonnegative matrix with index of imprimitivity greater than* 1, *then $W(A)$ cannot be a circular disk.*

It seems plausible that this is the case for any irreducible nonnegative matrix A. Here we verify it for 2-by-2 matrices.

Proposition 4.10. *No irreducible nonnegative* 2*-by-*2 *matrix can have a circular disk as its numerical range.*

Proof. Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

be a 2-by-2 irreducible nonnegative matrix whose numerical range is a circular disk with center λ and radius r. It is known that in this case both eigenvalues of A are equal to λ and A is unitarily similar to

$$\begin{bmatrix} \lambda & 2r \\ 0 & \lambda \end{bmatrix}.$$

We have $a + d = 2\lambda$ and $ad - bc = \lambda^2$. It follows that $ad - bc = (a + d)^2/4$, and thus $0 \leqslant (a - d)^2 = -4bc \leqslant 0$. This shows that $b = 0$ or $c = 0$. In either case, A is reducible contradicting our assumption. □

An alternative way to complete the proof of Proposition 4.10 is to apply the Perron–Frobenius theory: since A is unitarily similar to

$$\begin{bmatrix} \lambda & 2r \\ 0 & \lambda \end{bmatrix},$$

we have, λ equals $\rho(A)$ and is not a simple eigenvalue, contradiction.

The preceding argument can also be used to show that if A is a 3-by-3 or 4-by-4 primitive matrix with zero trace, then $W(A)$ cannot be a circular disk. For if $W(A)$ is a circular disk with center λ, then by [2, Remark 2] λ is a (real) eigenvalue of A with multiplicity at least two, hence $\lambda < \rho(A)$. On the other hand, by Corollary 3.4 and also the fact that $\rho(A)$ is the only eigenvalue of A with modulus $\rho(A)$, λ must be a positive number. If A is 3-by-3, then $\operatorname{tr} A = \rho(A) + 2\lambda > 0$, contradiction. If A is 4-by-4, then the eigenvalues of A are $\rho(A),\ \lambda,\ \lambda$, and $-(\rho(A) + 2\lambda)$, which is again a contradiction, as $|-(\rho(A) + 2\lambda)| > \rho(A)$.

In [12, Example 4.5], Nylen and Tam gave an example of an irreducible doubly stochastic matrix with index of imprimitivity two for which $W(A)$ is not a line segment (2-polygon). In related to that, we make the following simple observation:

Remark 4.11. Let A be a real matrix. Then $W(A)$ is a line segment if and only if either A is symmetric or A is the sum of a real scalar matrix and a skew-symmetric matrix. If A is nonnegative, then $W(A)$ is a line segment if only if A is symmetric.

Indeed, first note that $W(A)$ is symmetric with respect to the real axis for the real A. Hence if $W(A)$ is a line segment, it will either be lying in $\mathbb{R}$ or be perpendicular to $\mathbb{R}$. In the former case, A is symmetric. For the latter, assuming that $W(A)$ lies in the vertical line $x = a$, we have $W(\mathrm{i}(A - aI)) \subseteq \mathbb{R}$. Thus $A = aI + B$ with $B = -\mathrm{i}(\mathrm{i}(A - aI))$ skew-symmetric. The converse is trivial. If A is nonnegative and $W(A)$ is a line segment, then $W(A)$ cannot be perpendicular to the real axis for otherwise $w(A)$ would not be in $W(A)$. In this case, $W(A) \subseteq \mathbb{R}$ and hence A is symmetric.

The more general question of when the numerical range of an irreducible nonnegative matrix is a regular polygon with center at the origin is actually already answered by Tam and Yang [16, Remark 15]:

Remark 4.12. For any irreducible nonnegative matrix A with index of imprimitivity $m \geqslant 2$, $W(A)$ is a regular polygon (necessarily with center at the origin) if and only if $W(A) = \text{conv}\{\rho(A)e^{2\pi t i/m} : t = 0, 1, \ldots, m-1\}$.

Remark 4.12 also settles Issos's question, mentioned at the beginning of this section, for almost all cases except for the primitive matrix case. Our argument can also be used to show that, if A is a primitive matrix, then $W(A)$ can never be a regular polygon with center at the origin. Certainly, there are primitive matrices whose numerical ranges are regular polygons. For instance, take an irreducible nonnegative matrix A with index of imprimitivity $m > 1$ such that $W(A)$ is a regular polygon with center at the origin. Then for any $\alpha > 0$, $A + \alpha I$ is a primitive matrix whose numerical range is a regular polygon (with center at α). The problem of characterizing primitive matrices with regular polygons as their numerical ranges remains open.

Contrary to what is said in [16, p. 218, first paragraph], the condition given in Remark 4.12 can be transformed to a checkable condition. First, we observe the following:

Lemma 4.13. *Let $A \in M_n$. Let ρ be a positive real number and let $m \geqslant 2$ be a given positive integer. In order that*

$$W(A) \subseteq \text{conv}\left\{\rho e^{2\pi t i/m} : t = 0, 1, \ldots, m-1\right\},$$

it is necessary and sufficient that for $t = 0, 1, \ldots, m-1$, we have

$$\lambda_{\max}\left(\text{Re}(e^{-(2t-1)\pi i/m} A)\right) \leqslant \rho \cos\frac{\pi}{m}.$$

Proof. The polygon $\text{conv}\{\rho e^{2\pi t i/m} : t = 0, 1, \ldots, m-1\}$ can be expressed as $\bigcap_{t=0}^{m-1} H_t$, where H_t is the closed half-plane given by

$$H_t = \left\{ z = x + iy \in \mathbb{C} : x \cos\frac{(2t-1)\pi}{m} + y \sin\frac{(2t-1)\pi}{m} \leqslant \rho \cos\frac{\pi}{m} \right\}.$$

In order that $W(A)$ be included in the said polygon, it is necessary and sufficient that $W(A) \subseteq H_t$ for all t. Now $W(A) \subseteq H_t$ if and only if $W(e^{-(2t-1)\pi i/m} A)$ is included in the half-plane $\{z \in \mathbb{C} : \text{Re}\, z \leqslant \rho\cos(\pi/m)\}$, and the latter condition is fulfilled if and only if $\lambda_{\max}(\text{Re}(e^{-(2t-1)\pi i/m} A)) \leqslant \rho\cos(\pi/m)$. So our assertion follows. $\square$

Now we have the following:

Proposition 4.14. *Let A be an irreducible nonnegative matrix with index of imprimitivity m. In order that $W(A)$ be a regular polygon with center at the origin it is necessary and sufficient that the following conditions are both satisfied*:

(a) $\rho(A) = \rho(\text{Re}(A))$.

(b) *For* $t = 0, 1, \ldots, m-1$, $\lambda_{\max}(\text{Re}(e^{-(2t-1)\pi i/m} A)) = \rho(A)\cos(\pi/m)$.

(*In condition* (b), *we may replace the last equality by* "$\leqslant$".)

Proof. "Necessity" Suppose that $W(A)$ is a regular polygon with center at the origin. Then $w(A)$ must be a sharp point of $W(A)$ and, as noted before, condition (a) necessarily holds. Furthermore, by Remark 4.12, in this case we have

$$W(A) = \operatorname{conv}\{\rho(A)e^{2\pi t i/m} : t = 0, 1, \ldots, m-1\}.$$

It follows from Lemma 4.13 that for $t = 0, 1, \ldots, m-1$, we have

$$\lambda_{\max}\big(\operatorname{Re}(e^{-(2t-1)\pi i/m}A)\big) \leqslant \rho(A)\cos\frac{\pi}{m}.$$

Here we can replace each of the latter inequalities by an equality, because $W(A)$ is precisely the convex hull of the m points $\rho(A)e^{2\pi t i/m},\ t = 0, 1, \ldots, m-1$, not just a subset of it. So we have condition (b).

"Sufficiency" Again by Lemma 4.13, condition (b) implies that $W(A)$ is included in the regular polygon with vertices $\rho(A)e^{2\pi t i/m},\ t = 0, 1, \ldots, m-1$. These are the same as $w(A)e^{2\pi t i/m}$ by condition (a). But by Theorem A, $W(A)$ also contains each of these m points. Hence, $W(A)$ is equal to the said regular polygon. □

In view of Corollary 4.4, Theorem 4.7 and Proposition 4.14, in theory (assuming that all numerical quantities can be computed exactly), we can determine whether the numerical range of a nonnegative matrix is a regular polygon with center at the origin in the following way:

By a permutation similarity, we may rewrite the given nonnegative matrix A as $A_1 \oplus \cdots \oplus A_k$, where $A_1, \ldots, A_k$ are each nonnegative matrices with connected undirected graph. Then we follow the steps given below. If we obtain positive answers at each step, then $W(A)$ is a regular polygon with center at the origin. Otherwise, $W(A)$ is not.

Step 1. For each $j = 1, \ldots, k$, determine the values of $\rho(A_j)$ and $w(A_j)$ $(= \rho(\operatorname{Re}A_j))$. (Then $\rho(A) = \max_{1\leqslant j\leqslant k}\rho(A_j)$ and $w(A) = \max_{1\leqslant j\leqslant k} w(A_j)$.) Answer the following question: Is there a j such that $\rho(A_j) = \rho(A)$ and A_j satisfies the criterion for $W(A_j)$ to be a regular polygon with center at the origin, as given by Proposition 4.14?

Step 2. Let Λ denote the set of all j for which $\rho(A_j) = \rho(A)$ and $W(A_j)$ is a regular polygon with center at the origin. For each $j \in \Lambda$, determine the index of imprimitivity m_j of A_j (for instance, by finding the greatest common divisor of the circuit lengths of $G(A_j)$). Answer the following question: Is there a $j_0 \in \Lambda$ such that m_j divides m_{j_0} for each $j \in \Lambda$? (If such j_0 exists, hopefully $W(A)$ equals $W(A_{j_0})$.)

Step 3. Answer the following question: Is it true that, for each j for which $\rho(A_j) < \rho(A)$ or $\rho(A_j) = \rho(A)$ but $j \notin \Lambda$, we have $W(A_j) \subseteq W(A_{j_0})$? (Use Lemma 4.13 here.)

If we expect that $W(A)$ is not a regular polygon with center at the origin, we may also add the following step at the beginning:

Step 0. For $j = 1, \ldots, k$, determine $\rho(A_j)$ and $w(A_j)$ $(= \rho(\operatorname{Re} A_j))$. Answer the following questions:

(i) Is $\rho(A) = w(A)$?

(ii) For each j for which $\rho(A_j) = \rho(A)$, is A_j an irreducible matrix and do we have $\rho(A_j) = w(A_j)$?

(iii) For each j for which $\rho(A_j) < \rho(A)$, do we have $w(A_j) < \rho(A)$?

(If the answers are all "yes", then $w(A)$ is a sharp point of $W(A)$.)

Now we would also like to address the question of when the numerical range of a nonnegative matrix A has weak circular symmetry, i.e., $e^{2\pi i/m}W(A) = W(A)$ for some integer $m,\ 2 \leqslant m \leqslant n$, where n is the size of A. The question was solved for the special case when the undirected graph of A is connected (see [16, Theorem 2]). Clearly, the convex sets $W(A)$ and $W(e^{2\pi i/m}A)$ are equal if and only if they have same supporting lines in all directions. So one may give the following answer to the above question:

$$\begin{aligned}&e^{2\pi i/m}W(A) = W(A) \text{ if and only if}\\ &\quad \lambda_{\max}\big(\mathrm{Re}(e^{i\theta}A)\big) = \lambda_{\max}\big(\mathrm{Re}(e^{i(\theta+2\pi/m)}A)\big) \quad \forall\theta \in [0, 2\pi).\end{aligned}$$

But this is not a satisfactory answer, as there are infinitely many conditions we need to check. One may also try to reduce the problem to the case of a nonnegative matrix with connected undirected graph, and suspect that a result similar to Theorem 4.7 or [16, Theorem 3] also holds for the question of weak circular symmetry. The following example shows that this is not the case.

Example 4.15. Choose an irreducible nonnegative matrix A_1 whose numerical range is the triangle $\triangle = \mathrm{conv}\{1, e^{2\pi i/3}, e^{4\pi i/3}\}$. Also choose a nonnegative block-shift matrix A_2 with connected undirected graph such that the numerical range of A_2 is a circular disk, centered at origin, radius r, where r is greater than the radius of the inscribed circle of $\triangle$ but is less than that of the circumscribed circle. Now let $A = A_1 \oplus A_2$. Then $e^{2\pi i/3}W(A) = W(A)$, but we have neither $W(A_1) \subseteq W(A_2)$ nor $W(A_2) \subseteq W(A_1)$.

Note that in the above example $W(A)$ is not a polygon. But by modifying the example, we can easily construct one in which $W(A)$ is a (nonregular) polygon. The method of construction of our examples also suggests the following question:

Question 4.16. Let A be a nonnegative matrix which is permutationally similar to $A_1 \oplus \cdots \oplus A_k$, where $A_1, \ldots, A_k$ are nonnegative matrices each with a connected undirected graph. If, for some positive integer $m \geqslant 2$, we have $e^{2\pi i/m}W(A) = W(A)$, does it follow that there exist distinct indices $i_1, \ldots, i_p \in \langle k\rangle,\ p \geqslant 1$, such that $e^{2\pi i/m}W(A_{i_r}) = W(A_{i_r})$ for $r = 1, \ldots, p$, and $W(A_j) \subseteq \mathrm{conv}(W(A_{i_1}) \cup \cdots \cup W(A_{i_p}))$ for all $j \neq i_1, \ldots, i_p$?

We do not know the answer to the above question even when $W(A)$ is assumed to be a polygon. Also, note that if we drop the nonnegativity of A, the answer to the above question is clearly "no".

Acknowledgements

This research started when the three authors were attending the Fifth Workshop on Numerical Ranges and Numerical Radii, Nafplio, Greece, June, 2000. We have learned that J. Maroulas, P.J. Psarrakos and M.J. Tsatsomeros had shown in their paper "Perron–Frobenius type results on the numerical range" that no irreducible nonnegative matrix can have a circular disk or an elliptic disk with foci off the real axis as its numerical range. This answers affirmatively the question we ask, preceding Proposition 4.10. We thank them for sending us their preprint. Thanks are also due to John Drew for a stimulating discussion with the first author that led to Example 3.5.

References

[1] A. Berman, R.J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, vol. 9, SIAM, Philadelphia, 1994 (Revised reprint of the 1979 original, Classics in Applied Mathematics).
[2] M.T. Chien, B.S. Tam, Circularity of the numerical range, Linear Algebra Appl. 201 (1994) 113–133.
[3] A. Dritschel, H.J. Woerdeman, Model theory and linear extreme points in the numerical radius unit ball, Memoirs Amer. Math. Soc. 129 (1997) 615.
[4] M.R. Embry, The numerical range of an operator, Pacific J. Math. 32 (1970) 647–650.
[5] M. Fiedler, Geometry of the numerical range of matrices, Linear Algebra Appl. 37 (1981) 81–96.
[6] K.E. Gustafson, D.K.M. Rao, Numerical Range: The Field of Values of Linear Operators and Matrices, Springer, New York, 1996.
[7] M. Goldberg, E. Tadmer, On the numerical radius and its applications, Linear Algebra Appl. 42 (1982) 263–284.
[8] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1985.
[9] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge, 1991.
[10] J.N. Issos, The field of values of non-negative irreducible matrices, Ph.D. thesis, Auburn University, 1966.
[11] H. Minc, Nonnegative Matrices, Wiley, New York, 1988.
[12] P. Nylen, T.Y. Tam, Numerical range of a doubly stochastic matrix, Linear Algebra Appl. 153 (1991) 161–176.
[13] V. Pták, Isometric parts of an operator and the critical exponent, Časopis Pěst. Mat. 101 (1976) 383–388.
[14] B.S. Tam, Circularity of numerical ranges and block-shift matrices, Linear and Multilinear Algebra 37 (1994) 93–109.
[15] B.S. Tam, On matrices with cyclic structure, Linear Algebra Appl. 302–303 (1999) 377–410.
[16] B.S. Tam, S. Yang, On matrices whose numerical ranges have circular or weak circular symmetry, Linear Algebra Appl. 302–303 (1999) 193–221.
[17] H. Wielandt, Unzerlegbare, nicht negative matrizen, Math. Z. 52 (1950) 642–648.