Linear Algebra and its Applications 435 (2011) 2639–2657

Contents lists available at ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

Numerical ranges of cube roots of the identity

Thomas Ryan Harris [a], Michael Mazzella [a], Linda J. Patton [a,*], David Renfrew [a,1], Ilya M. Spitkovsky [b]

[a] Mathematics Department, California Polytechnic State University, San Luis Obispo, CA 93407, United States
[b] Department of Mathematics, College of William and Mary, Williamsburg, VA 23187-8795, United States

ARTICLE INFO

Article history:
Received 26 August 2010
Accepted 10 March 2011
Available online 1 June 2011

Submitted by L. Rodman

Keywords:
Numerical range
Algebraic operator
Threefold symmetry

ABSTRACT

The numerical range of a bounded linear operator T on a Hilbert space H is defined to be the subset $W(T) = \{\langle Tv, v\rangle : v \in H, \|v\| = 1\}$ of the complex plane. For operators on a finite-dimensional Hilbert space, it is known that if $W(T)$ is a circular disk then the center of the disk must be a multiple eigenvalue of T. In particular, if T has minimal polynomial $z^3 - 1$, then $W(T)$ cannot be a circular disk. In this paper we show that this is no longer the case when H is infinite dimensional. The collection of 3×3 matrices with threefold symmetry about the origin are also classified.

1. Introduction

If H is a complex Hilbert space and T is a bounded linear operator on H, the numerical range of T is the subset $W(T)$ of the complex plane defined by

$$W(T) = \{\langle Tv, v\rangle \mid v \in H, \|v\| = 1\}.$$

Since the quadratic forms in the definition of $W(T)$ arise naturally in, or are the primary object of study in, so many problems involving the operator T, properties of the numerical range have been extensively developed. Some standard references about the numerical range include [4] and Chapter I of [6]. The numerical radius of T, written as $w(T)$, is the supremum of the moduli of values in $W(T)$:

$$w(T) = \sup\{|z| : z \in W(T)\}.$$

* Corresponding author.
E-mail addresses: thomasryanharris@gmail.com (T.R. Harris), relations13@gmail.com (M. Mazzella), lpatton@calpoly.edu (L.J. Patton), drenfrew@math.ucdavis.edu (D. Renfrew), ilya@math.wm.edu (I.M. Spitkovsky).
1 Present address: Mathematical Sciences Building, One Shields Ave., University of California, Davis, CA 95616, United States.

0024-3795/$ - see front matter
doi:10.1016/j.laa.2011.03.020

Of course, $w(T) \leqslant \|T\|$. For finite-dimensional H, the numerical range $W(T)$ is always closed, so that sup in the definition of $w(T)$ can be replaced by max. This is no longer the case in the infinite dimensional setting.

The most famous result about the numerical range is the Toeplitz–Hausdorff theorem, going back to [5,18], according to which $W(T)$ is always convex. If T acts on a 2-dimensional space, then $W(T)$ is an ellipse with the foci at the eigenvalues of T. For a more detailed description and the proof of this result, as well as of the Toeplitz–Hausdorff theorem, see e.g. [6].

As it turns out, the elliptical shape of the numerical range is actually determined not by the dimension of the underlying space but by the fact that the operator is annihilated by a second degree polynomial. Such operators are called *quadratic*, and the respective result was established by Tso and Wu [19] (see also [15], where this result was extended to show that several types of generalized numerical ranges of quadratic operators are also ellipses, open or closed). For our purposes, the following particular case is relevant.

Theorem 1. *If T is an operator on a Hilbert space H and $T^2 = I$ with $T \neq \pm I$, then $W(T)$ is an elliptical disk (open or closed) with foci at ± 1 or the closed interval $[-1, 1]$. In particular, it is not a circular disk.*

In [2], the numerical ranges of composition operators induced by disk automorphisms were classified for many types of automorphisms; in many cases the numerical ranges are disks centered at the origin. In agreement with Theorem 1, this never is the case for elliptic automorphisms with rotation parameter ω satisfying $\omega^2 = 1$.

The authors of [2] conjectured that for an automorphic composition operator satisfying $T^n = I$ for any natural n, $W(T)$ is not a disk. Note that for T acting on a finite dimensional space, $W(T)$ can be a circular disk only if the center of this disk is an multiple eigenvalue of T (see, e.g. [11, Corollary 4.4], and also [20] for stronger more recent results and historical comments). So, operators satisfying $T^n = I$ cannot have circular numerical ranges in the finite dimensional setting. However, in Theorem 14, we construct an operator T on an infinite-dimensional Hilbert space where $T^3 = I$ and $W(T)$ is a disk. This does not answer the original question about composition operators, but it suggests techniques specific to composition operators may be required for $n \geqslant 3$.

Some basic definitions used throughout the paper follow.

Definition 2. If T is any operator on a Hilbert space that satisfies $q(T) = 0$ for a non-zero polynomial q, then T is algebraic.

Every finite matrix, but not every operator on an infinite dimensional Hilbert space, is algebraic. If T is algebraic, then the unique monic polynomial q of lowest degree for which $q(T) = 0$ is called the minimal polynomial of T.

The support function of a convex set is used in the analysis that follows, so its definition and properties will be briefly developed.

Definition 3. If E is a convex subset of the complex plane, then the support function p_E is defined for all real θ by:

$$p_E(\theta) = \sup \left\{ \Re e^{-i\theta} z \mid z \in E \right\}.$$

The value $p_E(\theta)$ is the maximum scalar projection of the set E in the direction of θ. Assume E contains the origin. Clearly the definition of E will then imply $p_E(\theta) \geqslant 0$ for all θ. If a line L is extended from the origin in the direction of the vector $(\cos(\theta), \sin(\theta))$ then $p_E(\theta)$ will be the distance from the origin to the point on ∂E where a line L' perpendicular to L is tangent to the boundary of E.

The support function completely determines the closure of E; that is, if $p_E(\theta) = p_F(\theta)$ for convex sets E and F and all real θ, then $\overline{E} = \overline{F}$.

When the set E is an ellipse, the support function has a simple formula. This result appears in many standard references about convex sets as well as in [2,3]:

Proposition 4. *If $a, b > 0$ and E is the elliptical disk determined by $\frac{x^2}{a^2} + \frac{y^2}{b^2} \leqslant 1$, then*

$$p_E(\theta) = \sqrt{a^2 \cos^2(\theta) + b^2 \sin^2(\theta)},$$

for all real θ.

In the remaining discussion, $E = W(T)$ for an operator T on a Hilbert space. In this case we will abbreviate $p_{W(T)}$ to simply p_T.

In Section 2, a support function for 3×3 matrices with minimal polynomial $z^3 - 1$ is derived; it is used in Section 3 to produce an operator T on an infinite dimensional Hilbert space that satisfies $T^3 = I$ and also has a disk as its numerical range. Section 4 provides necessary and sufficient conditions for a 3×3 matrix to have threefold symmetry about the origin.

2. Three by three matrices

The numerical ranges of 3×3 matrices were classified by Kippenhahn [10]. The numerical range of a 3×3 matrix M is either (1) the convex hull of its eigenvalues, (2) the convex hull of an ellipse and a point (which reduces to an ellipse if the point is inside the ellipse), (3) a shape with a flat portion on the boundary, and (4) an ovular shape. This classification is in terms of the *associated curve* of M. The latter is defined by the equation $L_M(u, v, w) = 0$ in homogenous line coordinates, where

$$L_M(u, v, w) = \det(uH + vK + wI), \tag{1}$$

and H and K are the real and imaginary Hermitian parts of M, respectively.

An alternative classification in terms of the entries of M and its standard canonical forms was given in [9], and further analysis about 3×3 matrices with flat portions was provided in [14]. Since a matrix satisfying $M^3 = I$ (and satisfying no lower degree polynomial equation) must have all three distinct cube roots of unity as its eigenvalues, results in [9] or [11] show the numerical range of such a matrix is not a disk. A natural question is whether having a minimal polynomial of $z^3 - 1$ prevents any of the other types of numerical ranges in a 3×3 matrix. The classification results in [9] and [14] can also be used to show that any of the four numerical range possibilities above can occur for a matrix M satisfying $M^3 = I$; the elliptical numerical ranges just cannot reduce to disks. Specific examples of matrices that have each possible type of numerical range follow.

Let $\omega_1 = e^{i\frac{2\pi}{3}}$ and $\omega_2 = e^{i\frac{4\pi}{3}}$, so all of the matrices in the four examples below have minimal polynomial $z^3 - 1$.

Example 5. If M is a diagonal matrix with diagonal entries $1, \omega_1, \omega_2$, then $M^3 = I$ and the numerical range of M is the triangle with vertices at 1, ω_1, and ω_2. Clearly any normal matrix with eigenvalues $1, \omega_1, \omega_2$ will have the same numerical range.

Example 6. If

$$M = \begin{pmatrix} 1 & x & 0 \\ 0 & \omega_1 & 0 \\ 0 & 0 & \omega_2 \end{pmatrix},$$

and $|x| \geqslant 3$, then $W(M)$ is an ellipse. If $0 < |x| < 3$, then $W(M)$ is a cone-shaped convex hull of an ellipse and a point external to the ellipse.

The facts about the example above follow directly from the conditions in Theorems 2.2 and 2.4 of [9].

Example 7. If

$$M = \begin{pmatrix} 1 & 2 & 2 \\ 0 & \omega_1 & 1 \\ 0 & 0 & \omega_2 \end{pmatrix},$$

then $W(M)$ has a flat portion on its boundary.

The facts in Example 7 follow from the conditions in Theorem 1.2 of [14].

Example 8. If x and y are both nonzero complex numbers and

$$M = \begin{pmatrix} 1 & x & y \\ 0 & \omega_1 & 0 \\ 0 & 0 & \omega_2 \end{pmatrix},$$

then $W(M)$ is ovular.

Since M in Example 8 is not normal, the numerical range of M is not the convex hull of its eigenvalues. The same theorems from [9] and [14] that were mentioned above show that $W(M)$ is not the convex hull of a point and an ellipse and has no flat part. The only remaining possibility is that $W(M)$ is ovular.

Our analysis of 3×3 matrices repeatedly uses the same functions of the entries of the matrix, so we define these quantities here. First, note that by Schur's Lemma, any 3×3 matrix with minimal polynomial $z^3 - 1$ is unitarily equivalent to an upper triangular matrix of the form below. Since numerical ranges are preserved under unitary equivalence, we can assume M equals this matrix:

$$M = \begin{pmatrix} 1 & a & b \\ 0 & \omega_1 & c \\ 0 & 0 & \omega_2 \end{pmatrix}, \tag{2}$$

where $\omega_1 = e^{i\frac{2\pi}{3}}$, $\omega_2 = e^{i\frac{4\pi}{3}}$ and a, b, and c are arbitrary complex numbers.

Let

$$H_\theta = \Re(e^{-i\theta}M) = \begin{pmatrix} \cos(\theta) & e^{-i\theta}\frac{a}{2} & e^{-i\theta}\frac{b}{2} \\ e^{i\theta}\frac{\overline{a}}{2} & \cos\left(\theta - \frac{2\pi}{3}\right) & e^{-i\theta}\frac{c}{2} \\ e^{i\theta}\frac{\overline{b}}{2} & e^{i\theta}\frac{\overline{c}}{2} & \cos\left(\theta - \frac{4\pi}{3}\right) \end{pmatrix}.$$

The support function of M is computed in terms of H_θ:

$$\begin{aligned} p_M(\theta) &= \sup\left\{\Re(e^{-i\theta}z) \mid z \in W(M)\right\} \\ &= \sup\left\{\Re(e^{-i\theta}\langle Mv, v\rangle) \mid v \in \mathbb{C}^3, \|v\| = 1\right\} \\ &= \sup\left\{\langle H_\theta v, v\rangle \mid v \in \mathbb{C}^3, \|v\| = 1\right\} \end{aligned}$$

and since H_θ is hermitian, the last supremum is the maximum eigenvalue of H_θ. That is, for every value of θ, the maximum root of the characteristic polynomial of H_θ is $p_M(\theta)$.

The characteristic polynomial of H_θ as a function of x is

$$q_\theta(x) = -x^3 + sx + t(\theta) \tag{3}$$

where

$$s = \frac{|a|^2 + |b|^2 + |c|^2 + 3}{4} \tag{4}$$

and $t(\theta) = \det H_\theta$. Directly computing this determinant leads to

$$\begin{aligned} t(\theta) = \cos(\theta)\cos\left(\theta - \frac{2\pi}{3}\right)\cos\left(\theta - \frac{4\pi}{3}\right) - \frac{|a|^2}{4}\cos\left(\theta - \frac{4\pi}{3}\right) \\ - \frac{|b|^2}{4}\cos\left(\theta - \frac{2\pi}{3}\right) - \frac{|c|^2}{4}\cos(\theta) + 2\Re e^{-i\theta}\frac{a\bar{b}c}{8}. \end{aligned}$$

Trigonometric identities show that $t(\theta)$ simplifies to:

$$t(\theta) = \frac{1}{4}\cos(3\theta) + f\cos(\theta) + g\sin(\theta), \tag{5}$$

with

$$f = \frac{1}{8}(|a|^2 + |b|^2 - 2|c|^2 + 2\Re a\bar{b}c) \tag{6}$$

$$g = \frac{1}{8}\left(\sqrt{3}|a|^2 - \sqrt{3}|b|^2 + 2\Im a\bar{b}c\right). \tag{7}$$

Further calculations show that we can find a formula for the support function of the numerical range for any 3×3 matrix of the form (2).

Proposition 9. *Let M be any 3×3 matrix of the form (2). The support function for M is*

$$p_M(\theta) = \frac{2}{\sqrt{3}}\sqrt{s}\cos\left(\frac{1}{3}\arccos\left(\frac{t(\theta)}{2}\sqrt{\frac{27}{s^3}}\right)\right), \tag{8}$$

where s and $t(\theta)$ are defined as in (4) and (5).

Proof. Fix $\theta \in [0, 2\pi)$. Substitution, or the Chebyshev cube root formula, shows that $p_M(\theta)$ is a root of the characteristic polynomial q_θ given in (3). To show that (8) delivers a formula for the support function, it remains to observe that $p_M(\theta)$ is the maximum root of q_θ (i.e. that $p_M(\theta)$ is the maximum eigenvalue of H_θ). The local maximum value of q_θ occurs at $x = \frac{\sqrt{s}}{\sqrt{3}}$ and the local minimum value of q_θ occurs at $x = -\frac{\sqrt{s}}{\sqrt{3}}$. Therefore, two distinct roots of q_θ cannot both be greater than or equal to $\frac{\sqrt{s}}{\sqrt{3}}$ or there would be another local extreme value at a location greater than $\frac{\sqrt{s}}{\sqrt{3}}$. Since the range of arccosine is $[0, \pi]$, the value of $p_M(\theta)$ is greater than or equal to $\frac{2\sqrt{s}}{\sqrt{3}}\frac{1}{2} = \frac{\sqrt{s}}{\sqrt{3}}$, so $p_M(\theta)$ is the maximum root. $\square$

3. Counterexample on an infinite-dimensional space

As discussed in the introduction, if an operator T is quadratic with distinct eigenvalues, then even if T is defined on an infinite dimensional space, the numerical range of T is an ellipse (possibly degenerate) with foci at the eigenvalues of T. Therefore the numerical range of T is not a disk since the foci are distinct. See [15], [19] or [2] for details. For example, if $T^2 = I$ (and $T \neq \pm I$), then the eigenvalues of T are exactly the values -1 and 1, and the numerical range of T is an ellipse with major axis on the x-axis and minor axis on the y-axis. Consequently, the maximum value of the support function of T is always attained at the angles $\theta = 0$ and $\theta = \pi$ and at no other values of θ.

In contrast, although an operator with minimal polynomial $z^3 - 1$ has eigenvalues equal to the cube roots of unity, there are no fixed angles at which the support function of $W(T)$ is always maximized. Even a 3×3 matrix M with $M^3 = I$ can have a support function for $W(M)$ which is maximized at any given angle. In Proposition 11, a collection of 3×3 matrices illustrating this fact is constructed. In Theorem 14, this collection is used to construct an operator T on an infinite dimensional Hilbert space that satisfies $T^3 = I$ and has numerical range equal to a disk.

Recall that in [2], the authors showed that a composition operator (with an elliptic automorphism as symbol) which satisfied $C_\varphi^2 = I$ could not have a disk as its numerical range because this operator is quadratic. Although our counterexample in this section shows that no such general argument can be used to prove that a composition operator satisfying $C_\varphi^3 = I$ does not have a disk as its numerical range, the special case of a composition operator is still open.

The first lemma is a technical result that permits creating a support function with absolute maximum at a given value by constructing a critical point at that value.

Lemma 10. *If $B > 9$ and β is in $[0, 2\pi)$, then the function $\tau(\theta) = \cos(3\theta) + B\cos(\theta - \beta)$ achieves its absolute maximum value on $[0, 2\pi)$ at exactly one point; namely, at the unique value of θ in $[0, 2\pi)$ where $\tau'(\theta) = 0$ and $\tau''(\theta) < 0$.*

Proof. Due to the symmetry about the value $\theta = \pi$, it suffices to show there is exactly one critical value θ_0 in $[0, \pi)$ because there is a one to one correspondence between critical values in $[0, \pi)$ and $[\pi, 2\pi)$ where absolute maxima in $[0, \pi)$ correspond to absolute minima in $[\pi, 2\pi)$ and vice versa. Furthermore, the argument is particularly straightforward if $\beta = 0$ or $\beta = \pi$, so we may fix $\beta \in (0, 2\pi)$ with $\beta \neq \pi$.

We wish to show that when $B > 9$, there is exactly one value of θ in $(0, \pi)$ such that

$$\tau'(\theta) = -3\sin(3\theta) - B\sin(\theta - \beta) = 0,$$

but this is equivalent to showing that when $M \in \left(0, \frac{1}{3}\right)$, there is exactly one value of θ in $(0, \pi)$ such that

$$\eta_M(\theta) = M\sin(3\theta) + \sin(\theta - \beta) = 0.$$

Define Ω to be the set of all $M \in \left(0, \frac{1}{3}\right)$ such that η_M has two or more zeroes in $(0, \pi)$. If Ω is not empty, then $M_0 = \inf \Omega$ is in $[0, \frac{1}{3})$. If M is sufficiently small (for example if $M < \frac{1}{\sqrt{10}}$), then η_M cannot have two roots in the same small interval because that would lead to a contradiction via Rolle's Theorem. Therefore $M_0 > 0$.

Now consider the function η_{M_0}. Since $\eta_{M_0}(0)$ and $\eta_{M_0}(\pi)$ have opposite signs, η_{M_0} has at least one root in $(0, \pi)$. If η_{M_0} has a double root θ_0, we are done, because then $\sin(\theta_0 - \beta) = -M_0\sin(3\theta_0)$ and $\cos(\theta_0 - \beta) = -3M_0\cos(3\theta_0)$, so we obtain the contradiction

$$1 = M_0^2\sin^2(3\theta_0) + 9M_0^2\cos^2(3\theta_0) = M_0^2 + 8M_0^2\cos^2(3\theta_0) < \frac{1}{9} + \frac{8}{9} = 1.$$

To see that η_{M_0} has a double root, note that if it has only one root in $(0, \pi)$, then there is a sequence M_n decreasing toward M_0 such that η_{M_n} has two distinct roots in $(0, \pi)$ for each n and a value $\tilde{\theta}$ between the distinct roots where $\eta'_{M_n}(\tilde{\theta}) = 0$. As n goes toward ∞, subsequential limits of these roots all coincide, which proves that the root of η_{M_0} is double.

If η_{M_0} has exactly two roots in $(0, \pi)$, then one of them must be a double root since η_{M_0} has opposite signs on endpoints. Finally, if η_{M_0} has three or more distinct single roots (i.e. roots where η_{M_0} has a sign change) in $(0, \pi)$, then a continuity argument shows that η_M must also have three or more sign changes for $M < M_0$ with M sufficiently close to M_0, contradicting the definition of M_0. Therefore the only possibility is that η_{M_0} has a double root, so Ω is empty. This proves the proposition. □

Proposition 11. *Let $\alpha \in [0, 2\pi)$ and let $a > \sqrt[3]{9}$. There exists $\beta \in [0, 2\pi)$ such that the maximum value of the support function of the numerical range of the matrix*

$$M(a, \beta) = \begin{pmatrix} 1 & ae^{i\beta} & a \\ 0 & \omega_1 & a \\ 0 & 0 & \omega_2 \end{pmatrix}, \tag{9}$$

is achieved at the value α. This maximum value is given by

$$\sqrt{a^2+1}\cos\left(\frac{1}{3}\arccos\left(\frac{\cos(3\alpha)+\sqrt{a^6-9\sin^2(3\alpha)}}{(\sqrt{a^2+1})^3}\right)\right).$$

Recall that the maximum value of $p_M(\theta)$ is the numerical radius of M.

Proof. First define functions u and v on $[0, 2\pi] \times (\sqrt[3]{9}, \infty)$ as follows.

$$u(\alpha, a) = -3\sin(\alpha)\sin(3\alpha) + \cos(\alpha)\sqrt{a^6 - 9\sin^2(3\alpha)},$$

and

$$v(\alpha, a) = 3\cos(\alpha)\sin(3\alpha) + \sin(\alpha)\sqrt{a^6 - 9\sin^2(3\alpha)}.$$

A straightforward computation shows that

$$u(\alpha, a)^2 + v(\alpha, a)^2 = a^6.$$

Now fix $\alpha \in [0, 2\pi)$ and $a > \sqrt[3]{9}$. Since $\sqrt{u(\alpha, a)^2 + v(\alpha, a)^2} = a^3$, we can set β equal to the angle in $[0, 2\pi)$ such that

$$u(\alpha, a) + iv(\alpha, a) = a^3(\cos(\beta) + i\sin(\beta)). \tag{10}$$

Multiplying the real and imaginary parts of Eq.(10) by $\cos(\alpha)$ and $\sin(\alpha)$ respectively and adding them, we obtain:

$$a^3\cos(\alpha - \beta) = \sqrt{a^6 - 9\sin^2(3\alpha)}. \tag{11}$$

Similarly, multiplying the real and imaginary parts of (10) by $\sin(\alpha)$ and $-\cos(\alpha)$:

$$a^3\sin(\alpha - \beta) = -3\sin(3\alpha). \tag{12}$$

Define $M(a, \beta)$ as in Eq. (9). Substituting the entries of (9) into the definitions (4) and (5) and then into the formula for (8) yields

$$p_{M(a,\beta)}(\theta) = \sqrt{a^2+1}\cos\left(\frac{1}{3}\arccos\left(\frac{\cos(3\theta)+a^3\cos(\theta-\beta)}{(\sqrt{a^2+1})^3}\right)\right). \tag{13}$$

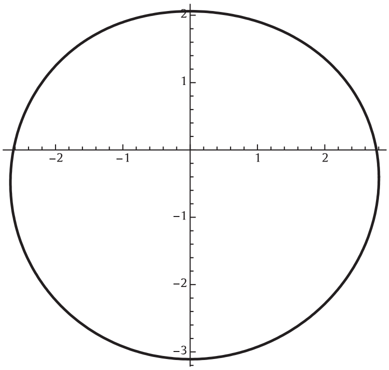

Fig. 1. $W\left(M\left(3, \frac{3\pi}{2}\right)\right)$.

The function $\cos\left(\frac{1}{3}\arccos(x)\right)$ is an increasing function of x. Therefore the function $p_{M(a,\beta)}(\theta)$ will achieve its maximum value on $[0, 2\pi)$ at the value of θ where

$$\tau(\theta) = \cos(3\theta) + a^3\cos(\theta - \beta)$$

is maximized. The identities (11) and (12) show that $\tau(\theta)$ satisfies $\tau'(\alpha) = 0$ and

$$\tau''(\alpha) = -9\cos(3\alpha) - \sqrt{a^6 - 9\sin^2(3\alpha)}.$$

The right side of the expression above is negative because $a > \sqrt[3]{9}$. Therefore $\tau(\theta)$ has a local maximum at $\theta = \alpha$.

Furthermore, the conditions required for Lemma 10 apply because $a^3 > 9$. Therefore $\tau(\theta)$ achieves its maximum at exactly one value in $[0, 2\pi)$, and since $p_{M(a,\beta)}(\theta)$ is maximized when $\tau(\theta)$ is maximized, this unique value must be $\theta = \alpha$.

The identity (11) also shows that $\tau(\alpha) = \cos(3\alpha) + \sqrt{a^6 - 9\sin^2(3\alpha)}$. Substituting $\theta = \alpha$ and this expression into (13) results in the formula for the maximum value of $p_{M(a,\beta)}(\alpha)$ as stated in the theorem. □

Example 12. If $a = 3$ and $\alpha = \frac{3\pi}{2}$, then (10) shows that

$$e^{i\beta} = \frac{1}{9} - \frac{4\sqrt{5}i}{9}$$

and the matrix

$$M\left(3, \frac{3\pi}{2}\right) = \begin{pmatrix} 1 & \frac{1}{9}(3 - 12\sqrt{5}i) & 3 \\ 0 & e^{\frac{2\pi i}{3}} & 3 \\ 0 & 0 & e^{\frac{4\pi i}{3}} \end{pmatrix}$$

has numerical range with maximum support value of approximately 3.10781 at $\theta = \frac{3\pi}{2}$ as shown in Fig. 1.

Proposition 13. *For any $\alpha \in [0, 2\pi)$ and any $x \in (\sqrt{9^{\frac{2}{3}} + 1}, \infty)$ there exists a 3×3 matrix M such that the minimal polynomial of M is $z^3 - 1$, the support function of M achieves its maximum value at α, and $p_M(\alpha) = x$. Furthermore, the operator norm of M satisfies $||M|| \leqslant 2x$.*

Proof. Let

$$m(a, \alpha) = \sqrt{a^2 + 1} \cos \left(\frac{1}{3} \arccos \left(\frac{\cos(3\alpha) + \sqrt{a^6 - 9 \sin^2(3\alpha)}}{(\sqrt{a^2 + 1})^3} \right) \right).$$

It is straightforward to verify that for any $\alpha \in [0, 2\pi)$, $m(a, \alpha)$ is an increasing function of a for $a \in \left(9^{\frac{1}{3}}, \infty\right)$. Furthermore, $m(a, \alpha)$ goes to infinity as a approaches infinity.

Fix $\alpha \in [0, 2\pi)$. If $x \in \left(\sqrt{9^{\frac{2}{3}} + 1}, \infty\right)$, then $a_0 = \sqrt{x^2 - 1} \in (9^{\frac{1}{3}}, \infty)$ and $m(a_0, \alpha) \leqslant \sqrt{a_0^2 + 1} = x$. By increasing a_0 to some value a, $x = m(a, \alpha)$ can be attained. Assume such an a that results in a maximum support value of x is produced. If the corresponding β is chosen by (10) and $M = M(a, \beta)$, then $M^3 = I$, the numerical radius of M is $p_M(\alpha) = x$ and since $\|M\| \leqslant 2\omega(M)$ (see [4]), it follows that $\|M\| \leqslant 2x$. $\square$

Theorem 14. *There exists a Hilbert space H and a bounded linear operator T on H such that $W(T)$ is an open disk centered at the origin and the minimal polynomial of T is $z^3 - 1$.*

Proof. We proceed in two steps.

Step 1. Let us construct a Hilbert space $\hat{H}$ and an operator $\hat{T}$ on $\hat{H}$ with minimal polynomial $z^3 - 1$ such that $W(\hat{T})$ is the union of an open disk centered at the origin and a set of points on the boundary of the disk. Define

$$\hat{H} = \mathbb{C}^3 \oplus \mathbb{C}^3 \oplus \mathbb{C}^3 \oplus \cdots$$

Fix $x > \sqrt{9^{\frac{2}{3}} + 1}$. Let $\{\alpha_n\}_{n=1}^{\infty}$ denote a dense collection of angles in $[0, 2\pi)$. For each α_n, let M_n denote the 3×3 matrix whose support function is guaranteed by Proposition 13 to attain a maximum value of x at α_n. Recall that since the spectrum of M_n consists of the cube roots of unity and the origin is in their convex hull, the origin is in each $W(M_n)$. Now define the block diagonal operator $\hat{T}$ on $\hat{H}$ by

$$\hat{T} = \begin{pmatrix} M_1 & 0 & 0 & 0 & 0 \\ 0 & M_2 & 0 & 0 & 0 \\ 0 & 0 & M_3 & 0 & 0 \\ 0 & 0 & 0 & M_4 & 0 \\ 0 & 0 & 0 & 0 & \ddots \end{pmatrix}.$$

Since each M_n has norm bounded by $2x$, the operator $\hat{T}$ is bounded on $\hat{H}$. Let $\mathrm{co}(\Omega)$ denote the convex hull of the set Ω in $\mathbb{C}$. Then (see, e.g. [13])

$$W(\hat{T}) = \mathrm{co}\left(\bigcup_{n=1}^{\infty} W(M_n) \right). \tag{14}$$

To see that the closure of the set (14) is a disk, note that for each n, M_n has compact numerical range containing the origin and with a maximum support function value of x at α_n. Consequently each $W(M_n)$ is contained in the closure of the disk $D(0, x)$ of radius x and center 0, and therefore the union $\bigcup_{n=1}^{\infty} W(M_n)$ and the closure of its convex hull are contained in $\overline{D(0, x)}$. This shows that $W(\hat{T}) \subseteq \overline{D(0, x)}$.

Conversely, if $z \in D(0, x)$, then $z = xre^{it}$ for $0 < r < 1$ and $t \in [0, 2\pi)$. Since $\{\alpha_n\}_{n=1}^{\infty}$ is dense in $[0, 2\pi)$, there exist α_k and α_m such that $\alpha_k < t \leqslant \alpha_m$ and $\cos(\alpha_m - \alpha_k) > 2r^2 - 1$. This choice guarantees that the minimum distance from the origin to the line segment from $xe^{i\alpha_m}$ to $xe^{i\alpha_k}$ (i.e. the magnitude of the midpoint of $xe^{i\alpha_m}$ and $xe^{i\alpha_k}$) is greater than $|z| = xr$, so z will be contained in the convex hull of the points 0, $xe^{i\alpha_m}$ and $xe^{i\alpha_k}$. Since these three points are all in the convex set $W(\hat{T})$, z must also be in $W(\hat{T})$. This proves that $D(0, x) \subseteq W(\hat{T})$. Therefore, $W(\hat{T})$ consists of the open disk $D(0, x)$ along with some of the boundary points of $D(0, x)$; namely the $\{\alpha_n\}$ values. Finally, the minimal polynomial of $\hat{T}$ is identical to the minimal polynomial of each block M_j, namely $z^3 - 1$.

Step 2. To obtain an operator T whose numerical range is an open disk, let $x > \sqrt{9^{\frac{2}{3}} + 1}$ and let $\{x_n\}$ be a increasing sequence in the interval $\left(\sqrt{9^{\frac{2}{3}} + 1}, x\right)$ that converges to x. By the construction described in Step 1, there exists a Hilbert space $\hat{H}_n$ and an operator $\hat{T}_n$ with the minimal polynomial $z^3 - 1$ such that $W(\hat{T}_n)$ is the union of the open disk $D(0, x_n)$ along with some of the boundary points of $D(0, x_n)$. Define the Hilbert space H by

$$H = \hat{H}_1 \oplus \hat{H}_2 \oplus \hat{H}_3 \oplus \ldots,$$

and define the operator T on H as a block diagonal operator in terms of the $\hat{T}_n$ operators:

$$T = \begin{pmatrix} \hat{T}_1 & 0 & 0 & 0 & 0 \\ 0 & \hat{T}_2 & 0 & 0 & 0 \\ 0 & 0 & \hat{T}_3 & 0 & 0 \\ 0 & 0 & 0 & \hat{T}_4 & 0 \\ 0 & 0 & 0 & 0 & \ddots \end{pmatrix}.$$

Clearly $D(0, x_n) \subset W(\hat{T}_n) \subset \overline{D(0, x_n)}$ for each positive integer n. Consequently,

$$\bigcup_{n=1}^{\infty} D(0, x_n) \subset \operatorname{co} \bigcup_{n=1}^{\infty} \left(W(\hat{T}_n)\right) \subset \bigcup_{n=1}^{\infty} \overline{D(0, x_n)}.$$

Since the left hand side and the right hand side of the latter chain of inclusions coincide with $D(0, x)$, so does the middle term, which in turn equals $W(T)$. As at step 1, the minimal polynomial of T is the same as the minimal polynomial of each block $\hat{T}_j$, and therefore still equals $z^3 - 1$. □

Note that the Hilbert space H in the previous theorem is separable. If the first part of the proof above is modified by replacing the dense set $\{\alpha_n\}$ with the entire boundary of the the disk $D(0, x)$, then a non-separable space H and an operator T on H could be constructed so that the minimal polynomial of T is $z^3 - 1$ and $W(T)$ is the closed disk $\overline{D(0, x)}$.

4. Threefold symmetry

The explicit formula (8) for the support function of a 3×3 matrix satisfying $M^3 = I$ also allows derivation of a simple condition that determines whether or not the numerical range $W(M)$ has a certain kind of symmetry about the origin.

Definition 15. A set S has threefold symmetry about the origin if $z \in S$ implies $e^{\frac{2\pi i}{3}} z \in S$.

Clearly the spectrum of a matrix M with minimal polynomial $z^3 - 1$ has threefold symmetry about the origin, but the numerical range might not as the examples from Section 2 show. The property of having n-fold symmetry about the origin is the obvious generalization of threefold symmetry.

Definition 16. Let M and C be $n \times n$ complex matrices. The C-numerical range of M is the set

$$W_C(M) = \{\mathrm{Tr}(CUMU^*) : U^*U = I\}.$$

The C-numerical range of a matrix is one of several generalizations of the classical numerical range. The classical numerical range satisfies the identity $W(M) = W_{E_{11}}(M)$ where E_{11} is the matrix with 1 in the upper left corner and zeroes elsewhere. Therefore any result that applies to $W_C(M)$ for all C also applies to the classical numerical range.

In [8], Li and Tsing proved a number of results about which $n \times n$ matrices (and general linear operators) have C-numerical ranges with different types of circular symmetry. For example, they showed that the C-numerical range of M has n-fold symmetry about the origin for all $n \times n$ matrices C if and only if M is unitarily equivalent to a special block matrix. These conditions are also equivalent to the unitary orbit of M having n-fold symmetry about the origin. Block versions of these results hold as well.

In the $n = 3$ case, their results show that a 3×3 matrix M that satisfies the condition $W_C(M)$ has threefold symmetry about the origin for all 3×3 matrices C if and only if M is unitarily equivalent to a matrix V of the form

$$V = \begin{pmatrix} 0 & 0 & p \\ q & 0 & 0 \\ 0 & r & 0 \end{pmatrix}. \tag{15}$$

Therefore, the results in [8] show that if there exist $p, q, r \in \mathbb{C}$ such that M is unitarily equivalent to V above, then $W_C(M)$ has has threefold symmetry about the origin for all 3×3 matrices C and therefore $W(M)$ has threefold symmetry about the origin. However, the results in [8] do not determine whether a 3×3 matrix M for which it is only known that its classical numerical range has threefold symmetry about the origin must be unitarily equivalent to a matrix of the form (15). Theorem 19 answers that question in the affirmative.

Proposition 17. *If the numerical range of a* 3×3 *matrix M has threefold symmetry about the origin but is not a disk, then the spectrum of M has threefold symmetry about the origin.*

Proof. Kippenhahn's classification shows that the only possible numerical ranges of 3×3 matrices with threefold symmetry about the origin are disks, equilateral triangles, or ovular shapes. By assumption, $W(M)$ is not a disk. If $W(M)$ is a triangle with threefold symmetry about the origin, then the eigenvalues of M are the vertices of the triangle and therefore the eigenvalues also have threefold symmetry. So it will suffice to prove this result when $W(M)$ is ovular, and in this case Kippenhahn showed that the associated curve as defined with (1) is irreducible and consists of two components: an outer portion and an inner portion. The outer component of the curve is the boundary of $W(M)$ and therefore has threefold symmetry about the origin since $W(M)$ does.

The matrix $\widetilde{M} = e^{i\frac{2\pi}{3}}M$ satisfies $W(\widetilde{M}) = e^{i\frac{2\pi}{3}}W(M) = W(M)$, so $W(\widetilde{M})$ is also ovular. The associated curve for $\widetilde{M}$ is defined as before and because $W(\widetilde{M})$ is ovular this curve is also irreducible. The outer portion of the associated curve of $\widetilde{M}$ is the boundary of $W(\widetilde{M}) = W(M)$. Since the outer portions of the associated curves for M and $\widetilde{M}$ (consisting of infinitely many points) coincide and the curves are irreducible, they must be the same curve. Therefore these curves have the same real foci. According to [10, Theorem 11], the real foci of the associated curve of a matrix are the eigenvalues of the matrix, so the eigenvalues of M and $\widetilde{M}$ are identical, which proves that the eigenvalues of M have threefold symmetry about the origin in the ovular case. □

In the proof of the main result in this section, a 3×3 matrix M is shown to be unitarily equivalent to a matrix of the form (15) by proving a sufficient collection of identities involving unitary invariants for 3×3 matrices. In general, two $n \times n$ matrices M and V are unitarily equivalent if $\mathrm{Tr}Y(M, M^*) = \mathrm{Tr}Y(V, V^*)$ for a sufficiently large collection of words $Y(s, t)$ in two noncommuting variables. When

$n = 3$, it was shown by Pearcy [12] that checking equality of traces for a certain collection of nine words is sufficient to guarantee unitary equivalence. This result was improved upon by Sibirskiĭ [16] (see also [17] and the related discussion in [7]) who brought the number of words down to seven. For convenience of reference, we state the result below.

Theorem 18. *The 3×3 matrices M and V are unitarily equivalent if and only if the following seven trace identities hold:*

$$\mathrm{Tr}(M) = \mathrm{Tr}(V), \tag{16a}$$

$$\mathrm{Tr}(M^2) = \mathrm{Tr}(V^2), \tag{16b}$$

$$\mathrm{Tr}(MM^*) = \mathrm{Tr}(VV^*), \tag{16c}$$

$$\mathrm{Tr}(M^3) = \mathrm{Tr}(V^3), \tag{16d}$$

$$\mathrm{Tr}(M^2M^*) = \mathrm{Tr}(V^2V^*), \tag{16e}$$

$$\mathrm{Tr}(M^2(M^*)^2) = \mathrm{Tr}(V^2(V^*)^2), \tag{16f}$$

$$\mathrm{Tr}(M^2(M^*)^2MM^*) = \mathrm{Tr}(V^2(V^*)^2VV^*). \tag{16g}$$

Furthermore, any proper subcollection of the preceding identities is not sufficient to guarantee unitary equivalence.

Theorem 19. *Let M be any 3×3 matrix. Assume $W(M)$ is not a disk. Then the following are equivalent:*

(i) *$W(M)$ has threefold symmetry about the origin.*
(ii) *$\mathrm{Tr}(M^2M^*) = 0$ and the spectrum $\sigma(M)$ has threefold symmetry about the origin.*
(iii) *There exist $p, q, r \in \mathbb{C}$ such that M is unitarily equivalent to the matrix V in (15).*

Proof. Condition (iii) implies condition (i) by the results in [8], so that we need only to establish the implications (i) $\to$ (ii) $\to$ (iii). This will first be done under the additional assumption that the minimal polynomial of M is $z^3 - 1$.

The general case will follow directly from this special case.

(i) $\to$ (ii): Assume M is a 3×3 matrix with minimal polynomial $z^3 - 1$ such that $W(M)$ has threefold symmetry about the origin. M can be represented in the form (2) and the support function for M is $p_M(\theta)$ as given in (8). The function $p_M(\theta)$ must satisfy

$$p_M(\theta) = p_M\left(\theta + \frac{2\pi}{3}\right)$$

for all $\theta \in [0, 2\pi)$. Cancelling injective composed functions in the expression for $p_M(\theta)$ results in $t(\theta) = t\left(\theta + \frac{2\pi}{3}\right)$ for all $\theta \in [0, 2\pi)$. With the use of formulas (6) and (7) we obtain

$$\frac{1}{4}\cos(3\theta) + f\cos(\theta) + g\sin(\theta) = \frac{1}{4}\cos(3\theta + 2\pi) + f\cos\left(\theta + \frac{2\pi}{3}\right) + g\sin\left(\theta + \frac{2\pi}{3}\right).$$

Clearly the $\cos(3\theta)$ terms cancel and the function $f\cos(\theta) + g\sin(\theta)$ has period greater than $\frac{2\pi}{3}$ unless $f = g = 0$.

Therefore

$$0 = \frac{1}{8}(|a|^2 + |b|^2 - 2|c|^2 + 2\Re a\bar{b}c), \text{ and}$$

$$0 = \frac{1}{8}\left(\sqrt{3}|a|^2 - \sqrt{3}|b|^2 + 2\Im a\bar{b}c\right).$$

Computing directly from (2) yields

$$\begin{aligned}\mathrm{Tr}(M^2M^*) &= -\omega_2|a|^2 - \omega_1|b|^2 - |c|^2 + a\bar{b}c \\ &= \frac{1}{2}\left(|a|^2 + |b|^2 - 2|c|^2 + 2\Re a\bar{b}c\right) + i\frac{1}{2}\left(\sqrt{3}|a|^2 - \sqrt{3}|b|^2 + 2\Im a\bar{b}c\right) \\ &= 4f + i4g \\ &= 0.\end{aligned}$$

Since $\sigma(M)$ consists of the cube roots of unity, the spectrum has threefold symmetry about the origin so (i) $\rightarrow$ (ii) is established in the special case that M is a 3×3 matrix with minimal polynomial $z^3 - 1$.

(ii) $\rightarrow$ (iii): Now assume M is a 3×3 matrix with minimal polynomial $z^3 - 1$ such that $\mathrm{Tr}(M^2M^*) = 0$. In [1], it is shown that every 3×3 matrix is unitarily equivalent to a matrix of the form

$$M = \begin{pmatrix} \lambda_1 & 0 & x \\ y & \lambda_2 & 0 \\ 0 & z & \lambda_3 \end{pmatrix}.$$

If $M^3 = I$, then

$$\lambda_1^3 + xyz = \lambda_2^3 + xyz = \lambda_3^3 + xyz = 1, \tag{17}$$

and

$$\begin{aligned}(\lambda_1 + \lambda_2 + \lambda_3)xy = (\lambda_1 + \lambda_2 + \lambda_3)xz = (\lambda_1 + \lambda_2 + \lambda_3)yz = 0, \\ (\lambda_1^2 + \lambda_1\lambda_3 + \lambda_3^2)x = (\lambda_1^2 + \lambda_1\lambda_2 + \lambda_2^2)y = (\lambda_2^2 + \lambda_2\lambda_3 + \lambda_3^2)z = 0.\end{aligned} \tag{18}$$

The equations in (17) show that $|\lambda_i| = |\sqrt[3]{1 - xyz}|$ for $i = 1, 2, 3$, so the λ_i values all have the same magnitude.

If any of the values x, y, or z is zero, then $\lambda_i^3 = 1$ for $i = 1, 2, 3$ and the lambda values are the eigenvalues of M which are the distinct cube roots of unity by the minimal polynomial hypothesis. By assumption,

$$0 = \mathrm{Tr}(M^2M^*) = |\lambda_1|^2\lambda_1 + |\lambda_2|^2\lambda_2 + |\lambda_3|^2\lambda_3 + |x|^2(\lambda_1 + \lambda_3) + |y|^2(\lambda_1 + \lambda_2) + |z|^2(\lambda_2 + \lambda_3),$$

from which it follows that $x = y = z = 0$. Therefore, if $xyz = 0$ then M is normal and unitarily equivalent to V with $p = 1$, $q = \omega_1$, and $r = \omega_2$. Thus in this special case (iii) holds.

Therefore we may assume without loss of generality that $xyz \neq 0$. In this case the equations in (18) show that there exists $\xi \in \mathbb{C}$ such that $\lambda_1 = \xi$, $\lambda_2 = \xi\omega_1$, and $\lambda_3 = \xi\omega_2$.

If $\xi = 0$, then we are done because M is already of form V with $p = x$, $q = y$, and $r = z$. Thus, we may assume

$$M = \begin{pmatrix} \xi & 0 & x \\ y & \xi\omega_1 & 0 \\ 0 & z & \xi\omega_2 \end{pmatrix},$$

with $\xi \neq 0$ and $xyz \neq 0$.

We will show that there exists a matrix V of form (15) such that each of the seven corresponding unitary invariants in Theorem 18 are equal for V and M. The associated matrices that are required for the trace calculations are computed below:

$$M^2 = \begin{pmatrix} \xi^2 & xz & -x\xi\omega_1 \\ -y\xi\omega_2 & \xi^2\omega_2 & xy \\ zy & -z\xi & \xi^2\omega_1 \end{pmatrix},$$

$$M^3 = I,$$

$$MM^* = \begin{pmatrix} |\xi|^2 + |x|^2 & \bar{y}\xi & x\bar{\xi}\omega_1 \\ y\bar{\xi} & |\xi|^2 + |y|^2 & \bar{z}\xi\omega_1 \\ \bar{x}\xi\omega_2 & z\bar{\xi}\omega_2 & |\xi|^2 + |z|^2 \end{pmatrix}, \tag{19}$$

$$M^2M^* = \begin{pmatrix} \xi(|\xi|^2 - |x|^2\omega_1) & \bar{y}\xi^2 + xz\bar{\xi}\omega_2 & x(|z|^2 - |\xi|^2\omega_2) \\ y(|x|^2 - |\xi|^2\omega_2) & \xi(|\xi|^2\omega_1 - |y|^2\omega_2) & \bar{z}\xi^2\omega_2 + xy\bar{\xi}\omega_1 \\ yz\bar{\xi} + \bar{x}\xi^2\omega_1 & z(|y|^2 - |\xi|^2\omega_2) & \xi(|\xi|^2\omega_2 - |z|^2) \end{pmatrix}, \tag{20}$$

and finally $M^2(M^*)^2 =$

$$\begin{pmatrix} |\xi|^4 + |zx|^2 + |\xi x|^2 & (-\bar{y}\xi|\xi|^2 - \bar{y}\xi|x|^2 + xz\bar{\xi}^2)\omega_1 & -x\bar{\xi}|\xi|^2 - x\bar{\xi}|z|^2 + \overline{yz}\xi^2 \\ (-y\bar{\xi}|\xi|^2 - y\bar{\xi}|x|^2 + \overline{xz}\xi^2)\omega_2 & |\xi|^4 + |xy|^2 + |\xi y|^2 & (-\bar{z}\xi|\xi|^2 - \bar{z}\xi|y|^2 + xy\bar{\xi}^2)\omega_2 \\ -\bar{x}\xi|\xi|^2 - \bar{x}\xi|z|^2 + yz\bar{\xi}^2 & (-z\bar{\xi}|\xi|^2 - z\bar{\xi}|y|^2 + \overline{xy}\xi^2)\omega_1 & |\xi|^4 + |yz|^2 + |\xi z|^2 \end{pmatrix}. \tag{21}$$

Also,

$$V^2 = \begin{pmatrix} 0 & pr & 0 \\ 0 & 0 & pq \\ rq & 0 & 0 \end{pmatrix},$$

$$V^3 = pqrI,$$

$$VV^* = \begin{pmatrix} |p|^2 & 0 & 0 \\ 0 & |q|^2 & 0 \\ 0 & 0 & |r|^2 \end{pmatrix},$$

$$V^2V^* = \begin{pmatrix} 0 & 0 & p|r|^2 \\ q|p|^2 & 0 & 0 \\ 0 & r|q|^2 & 0 \end{pmatrix},$$

and

$$V^2(V^*)^2 = \begin{pmatrix} |pr|^2 & 0 & 0 \\ 0 & |pq|^2 & 0 \\ 0 & 0 & |rq|^2 \end{pmatrix}.$$

Also note that

$$\mathrm{Tr}(V^2(V^*)^2VV^*) = |p|^4|r|^2 + |q|^4|p|^2 + |r|^4|q|^2.$$

No matter what values of p, q, and r are chosen for V, three of the unitary invariants for M and V corresponding to (16a), (16b), and (16e) are automatically equal:

$$\mathrm{Tr}(M) = \xi + \xi\omega_1 + \xi\omega_2 = 0 = \mathrm{Tr}(V),$$

$$\mathrm{Tr}(M^2) = \xi^2 + \xi^2\omega_2 + \xi^2\omega_1 = 0 = \mathrm{Tr}(V^2),$$

and by assumption

$$\mathrm{Tr}(M^2M^*) = 0 = \mathrm{Tr}(V^2V^*).$$

Since

$$\mathrm{Tr}(M^2M^*) = \xi(-|x|^2\omega_1 - |y|^2\omega_2 - |z|^2) = \xi\left(\frac{1}{2}(|x|^2 + |y|^2 - 2|z|^2) + \frac{i}{2}(|y|^2 - |x|^2)\right),$$

the assumption $\mathrm{Tr}(M^2M^*) = 0$ also implies that $|y| = |z| = |x|$. Since (17) implies $xyz = 1 - \xi^3$, it follows that

$$|x|^6 = |1 - \xi^3|^2 = 1 - \overline{\xi^3} - \xi^3 + |\xi|^6. \tag{22}$$

The condition $|y| = |x| = |z|$ also simplifies the calculation of the last trace involving M and M^*, because when the diagonal entries of the product $M^2(M^*)^2MM^*$ are computed from (19) and (21) with $|x|$ replacing the values $|z|$ and $|y|$ everywhere, it turns out that each diagonal entry of $M^2(M^*)^2MM^*$ is

$$\begin{aligned}&(|\xi|^2 + |x|^2)(|\xi|^4 + |\xi x|^2 + |x|^4) - |\xi x|^2(|x|^2 + |\xi|^2)\omega_1 - |\xi x|^2(|x|^2 + |\xi|^2)\omega_2\\ &\quad + \overline{\xi}^3(1 - \xi^3)\omega_1 + \xi^3(1 - \overline{\xi}^3)\omega_2,\end{aligned}$$

which simplifies to

$$\begin{aligned}&(|\xi|^2 + |x|^2)(|\xi|^4 + 2|\xi x|^2 + |x|^4) + \overline{\xi}^3(1 - \xi^3)\omega_1 + \xi^3(1 - \overline{\xi}^3)\omega_2\\ &\quad = (|\xi|^2 + |x|^2)^3 + \overline{\xi}^3(1 - \xi^3)\omega_1 + \xi^3(1 - \overline{\xi}^3)\omega_2.\end{aligned}$$

Therefore

$$\begin{aligned}\mathrm{Tr}(M^2(M^*)^2MM^*) &= 3((|\xi|^2 + |x|^2)^3 + \overline{\xi}^3(1 - \xi^3)\omega_1 + \xi^3(1 - \overline{\xi}^3)\omega_2)\\ &= 3(|x|^2 + |\xi|^2)^3 + 3|\xi|^6 + 3\omega_1\overline{\xi}^3 + 3\omega_2\xi^3.\end{aligned} \tag{23}$$

It is now possible to show that there exist values of p, q, and r that make the remaining four invariants for V equal to those for M. To do so, form the polynomial

$$h(t) = t^3 - 3(|\xi|^2 + |x|^2)t^2 + 3(|\xi|^4 + |\xi x|^2 + |x|^4)t - 1.$$

The cubic polynomial h has local extreme values at the critical values $|\xi|^2 \pm |\xi x| + |x|^2$, which are both positive, and $h(0) = -1$. In addition, $h(|\xi|^2 - |\xi x| + |x|^2) = -1 + (|\xi|^3 + |x|^3)^2$, which is non-negative, and $h(|\xi|^2 + |\xi x| + |x|^2) = -1 + (|\xi|^3 - |x|^3)^2$, which is non-positive. Therefore h must have a root between $t = 0$ and the critical value $t = |\xi|^2 - |\xi x| + |x|^2$, another root between the two critical values, and (because $h(t)$ goes to infinity as t goes to infinity) its third root greater than the critical value $t = |\xi|^2 + |\xi x| + |x|^2$. If one of the extreme values is zero, then h has a double root, but all roots are still positive. If these three positive roots of h are denoted u, v, and w, then

$$\begin{aligned} t^3 - 3(|\xi|^2 + |x|^2)t^2 + 3(|\xi|^4 + |\xi x|^2 + |x|^4)t - 1 &= (t-u)(t-v)(t-w) \\ &= t^3 - (u+v+w)t^2 \\ &\quad +(uv+uw+vw)t - uvw. \end{aligned}$$

Therefore there exist positive numbers u, v and w (uniquely determined up to permutation) such that

$$u+v+w = 3(|\xi|^2+|x|^2), \quad uvw = 1, \quad \text{and} \quad uv+uw+vw = 3\left(|\xi|^4 + |\xi x|^2 + |x|^4\right). \tag{24}$$

If $p = \sqrt{u}$, $q = \sqrt{v}$, and $r = \sqrt{w}$, then these are exactly the trace identities needed to satisfy (16c), (16d), and (16f), respectively. It follows that there exists V of form (15) such that the first six invariants (16a) through (16f) hold. It remains to show our assumptions, now including equality of these first six traces, imply (16g), i.e. that

$$\mathrm{Tr}(M^2(M^*)^2MM^*) = \mathrm{Tr}(V^2(V^*)^2VV^*)$$

as well. Fix any cube root of $\xi^3 - 1$ and let $\rho = \overline{\xi}\sqrt[3]{\xi^3-1}$, so $\overline{\rho} = \xi\overline{\sqrt[3]{\xi^3-1}} = \xi\sqrt[3]{\overline{\xi}^3-1}$ if the cube root of $\overline{\xi}^3 - 1$ is chosen consistently. Therefore $\rho\overline{\rho} = |\xi|^2|x|^2$ and $\rho^3 + \overline{\rho}^3 = |\xi|^6 + |x|^6 - 1$ by (22). It is straightforward, although tedious, to check that u, v, and w defined by

$$\begin{aligned} u &= |\xi|^2 + |x|^2 - \rho - \overline{\rho}, \\ v &= |\xi|^2 + |x|^2 - \omega_2\rho - \omega_1\overline{\rho}, \\ w &= |\xi|^2 + |x|^2 - \omega_1\rho - \omega_2\overline{\rho}, \end{aligned} \tag{25}$$

satisfy all the equations in (24), so these are the three positive roots of the polynomial h.

With these values determined, it follows that

$$\begin{aligned} u^2 &= (|\xi|^2 + |x|^2)^2 - 2(\rho+\overline{\rho})(|\xi|^2+|x|^2) + (\rho+\overline{\rho})^2, \\ v^2 &= (|\xi|^2 + |x|^2)^2 - 2(\omega_2\rho+\omega_1\overline{\rho})(|\xi|^2+|x|^2) + (\omega_2\rho+\omega_1\overline{\rho})^2, \\ w^2 &= (|\xi|^2 + |x|^2)^2 - 2(\omega_1\rho+\omega_2\overline{\rho})(|\xi|^2+|x|^2) + (\omega_1\rho+\omega_2\overline{\rho})^2, \end{aligned}$$

and

$$\begin{aligned} u^2w &= (|\xi|^2+|x|^2)^3 - (2(\rho+\overline{\rho}) + \omega_1\rho + \omega_2\overline{\rho})(|\xi|^2+|x|^2)^2 \\ &\quad + \left(2(\rho+\overline{\rho})(\omega_1\rho+\omega_2\overline{\rho}) + (\rho+\overline{\rho})^2\right)(|\xi|^2+|x|^2) - (\rho+\overline{\rho})^2(\omega_1\rho+\omega_2\overline{\rho}), \\ v^2u &= (|\xi|^2+|x|^2)^3 - (2(\omega_2\rho+\omega_1\overline{\rho}) + \rho+\overline{\rho})(|\xi|^2+|x|^2)^2 \\ &\quad + \left(2(\rho+\overline{\rho})(\omega_2\rho+\omega_1\overline{\rho}) + (\omega_2\rho+\omega_1\overline{\rho})^2\right)(|\xi|^2+|x|^2) - (\rho+\overline{\rho})(\omega_2\rho+\omega_1\overline{\rho})^2, \\ w^2v &= (|\xi|^2+|x|^2)^3 - (2(\omega_1\rho+\omega_2\overline{\rho}) + (\omega_2\rho+\omega_1\overline{\rho}))(|\xi|^2+|x|^2)^2 \\ &\quad + \left(2(\omega_1\rho+\omega_2\overline{\rho})(\omega_2\rho+\omega_1\overline{\rho}) + (\omega_1\rho+\omega_2\overline{\rho})^2\right)(|\xi|^2+|x|^2) \\ &\quad -(\omega_2\rho+\omega_1\overline{\rho})(\omega_1\rho+\omega_2\overline{\rho})^2. \end{aligned}$$

Therefore the final invariant for V satisfies

$$\mathrm{Tr}(V^2(V^*)^2VV^*) = u^2w+v^2u+w^2v = 3(|\xi|^2+|x|^2)^3 + d_2(|\xi|^2+|x|^2)^2 + d_1(|\xi|^2+|x|^2) + d_0,$$

where

$$d_2 = -3(\omega_2\rho + \omega_1\overline{\rho}) - 3(\omega_1\rho + \omega_2\overline{\rho}) - 3(\rho + \overline{\rho}) = 0,$$

$$d_1 = ((\omega_2\rho + \omega_1\overline{\rho}) + (\omega_1\rho + \omega_2\overline{\rho}) + (\rho + \overline{\rho}))^2 = 0,$$

and

$$d_0 = -3\omega_1\rho^3 - 3\omega_2\overline{\rho}^3.$$

Using (22) again,

$$\begin{aligned}\mathrm{Tr}(V^2(V^*)^2VV^*) &= 3(|\xi|^2 + |x|^2)^3 - 3\omega_1\rho^3 - 3\omega_2\overline{\rho}^3 \\ &= 3(|\xi|^2 + |x|^2)^3 + 3|\xi|^6 + 3\omega_1\overline{\xi}^3 + 3\omega_2\xi^3 = \mathrm{Tr}(M^2(M^*)^2MM^*).\end{aligned}$$

The last equality follows from (23). This proves the equivalence of the last invariant (16g) for V and for M. Therefore when $p = \sqrt{u}$, $q = \sqrt{v}$, and $r = \sqrt{w}$ are defined as in (25), all seven of the traces that are needed to prove V and M are unitarily equivalent are equal. Therefore M is unitarily equivalent to a matrix of the form (15).

This concludes the proof of (i) $\to$ (ii) $\to$ (iii) when M is a 3×3 matrix with minimal polynomial $z^3 - 1$.

For the general case of (i) $\to$ (ii), assume N is any nonzero 3×3 matrix whose numerical range has threefold symmetry about the origin. Then the spectrum of N has threefold symmetry about the origin by Proposition 17, and the spectrum therefore consists of points $\lambda, \lambda\omega_1, \lambda\omega_2$ for some $\lambda \in \mathbb{C}$.

If $\lambda = 0$ then N would have a triple eigenvalue, but we will show this is not possible. By Theorem 4.1 from [9] along with our assumption that $W(N)$ is a non-disk with threefold symmetry about the origin, it follows that N is unitarily equivalent to, and can therefore be assumed to have the form

$$N = \begin{pmatrix} 0 & x & y \\ 0 & 0 & z \\ 0 & 0 & 0 \end{pmatrix} \tag{26}$$

with $xyz \neq 0$ and $|x|, |y|, |z|$ not all equal. In this case $N = H + iK$ with $H = \frac{N+N^*}{2}$ and $K = \frac{N-N^*}{2i}$. Since $W(N)$ has threefold symmetry about the origin, the support function $p_N(\theta)$ has period $\frac{2\pi}{3}$. Recall that $p_N(\theta)$ is the maximum eigenvalue of the hermitian matrix $H_\theta = \cos(\theta)H + \sin(\theta)K$. Since the trace of H_θ is zero and the maximum eigenvalue of H_θ is the negative of the minimum eigenvalue of $H_{\theta+\pi}$, the periodicity requirement for $p_N(\theta)$ forces all three eigenvalues of H_θ to equal the corresponding three eigenvalues of $H_{\theta+\frac{2\pi}{3}}$. Therefore

$$\det(H_\theta) = \det\left(H_{\theta+\frac{2\pi}{3}}\right)$$

for all θ. However, a straightforward computation shows that

$$\det(H_\theta) = \frac{1}{4}\Re e^{i\theta}\overline{x}y\overline{z}.$$

This function of θ cannot have period $\frac{2\pi}{3}$ if $\overline{x}y\overline{z} \neq 0$ which must hold since $xyz \neq 0$. Therefore if $W(N)$ has threefold symmetry then N cannot have a triple eigenvalue of 0.

So we may assume $\lambda \neq 0$ and therefore $N = \lambda M$ where the minimal polynomial of M is $z^3 - 1$. Clearly the set $W(M) = \frac{1}{\lambda} W(N)$ also has threefold symmetry about the origin. By the $M^3 = I$ case, this implies that $\mathrm{Tr}(M^2 M^*) = 0$, so $\mathrm{Tr}(N^2 N^*) = |\lambda|^2 \lambda \mathrm{Tr}(M^2 M^*) = 0$.

For the general proof that (ii) implies (iii), assume N is a 3×3 matrix where $\mathrm{Tr}(N^2 N^*) = 0$ and $\sigma(N)$ has threefold symmetry about the origin. In this case either N has a triple eigenvalue of zero or else $N = \lambda M$ where $\lambda \neq 0$ and the minimal polynomial of M is $z^3 - 1$. If $\lambda = 0$ is a triple eigenvalue, then N has the form (26). In this case it is easy to compute that $\mathrm{Tr}(N^2 N^*) = x\bar{y}z$. If $x\bar{y}z = 0$ then $W(N)$ is a disk by Theorem 4.1 of [9], contradicting our assumption. So $N = \lambda M$ and $\mathrm{Tr}(M^2 M^*) = \frac{1}{|\lambda|^2 \lambda} \mathrm{Tr}(N^2 N^*) = 0$. Therefore M is unitarily equivalent to some matrix V of the form (15), but this implies $N = \lambda M$ is unitarily equivalent to λV, which is also of form (15). This concludes the proof of the theorem. $\square$

Example 20. For an example illustrating the conditions in the previous theorem, note that the matrix

$$M_1 = \frac{1-i}{2} \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} & -1 \\ 2\sqrt{2} & 0 & 2\sqrt{2} \\ 1 & -\frac{1}{\sqrt{2}} & -1 \end{pmatrix}$$

is unitarily equivalent to the matrix

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

so M_1 satisfies part *(iii)* of the theorem, and

$$M_1^2 M_1^* = \begin{pmatrix} \frac{1}{4} - \frac{i}{4} & (2-2i)\sqrt{2} & -\frac{1}{4} + \frac{i}{4} \\ (2-2i)\sqrt{2} & 0 & (2-2i)\sqrt{2} \\ \frac{1}{4} - \frac{i}{4} & (-2+2i)\sqrt{2} & -\frac{1}{4} + \frac{i}{4} \end{pmatrix},$$

so $\mathrm{Tr}(M_1^2 M_1^*) = 0$. In addition, the spectrum of M_1 is $\left\{1-i, e^{i\frac{2\pi}{3}}(1-i), e^{i\frac{4\pi}{3}}(1-i)\right\}$, so part *(ii)* of the theorem holds. The plot of the boundary of $W(M_1)$ appears in Fig. 2.

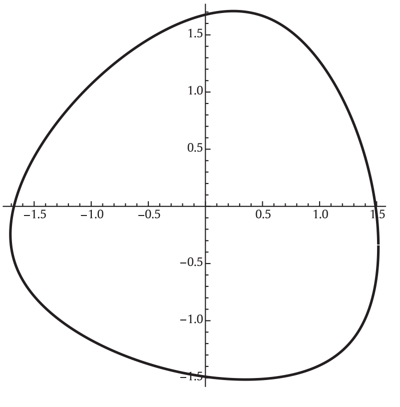

Fig. 2. $W(M_1)$.

Remark. A matrix M can satisfy $\mathrm{Tr}(M^2M^*) = 0$ without having numerical range with threefold symmetry about the origin if $\sigma(M)$ does not have threefold symmetry about the origin. For example, the matrix

$$M = \begin{pmatrix} 0 & -3 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

satisfies $\mathrm{Tr}(M^2M^*) = 0$ but $\sigma(M) = \{0, 0, 1\}$ and $W(M)$ does not have threefold symmetry about the origin.

Acknowledgements

The authors thank P.Y. Wu for making his preprint [20] available and for useful discussions which helped to remove some redundancies. The first four authors received support from Cal Poly's College Based Fee.

References

[1] J. An, D.Ž. Doković, Universal subspaces for compact Lie groups, J. Reine Angew. Math. 647 (2010) 151–173.
[2] P.S. Bourdon, J.H. Shapiro, The numerical ranges of automorphic composition operators, J. Math. Anal. Appl. 251 (2000) 839–854.
[3] M.-T. Chien, L. Yeh, Y.-T. Yeh, On geometric properties of the numerical range, Linear Algebra Appl. 274 (1998) 389–410.
[4] K.E. Gustafson, D.K.M. Rao, Numerical Range: The Field of Values of Linear Operators and Matrices, Springer-Verlag, New York, 1996.
[5] F. Hausdorff, Der Wertvorrat einer Bilinearform, Math. Z. 3 (1919) 314–316.
[6] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge, 1991.
[7] D.Ž. Doković, C.R. Johnson, Unitarily achievable zero patterns and traces of words in A and A*, Linear Algebra Appl. 421 (2007) 63–68.
[8] C.K. Li, N.K. Tsing, Matrices with circular symmetry on their unitary orbits and C-numerical ranges, Proc. Amer. Math. Soc. 111 (1) (1991) 19–28.
[9] D. Keeler, L. Rodman, I. Spitkovsky, The numerical range of 3×3 matrices, Linear Algebra Appl. 252 (1997) 115–139.
[10] R. Kippenhahn, Über den Wertevorrat einer Matrix, Math. Nachr. 6 (1951) 193–228.
[11] J. Maroulas, P.J. Psarrokos, M.J. Tsatsomeros, Perron–Frobenius type results on the numerical range, Linear Algebra Appl. 348 (2002) 49–62.
[12] C. Pearcy, A complete set of unitary invariants for 3×3 complex matrices, Trans. Amer. Math. Soc. 104 (1962) 425–429.
[13] M. Pollack, Numerical range and convex sets, Canad. Math. Bull. 17 (1974) 295–296.
[14] L. Rodman, I. Spitkovsky, 3×3 matrices with a flat portion on the boundary of the numerical range, Linear Algebra Appl. 397 (2005) 193–207.
[15] L. Rodman, I. Spitkovsky, On generalized numerical ranges of quadratic operators, Recent Advances in Matrix and Operator Theory, Oper. Theory Adv. Appl., vol. 179, Birkhäuser, Basel, 2008, pp. 241–256.
[16] K.S. Sibirskiĭ, A minimal polynomial basis of unitary invariants of a square matrix of order three, Mat. Zametki 3 (1968) 291–295 (in Russian).
[17] K.S. Sibirskiĭ, Algebraic Invariants of Differential Equations and Matrices, Stiintsa, Kishinev, 1976 (in Russian).
[18] O. Toeplitz, Das algebraische Analogon zu einem Satz von Fejér, Math. Z. 2 (1918) 187–197.
[19] S.-H. Tso, P.Y. Wu, Matricial ranges of quadratic operators, Rocky Mountain J. Math. 29 (3) (1999).
[20] P.Y. Wu, Numerical ranges as circular discs, preprint.