2011

The ratio field of values

Brian Lins

Eduard Einstein

Charles R. Johnson
William & Mary, crjohn@WM.EDU

Ilya Spitkovsky
William & Mary, imspitkovsky@gmail.com

Recommended Citation

Lins, Brian; Einstein, Eduard; Johnson, Charles R.; and Spitkovsky, Ilya, The ratio field of values (2011). *Linear Algebra and Its Applications*, 434(4), 1119-1136.
10.1016/j.laa.2010.10.022

Linear Algebra and its Applications 434 (2011) 1119–1136

Contents lists available at ScienceDirect

Linear Algebra and its Applications

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

The ratio field of values ☆

Eduard Einstein [a], Charles R. Johnson [b], Brian Lins [c,*], Ilya Spitkovsky [b]

[a] Pomona College, USA
[b] The College of William & Mary, USA
[c] Hampden-Sydney College, USA

ARTICLE INFO

Article history:
Received 12 April 2010
Accepted 26 October 2010
Available online 30 November 2010

Submitted by M. Tsatsomeros

AMS classification:
Primary: 15A60

Keywords:
Field of values
Generalized eigenvalue problem
Numerical range
Matrix pencil
Matrix polynomial

ABSTRACT

The ratio field of values, a generalization of the classical field of values to a pair of n-by-n matrices, is defined and studied, primarily from a geometric point of view. Basic functional properties of the ratio field are developed and used. A decomposition of the ratio field into line segments and ellipses along a master curve is given and this allows computation. Primary theoretical results include the following. It is shown (1) for which denominator matrices the ratio field is always convex, (2) certain other cases of convex pairs are given, and (3) that, at least for $n = 2$, the ratio field obeys a near convexity property that the intersection with any line segment has at most n components. Generalizations of the ratio field of values involving more than one matrix in both the numerator and denominator are also investigated. It is shown that generally such extensions need not be convex or even simply connected.

1. Introduction

The classical field of values of a single matrix $A \in M_n(\mathbb{C})$ is defined by

$$F(A) \equiv \{x^*Ax : x \in \mathbb{C}, \ x^*x = 1\}$$

It has been long and deeply studied (see e.g. [4]). Recently, an analogous field for *two* matrices has arisen in a numerical application [6]. It may also be a natural tool for generalized eigenvalue problems,

☆ This work was partially supported by NSF Grant DMS-0751964.
* Corresponding author.
E-mail addresses: Ece02006@mymail.pomona.edu (E. Einstein), crjohnso@math.wm.edu (C.R. Johnson), blins@hsc.edu (B. Lins), ilya@math.wm.edu (I. Spitkovsky).

0024-3795/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2010.10.022

since similar tools have been used to study eigenvalue problems as in [11]. For $A, B \in M_n(\mathbb{C})$ with $0 \notin F(B)$ we call:

$$R(A,B) = \left\{ \frac{x^*Ax}{x^*Bx} : x \in \mathbb{C}^n,\ x^*x = 1 \right\}$$

the **ratio field of values of** A **and** B**.** Of course,

$$R(A,B) = \left\{ \frac{x^*Ax}{x^*Bx} : x \in \mathbb{C}^n,\ x \neq 0 \right\}$$

is an equivalent description. The ratio field of values turns out to be a special case of the numerical range of a matrix polynomial as introduced by Li and Rodman [7]. Let $P(\lambda) = A_0 + A_1\lambda + \cdots + A_m\lambda^m$ where each $A_i \in M_n(\mathbb{C})$. The **numerical range of** $P(\lambda)$ is the set

$$W(P) = \{\lambda \in \mathbb{C} : 0 \in F(P(\lambda))\}.$$

If $P(\lambda)$ is the matrix pencil $A - \lambda B$, then $R(A,B) = W(A - \lambda B)$ if and only if $0 \notin F(B)$. The numerical range of matrix pencils has been studied by several authors [10,9,1,2]. A related study of numerical ranges in indefinite inner product spaces appears in [8]. Our purpose here is to develop theory for the ratio field, primarily its geometry.

In the next section, we mention several elementary properties of $R(A,B)$, many of which will be used frequently, and most of which are the appropriate analogues of properties of the usual field. One of these is the simultaneous congruential invariance of $R(A,B)$. Then, using the congruential canonical form for B [5], with $0 \notin F(B)$, we may give a congruential canonical form for the pair A, B that is crucial for the development of the theory of $R(A,B)$. In Section 3, we give a parametric description of $R(A,B)$, in general; in case $n = 2$, this description shows that $R(A,B)$ is the union of ellipses with centers lying on a curve that will either be a circular arc or a line segment. This observation provides a valuable tool for the remaining analysis.

While the usual field of values is always convex, the ratio field may or may not be convex. We call an ordered pair $A, B \in M_n(\mathbb{C})$, a **convex pair** if $R(A,B)$ is convex. In Section 4, several types of convex pairs are identified when $n = 2$. In addition, for general n, the matrices B for which $R(A,B)$ is convex for all A are characterized. Besides $A = 0$, there are no matrices A for which $R(A,B)$ is convex for all B.

Generally, $R(A,B)$ is not convex or even star-shaped as shown by examples. For 2-by-2 matrices, we will show that $R(A,B)$ satisfies a "near convexity" property in Section 5. We conjecture that a generalization of this property is satisfied in the n-by-n case as well.

In Section 6 we generalize the ratio field of values to include fields of values with more than one matrix in the numerator and in the denominator. We prove that in general these (k,m)-fields of values will not be simply connected.

Finally, in Appendix, we give a series of pictures of ratio fields, both as examples of our results and to exhibit the rich variety of shapes that can occur in low dimensions. These pictures were generated using the mentioned parametric description of $R(A,B)$.

2. Elementary properties and canonical form

In this section, we introduce a number of basic properties of the ratio field of values. Many of these properties are direct analogues of properties of the usual field of values. Some of these results are also known for numerical ranges of matrix polynomials [7].

Lemma 2.1. *Let $A, B \in M_n(\mathbb{C})$ and $0 \notin F(B)$. Then the ratio field of values $R(A,B)$ satisfies the following properties:*

(1) (Compactness). *$R(A,B)$ is a compact subset of $\mathbb{C}$.*
(2) (Connectedness). *$R(A,B)$ is connected.*

(3) (Ratio homogeneity). $R(\alpha A, \beta B) = \frac{\alpha}{\beta} R(A, B)$ *for any* $\alpha, \beta \in \mathbb{C}, \beta \neq 0$.
(4) (Translation). $R(A + \beta B, B) = R(A, B) + \beta$ *for* $\beta \in \mathbb{C}$.
(5) (Inversion). $R(A, B) = \dfrac{1}{R(B, A)}$ *if* $0 \notin F(A)$.
(6) (Numerator subadditivity). $R(A_1 + A_2, B) \subseteq R(A_1, B) + R(A_2, B)$.
(7) (Generalized eigenvalue inclusion). *The eigenvalues of* $B^{-1}A$ *are contained in* $R(A, B)$.
(8) (Degeneracy). $R(A, B)$ *is a single point if and only if* $A = \beta B$ *for some* $\beta \in \mathbb{C}$.

Proof. Let $f : \mathbb{C}^n \backslash \{0\} \to \mathbb{C}$ be the map $f(x) = \dfrac{x^*Ax}{x^*Bx}$. Let $S_{\mathbb{C}}^n = \{x \in \mathbb{C}^n : x^*x = 1\}$. By definition $R(A, B) = f(S_{\mathbb{C}}^n)$. Since $0 \notin F(B)$, f is continuous. Both compactness and connectedness of $R(A, B)$ follow immediately. The ratio homogeneity, translation, inversion, and numerator subadditivity properties all follow immediately from the definition of $R(A, B)$.

If λ is an eigenvalue of $B^{-1}A$, then there exists $x \in S_{\mathbb{C}}^n$ such that $B^{-1}Ax = \lambda x$. Therefore

$$Ax = \lambda Bx \Rightarrow x^*Ax = \lambda x^*Bx \Rightarrow \lambda = \frac{x^*Ax}{x^*Bx}$$

so $\lambda \in R(A, B)$.

We now prove the degeneracy property. It is well known (e.g. see [4]) that $F(A) = \{0\}$ if and only if $A = 0$. Consequently, $R(A, B) = \{0\}$ if and only if $A = 0$. It remains to invoke the translation property, according to which $R(A, B) = \{\beta\}$ if and only if $R(A - \beta B, B) = \{0\}$. $\square$

Since $R(A, B) = \{\lambda \in \mathbb{C} : 0 \in F(\lambda B - A)\}$, the principal submatrix inclusion property for the usual field of values [4, Property 1.2.11] implies the following analogue for the ratio field of values.

Lemma 2.2 (Principal submatrix inclusion)**.** *For* $\alpha \subseteq \{1, 2, \ldots, n\}$,

$$R(A[\alpha], B[\alpha]) \subseteq R(A, B).$$

The following lemma relates the usual field of values to the ratio field of values.

Lemma 2.3. *Let* $A, B \in M_n(\mathbb{C})$ *and* $0 \notin F(B)$. *Let* $U(n)$ *denote the set of unitary matrices in* $M_n(\mathbb{C})$. *Then*

(1) $R(A, B) \subseteq F(A)/F(B)$ *and*
(2) $\bigcup_{U \in U(n)} R(U^*AU, B) = F(A)/F(B)$.

Proof. Suppose that $z \in R(A, B)$. Then $z = \dfrac{x^*Ax}{x^*Bx}$ for some $x \in \mathbb{C}^n$. Let $z_1 = x^*Ax$ and let $z_2 = x^*Bx$. Then $z = z_1/z_2$, so $R(A, B) \subseteq F(A)/F(B)$. This implies that $\bigcup_{U \in U(n)} R(U^*AU, B) \subseteq F(A)/F(B)$.

Suppose $z \in F(A)/F(B)$. Then there exist $x, y \in \mathbb{C}^n$ with $x^*x = y^*y = 1$ such that

$$z = \frac{y^*Ay}{x^*Bx}.$$

Since $x^*x = y^*y = 1$, there exists a unitary matrix U_0 such that $y = U_0x$. Therefore

$$z = \frac{y^*Ay}{x^*Bx} = \frac{x^*U_0^*AU_0x}{x^*Bx} \in R\left(U_0^*AU_0, B\right)$$

so that $F(A)/F(B) \subseteq \bigcup_{U \in U(n)} R(U^*AU, B)$. $\square$

We now introduce two lemmas that restrict the class of matrices whose ratio fields we need to study. While the usual field of values of a matrix is invariant under unitary similarity, the ratio field of values is invariant under simultaneous congruence by an invertible matrix. This is a special case of Proposition 2.1(d) in [7].

Lemma 2.4 (Congruential invariance). *If $A, B, C \in M_n(\mathbb{C})$, $0 \notin F(B)$, and C is invertible, then $R(A, B) = R(C^*AC, C^*BC)$.*

However, the ratio field is not invariant under congruence or even unitary similarity in only the numerator or denominator, refer to Figs. 1 and 2 in Appendix for graphical examples.

Any $B \in M_n(\mathbb{C})$ with $0 \notin F(B)$ is congruent to a unimodular diagonal form [5].

Lemma 2.5. *If $B \in M_n(\mathbb{C})$ is non-singular, then B is congruent to a diagonal matrix $\operatorname{diag}(e^{i\theta_1}, \ldots, e^{i\theta_n})$ with each $\theta_i \in [0, 2\pi)$.*

Definition 2.1. We say that an ordered pair of matrices $A, B \in M_n(\mathbb{C})$ is in **congruential canonical form** if $B = \operatorname{diag}(1, e^{\theta_2}, \ldots, e^{i\theta_n})$ with $\theta_2, \ldots, \theta_n \in (-\pi, \pi)$. We refer to the angles θ_i as the **canonical angles** of B. By convention, we let $\theta_1 = 0$.

Given any two matrices $A, B \in M_n(\mathbb{C})$ with $0 \notin F(B)$, Lemma 2.5 together with the congruential invariance and ratio homogeneity properties imply that there are n-by-n matrices A_0, B_0 in congruential canonical form such that $R(A, B) = R(A_0, B_0)$.

In subsequent sections we will give parametric descriptions of the ratio field of values. The following lemma shows that in order to generate the entire ratio field of values, we may assume the first coordinate of each $x \in \mathbb{C}^n$ is real.

Lemma 2.6. *Let $A, B \in M_n(\mathbb{C})$ and $0 \notin F(B)$. Let $f : \mathbb{C}^n\backslash\{0\} \to \mathbb{C}$ be the map*

$$f(x) = \frac{x^*Ax}{x^*Bx}.$$

Let $S' = \{x \in \mathbb{C}^n : x^*x = 1 \text{ and } x_1 \text{ is real}\}$. Then f maps S' onto $R(A, B)$.

Proof. We already know that f maps the unit sphere $S^n_{\mathbb{C}}$ onto $R(A, B)$. Suppose that $z \in R(A, B)$ and $f(x) = z$ for some $x \in S^n_{\mathbb{C}}$. Choose $\alpha \in \mathbb{C}$ such that $|\alpha| = 1$ and αx_1 is real. Then

$$f(\alpha x) = \frac{(\alpha x)^*A(\alpha x)}{(\alpha x)^*B(\alpha x)} = \frac{x^*Ax}{x^*Bx} = z.$$

So every $z \in R(A, B)$ is in the image $f(S')$. $\square$

It is known that a bounded numerical range of a matrix pencil is simply connected (Theorem 4 [9]). Since the numerical range of a matrix pencil $W(A - \lambda B)$ coincides with the ratio field of values $R(A, B)$ precisely when $W(A - \lambda B)$ is bounded, this result applies directly to ratio fields of value. We include our own slightly shorter version of the proof given in [9].

Theorem 2.1. *Let $A, B \in M_n(\mathbb{C})$ and suppose that $0 \notin F(B)$. The ratio field of values $R(A, B)$ is simply connected.*

Proof. Choose any $\lambda_0 \in \mathbb{C}\backslash R(A, B)$. We will show that there is a ray originating from λ_0 that does not intersect $R(A, B)$. We may assume without loss of generality that $\lambda_0 = 0$ by replacing A with $A - \lambda_0 B$. Since $0 \notin R(A, B)$, it follows that $0 \notin F(A)$. The field of values $F(A)$ is a convex set; therefore it is contained in a sector of the plane with vertex at the origin and an angle of less that π. $F(B)$ is also contained in such a sector. The set $F(A)/F(B)$ will then be contained in a sector with vertex at the origin and an angle of less than 2π. In particular, there must be a ray from the origin that is disjoint from $F(A)/F(B)$. Since $R(A, B) \subseteq F(A)/F(B)$, the proof is complete. $\square$

3. Master curve and ellipse description of $R(A, B)$

In this section, we will derive a geometric description of the ratio field of values for 2-by-2 matrices. Suppose that $A, B \in M_2(\mathbb{C})$ and $0 \notin F(B)$. By Lemmas 2.4 and 2.5 we may assume without loss of generality that

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{bmatrix}.$$

By Lemma 2.6 every point in $F(A)$ has the form x^*Ax for some

$$x = \begin{bmatrix} r \\ \sqrt{1-r^2}e^{i\omega} \end{bmatrix}$$

in which $r \in [0, 1]$ and $\omega \in [0, 2\pi]$. Expanding this expression we get a parametric formula for the field of values

$$F(A) = \left\{ r^2 a_{11} + (1-r^2)a_{22} + r\sqrt{1-r^2}(a_{12}e^{i\omega} + a_{21}e^{-i\omega}) : 0 \leqslant \omega \leqslant 2\pi, 0 \leqslant r \leqslant 1 \right\}.$$

Let $\lambda = 1 - r^2$ and let

$$E(A) = \left\{ a_{12}e^{i\omega} + a_{21}e^{-i\omega} : 0 \leqslant \omega \leqslant 2\pi \right\}. \tag{3.1}$$

When the matrix A is understood, we will refer to the set $E(A)$ as E. If we let

$$E_\lambda = E_\lambda(A) = \lambda a_{22} + (1-\lambda)a_{11} + \sqrt{\lambda(1-\lambda)}E(A), \tag{3.2}$$

then we derive the following explicit formula for the field of values as a union of sets that are either ellipses or line segments:

$$F(A) = \bigcup_{0 \leqslant \lambda \leqslant 1} E_\lambda.$$

The following lemma gives explicit conditions for $E = E(A)$ to be an ellipse, and we include it for ease of reference later in the paper.

Lemma 3.1. *Let $E = E(A)$ be defined as in (3.1).*

(1) *E is a proper ellipse if $a_{12} \neq a_{21}$ and Re $\left(\frac{a_{12}+a_{21}}{a_{12}-a_{21}}\right) \neq 0$.*
(2) *E is a line segment if $a_{12} = a_{21}$ or Re $\left(\frac{a_{12}+a_{21}}{a_{12}-a_{21}}\right) = 0$ and a_{12}, a_{21} are nonzero.*
(3) *E is a single point if $a_{12} = a_{21} = 0$.*

Furthermore, any proper ellipse centered at the origin in $\mathbb{C}$ can be written in the form $\{be^{i\omega} + ce^{-i\omega} : 0 \leqslant \omega \leqslant 2\pi\}$ for some pair $b, c \in \mathbb{C}$.

Proof. Let $b = a_{21}$ and $c = a_{21}$. Let $T : \mathbb{C} \to \mathbb{C}$ be the real linear map defined by $T(1) = b + c$ and $T(i) = (b-c)i$. Since $be^{i\omega} + ce^{-i\omega} = (b+c)\cos(\omega) + (b-c)i\sin(\omega)$, T maps the unit circle in $\mathbb{C}$ onto the set $\{be^{i\omega} + ce^{-i\omega} : 0 \leqslant \omega \leqslant 2\pi\}$. If T is invertible, then it follows that the set $\{be^{i\omega} + ce^{-i\omega} : 0 \leqslant \omega \leqslant 2\pi\}$ is a proper ellipse. On the other hand, T will be singular if and only if $b + c$ and $(b - c)i$ both lie on the same line passing through the origin. In this case, the set $\{be^{i\omega} + ce^{-i\omega} : 0 \leqslant \omega \leqslant 2\pi\}$ will be a line segment or a single point, not an ellipse. Note that $0, b+c$, and $(b - c)i$ are co-linear if and only if $b = c$ or $\mathrm{Re}\,(\frac{b+c}{b-c}) = 0$. Furthermore $0 = b + c = (b - c)i$ if and only if $b = c = 0$. This proves conditions 1–3.

Now suppose that E is any proper ellipse centered at the origin. Then there is an invertible real linear map S such that $S(E)$ is the unit circle. Let $T = S^{-1}$, and choose $b, c \in \mathbb{C}$ so that $b + c = T(1)$ and $(b - c)i = T(i)$. Then E will equal $\{be^{i\omega} + ce^{-i\omega} : 0 \leqslant \omega \leqslant 2\pi\}$. □

For each $\lambda \in [0, 1]$ let

$$\tau_\lambda = \tau_\lambda(B) = \frac{1}{\lambda e^{i\theta} + (1 - \lambda)}. \tag{3.3}$$

Note that τ_λ is well-defined for every $\lambda \in \mathbb{R}$ since the requirement that $0 \notin F(B)$ implies that $0 \notin \mathrm{conv}\,\{1, e^{i\theta}\}$ so the line passing through 1 and $e^{i\theta}$ does not intersect the origin. Using τ_λ we may write the ratio field of values as

$$R(A, B) = \bigcup_{0 \leqslant \lambda \leqslant 1} \tau_\lambda E_\lambda$$

in which $\tau_\lambda E_\lambda = \{\tau_\lambda z : z \in E_\lambda\}$. In this way, we see that the ratio field of values is also a union of ellipses. Since the ratio field of values is simply connected (Theorem 2.1) we may replace the ellipses $\tau_\lambda E_\lambda$ in the equation above with the solid ellipses $\mathrm{conv}\,\tau_\lambda E_\lambda$. Thus the ratio field of values is a union of solid ellipses.

$$R(A, B) = \bigcup_{0 \leqslant \lambda \leqslant 1} \mathrm{conv}\,\tau_\lambda E_\lambda. \tag{3.4}$$

Each of the sets E_λ is centered at $\lambda a_{22} + (1 - \lambda)a_{11}$. Therefore the center of each $\tau_\lambda E_\lambda$ is $\tau_\lambda(\lambda a_{22} + (1 - \lambda)a_{11})$. It will be useful to define the following function:

$$\sigma(\lambda) = \tau_\lambda(\lambda a_{22} + (1 - \lambda)a_{11}). \tag{3.5}$$

We will refer to the set $\mathcal{M} = \{\sigma(\lambda) : 0 \leqslant \lambda \leqslant 1\}$ as the **master curve** of $R(A, B)$. Note that $\mathcal{M} \subseteq R(A, B)$ by Eq. (3.4). Furthermore, $\sigma(\lambda)$ is a linear fractional transformation in λ. Therefore the master curve is either a line segment, a circular arc, or a single point, since it is the image of the line segment $[0, 1]$ under a linear fractional transformation.

Although each ratio field of values is a union of convex sets, it is not hard to construct examples of ratio fields that are not themselves convex. In fact, there are 2-by-2 ratio fields of values, $R(A, B)$, with non-empty interior that have a **cut point**, that is, there is a point $z_0 \in R(A, B)$ such that $R(A, B)\backslash\{z_0\}$ is disconnected. The following lemma is useful to explain how this surprising behavior can occur.

Lemma 3.2. *Suppose that each E_λ is a line segment rather than an ellipse. Let l_λ denote the line containing E_λ for each $0 \leqslant \lambda \leqslant 1$. If $\theta \neq 0$, then the lines $\tau_\lambda l_\lambda$ all intersect at a single point, and furthermore, the union $\bigcup_{0 \leqslant \lambda \leqslant 1} \tau_\lambda l_\lambda$ has a cut point at that point of intersection.*

Proof. Choose $\alpha \in \mathbb{C}$ such that $|\alpha| = 1$ and such that the line $\{z \in \mathbb{C} : \mathrm{Re}\,(\alpha z) = 0\}$ is parallel to E. Let $b_\lambda = \mathrm{Re}\,(\alpha(\lambda a_{22} + (1 - \lambda)a_{11}))$. By construction $l_\lambda = \{z \in \mathbb{C} : \mathrm{Re}\,(\alpha z) = b_\lambda\}$ for all $0 \leqslant \lambda \leqslant 1$. For any $z \in \mathbb{C}$, the function

$$f_z(\lambda) = \mathrm{Re}\,\left(\alpha \tau_\lambda^{-1} z\right) - b_\lambda$$

is a real linear function of λ since $\tau_\lambda^{-1} = \lambda e^{i\theta} + (1 - \lambda)$. Furthermore, $z \in \tau_\lambda l_\lambda$ if and only if $f_z(\lambda) = 0$. Since $\theta \neq 0$, it follows that $\tau_0\lambda_0$ and $\tau_1\lambda_1$ cannot be parallel and instead must have a point of intersection z_0. Thus $f_{z_0}(0) = f_{z_0}(1) = 0$, and since f_{z_0} is linear, $f_{z_0}(\lambda) = 0$ for all $0 \leqslant \lambda \leqslant 1$. Therefore $z_0 \in \tau_\lambda l_\lambda$ for all $0 \leqslant \lambda \leqslant 1$. It is immediately clear that z_0 is a cut point for the set $\bigcup_{0\leqslant\lambda\leqslant 1} \tau_\lambda l_\lambda$. □

The above lemma gives us a sufficient condition for the ratio field of values of a pair of 2-by-2 matrices to have a cut point. We describe this condition along with a necessary condition in the theorem below. See Figs. 3, 4, and 5 for illustrations of the cut point for 2-by-2 ratio fields of value.

Theorem 3.1. *Let $A, B \in M_2(\mathbb{C})$ be in congruential canonical form with $0 \notin F(B)$ and suppose $E = E(A)$ is defined as in (3.1). If E is a proper ellipse, then $R(A, B)$ does not contain a cut point. If E is a nontrivial line segment, then $R(A, B)$ contains a cut point if and only if there exists distinct $\lambda_1, \lambda_2 \in (0, 1)$ such that the relative interiors of $\tau_{\lambda_1}E_{\lambda_1}$ and $\tau_{\lambda_2}E_{\lambda_2}$ have non-empty intersection.*

Proof. If E is a proper ellipse, then so is each set $\tau_\lambda E_\lambda$ with $\lambda \in (0, 1)$. Since $\tau_\lambda E_\lambda$ is a proper ellipse for $\lambda \in (0, 1)$, each of the corresponding solid ellipses conv $\tau_\lambda E_\lambda$ has non-empty interior.

Choose any two points $z_1, z_2 \in R(A, B)$. Let $\lambda_1, \lambda_2 \in [0, 1]$ be chosen so that $z_1 \in \text{conv}\, \tau_{\lambda_1}E_{\lambda_1}$ and $z_2 \in \text{conv}\, \tau_{\lambda_2}E_{\lambda_2}$. Let $\sigma(\lambda_1)$ be the center of $\tau_{\lambda_1}E_{\lambda_1}$ and similarly let $\sigma(\lambda_2)$ be the center of $\tau_{\lambda_2}E_{\lambda_2}$. Both $\sigma(\lambda_1)$ and $\sigma(\lambda_2)$ are elements of the master curve.

We can construct a path connecting z_1 to z_2 in $R(A, B)$ by taking the line segment from z_1 to $\sigma(\lambda_1)$ then traveling along the master curve from $\sigma(\lambda_1)$ to $\sigma(\lambda_2)$, and finally following a straight line from $\sigma(\lambda_2)$ to z_2. With the possible exception of z_1 and z_2, every point in this path is an interior point of $R(A, B)$, and therefore cannot be a cut point. Therefore $R(A, B)$ cannot have any cut points.

If E is a line segment, then so are each of the sets $\tau_\lambda E_\lambda$. If there are two distinct values λ_1 and λ_2 such that the line segments $\tau_{\lambda_1}E_{\lambda_1}$ and $\tau_{\lambda_2}E_{\lambda_2}$ intersect at a point in their relative interiors, then that point of intersection will be a cut point for the set $R(A, B)$ as a direct consequence of Lemma 3.2.

It remains to show that if the relative interiors of the line segments $\tau_\lambda E_\lambda$ do not intersect, then $R(A, B)$ does not have a cut point. Let us parametrize the line segment $E = E(t)$ with a parameter $t \in [0, 1]$. This induces a parametrization $E_\lambda(t)$ for each of the line segments E_λ. The map $\varphi(t, \lambda) := \tau_\lambda E_\lambda(t)$ is continuous. Since the relative interiors of the line segments E_λ do not intersect, it follows that φ is a continuous bijection of $(0, 1)\times(0, 1)$ into a subset of $R(A, B)$. Since closed subsets of $(0, 1)\times(0, 1)$ are compact and compact sets are mapped to compact sets, we see that φ is a homeomorphism onto its range, and we conclude that $R(A, B)$ cannot have a cut point. □

Remark 1. Let A, B be any n-by-n matrices with $0 \notin F(B)$. If $R(A, B)$ has empty interior, then $R(A, B)$ is either a circular arc, a line segment, or a single point (Theorem 7 [9]). In the first two cases, every point in $R(A, B)$ except the endpoints will be a cut point.

4. Convex pairs

The general problem of characterizing all pairs A, B with $R(A, B)$ convex is currently unsolved. The master curve and ellipse model for the ratio field when $n = 2$ illustrates the difficulty of developing specific criteria for convexity even in this low dimensional case. Modifying a single entry of A or B can radically distort $R(A, B)$. See the Figs. 6, 7 and 8 in Appendix at the end of this paper for examples showing some of the difficulties that arise.

Despite the difficulties, there are some special cases in which simple conditions on A and B can be given to ensure that $R(A, B)$ is convex or to ensure that $R(A, B)$ is not convex. In this section, we will focus on some of these special cases.

We say that a matrix $P \in M_n(\mathbb{C})$ is **rotationally positive definite** or **RPD** if there exists $\alpha \in [0, 2\pi)$ such that $e^{i\alpha}P$ is positive definite. When B is RPD, $R(A, B)$ is convex for all choices of $A \in M_n$ because $R(A, B)$ reduces to the field of values of a single matrix. Conversely, we will show that if B is not rotationally positive definite, there exist A such that $R(A, B)$ is not convex.

Theorem 4.1. *Let $B \in M_n(\mathbb{C})$ such that $0 \notin F(B)$. Then $R(A,B)$ is convex for all $A \in M_n(\mathbb{C})$ if and only if B is RPD.*

Proof. If B is RPD, then $e^{i\alpha}B$ is congruent to the identity for some $\alpha \in [0, 2\pi)$. Consequently we may let the pair $A_0, B_0 = I_n$ be the congruential canonical form of A, B. By Lemma 2.4, $R(A,B) = R(A_0,B_0) = R(A_0,I_n) = F(A_0)$. Since the field of values of a matrix is always convex [4], $R(A,B)$ is convex.

Consider the case in which B is not RPD and let C and α be given such that $e^{i\alpha}C^*BC$ is in congruential canonical form and $A = e^{-i\alpha}(C^*)^{-1}C$ so that $e^{i\alpha}C^*AC = I_n$. Then $R(A,B) = 1/F(e^{i\alpha}C^*BC)$. Let the canonical angles of $e^{i\alpha}C^*BC$ be enumerated as $\{\theta_j\}_1^n$ in which $\theta_1 = 0$. Note that for some j_0, $\theta_{j_0} \neq 0$, otherwise $C^*(e^{i\alpha}B)C = I_n$ which implies that $e^{i\alpha}B$ is positive definite by Sylvester's law of inertia [3].

Since the usual field of a diagonal matrix is the convex closure of its diagonal entries, $F(e^{i\alpha}C^*BC)$ is a polygon with two or more vertices all lying on the unit circle, so if $z \in e^{i\alpha}C^*BC$, then $|z| \leqslant 1$. Therefore, if $y \in R(A,B)$, $|y| \geqslant 1$. Consequently, the chord joining the distinct points 1 and $e^{-i\theta_{j_0}}$, which are elements of $1/F(e^{i\alpha}C^*BC)$, cannot lie in $R(A,B)$. Therefore, $R(A,B)$ is not convex. □

Unlike the case in which B is RPD, the ratio field is not always convex when A is RPD. By a result due to Wilker concerning the convexity of the inverse of an ellipse, it is possible to determine the convexity of the ratio field of values in the 2-by-2 case when A is RPD.

Theorem 4.2. *Let $A \in M_2(\mathbb{C})$ be RPD. For all $C \in M_2(\mathbb{C})$ with $0 \notin F(C)$, there is an $\alpha \in \mathbb{R}$ and a matrix $B = [b_{ij}]_{1\leqslant i,j\leqslant 2}$ with $b_{11} \in \mathbb{C}$ and $b_{21}, b_{12} \in \mathbb{R} \geqslant 0$ such that $R(A,C) = e^{i\alpha}R(I_2,B)$. Furthermore, $R(A,C) = e^{i\alpha}R(I_2,B)$ is convex if and only if*

$$|2b_{11}(b_{12}-b_{21}) + 4ib_{12}b_{21}| \geqslant (b_{12}+b_{21})^2 \text{ and}$$

$$|2b_{11}(b_{12}-b_{21}) - 4ib_{12}b_{21}| \geqslant (b_{12}+b_{21})^2.$$

Proof. Since A is RPD, A is congruent to a matrix $e^{i\alpha_1}I_2$ for some $\alpha_1 \in \mathbb{R}$. Using Lemma 2.4, we may find Q such that $R(A,C) = R(e^{i\alpha_1}I_2, Q) = e^{i\alpha_1}R(I_2,Q)$. There is a unitary matrix U and an $\alpha_2 \in \mathbb{R}$ such that $e^{i\alpha_2}U^*QU = B$ in which B has the form $B = [b_{ij}]_{1\leqslant i,j\leqslant 2}$ with $b_{11} \in \mathbb{C}$ and $b_{21}, b_{12} \in \mathbb{R} \geqslant 0$ (Lemma 1.3.1 in [4]). By congruential invariance $R(I_2,B) = e^{-i\alpha_2}R(I_2,Q) = e^{i(\alpha_1-\alpha_2)}R(A,C)$. Let $\alpha = \alpha_1 - \alpha_2$. Then,

$$R(A,C) = e^{i\alpha}R(I_2,B) = e^{i\alpha}\cdot 1/F(B).$$

Since $R(A,C)$ and $1/F(B)$ are rotations of each other, one is convex if and only if the other is. Using the description of the field of values given at the beginning of Section 3, we can see that $F(B)$ is a convex ellipse centered at b_{11} with major axis of length $b_{12}+b_{21}$ parallel to the real axis and minor axis $|b_{12}-b_{21}|$, so $R(A,B)$ is the inverse of a convex ellipse.

The inverse of the ellipse $F(B)$ is convex if and only if 0 lies on or outside the circles of curvature belonging to the endpoints of the minor axis of $F(B)$, or if the eccentricity of $F(B)$ does not exceed $\sqrt{2}$, 0 lies on or inside the circles of curvature belonging to the endpoints of the major axis of $F(B)$ [12]. We may eliminate the second case, because $0 \notin F(B)$.

We parametrize the boundary of $F(B)$ as

$$\partial F(B) = b_{11} + \frac{1}{2}(b_{12}+b_{21})\cos(\omega) + \frac{1}{2}i|b_{12}-b_{21}|\sin(\omega)$$

with $\omega \in [0, 2\pi]$. The curvature of $\partial F(B)$ as a function of ω is

$$\kappa(\omega) = \frac{2(b_{12}+b_{21})|b_{12}-b_{21}|}{((|b_{12}-b_{21}|\cos(\omega))^2 + ((b_{12}+b_{21})\sin(\omega))^2)^{\frac{3}{2}}}.$$

The curvature at the endpoints of the minor axes is given by

$$\kappa\left(\frac{\pi}{2}\right) = \kappa\left(\frac{3\pi}{2}\right) = \frac{2(b_{12}+b_{21})|b_{12}-b_{21}|}{(b_{12}+b_{21})^3} = \frac{2|b_{12}-b_{21}|}{(b_{12}+b_{21})^2}.$$

Recall that the radius of each circle of curvature is $1/\kappa(\omega)$. From this, we calculate the following expressions for the circles of curvature at the endpoints of the minor axis

$$\left\{ b_{11} \pm \frac{i}{2}\left(|b_{12}-b_{21}| - \frac{(b_{12}+b_{21})^2}{|b_{12}-b_{21}|} \right) + \frac{(b_{12}+b_{21})^2}{2|b_{12}-b_{21}|} e^{i\psi} \,:\, \psi \in [0, 2\pi] \right\}.$$

Using these expressions, we condense Wilker's condition to the following necessary and sufficient conditions for $R(I_2, B)$ to be convex.

$$\left| b_{11} + \frac{i}{2}\left(|b_{12}-b_{21}| - \frac{(b_{12}+b_{21})^2}{|b_{12}-b_{21}|} \right) \right| \geqslant \frac{(b_{12}+b_{21})^2}{2|b_{12}-b_{21}|}$$

and

$$\left| b_{11} - \frac{i}{2}\left(|b_{12}-b_{21}| - \frac{(b_{12}+b_{21})^2}{|b_{12}-b_{21}|} \right) \right| \geqslant \frac{(b_{12}+b_{21})^2}{2|b_{12}-b_{21}|}.$$

If we multiply both sides of the equations above by $2|b_{12}-b_{21}|$, we obtain the desired conditions for $R(A, C)$ to be convex. □

Another special case in which we can say for certain that $R(A, B)$ is not convex arises when $n = 2$ with A normal.

Theorem 4.3. *If $A \in M_2(\mathbb{C})$ is a nonzero normal matrix with $0 \in F(A)$ and $B = diag(1, e^{i\theta})$ with $\theta \in (-\pi, \pi)$, then $R(A, B)$ is not convex.*

Proof. Since A is normal, $F(A)$ is the convex hull of the eigenvalues of A (see e.g. [4]). Thus $F(A)$ is a line segment. From Section 2 we know that $F(A) = \bigcup_\lambda E_\lambda$ (see Eq. (3.2)). Let l denote the line containing $F(A)$. Note that every $E_\lambda \subseteq l$. Therefore, every $\tau_\lambda E_\lambda \subseteq \tau_\lambda l$. Since $0 \in l$ it follows that $0 \in \tau_\lambda l$ for all $0 \leqslant \lambda \leqslant 1$. Suppose that $A = (a_{ij})$. By construction $E_0 = \{a_{11}\}$ and $E_1 = \{a_{22}\}$. If $a_{11} = a_{22} = 0$, then 0 is the center of each of the line segments $\tau_\lambda E_\lambda$ for $0 \leqslant \lambda \leqslant 1$. In that case, 0 is a cut point for $R(A, B)$ and since $R(A, B)$ is not a subset of a line, this proves that $R(A, B)$ is not convex.

If $a_{11} \neq 0$, then the line $\tau_0 l$ contains a_{11} and it also contains 0, but it does not contain any other point in $R(A, B)$, since the other lines $\tau_\lambda l$ only intersect the line $\tau_0 l$ at the origin. Therefore $R(A, B)$ is not convex. If $a_{22} \neq 0$, then the line $\tau_1 l$ contains $e^{-i\theta} a_{22}$ and it also contains 0, but it does not contain any other point in $R(A, B)$. This completes the proof that $R(A, B)$ is not convex. □

The question of when two matrices have a convex ratio field of values seems very delicate, even in the simple 2-by-2 case in which A is normal. See Figs. 9, 10, 11 and 12 in Appendix for an illustration of the difficulties. This problem of classifying pairs of matrices with a convex ratio field of values certainly merits further investigation.

5. *n*-Convexity of *R*(*A*, *B*)

We begin this section with a definition that generalizes the notion of convexity.

Definition 5.1. For any integer $n > 0$, we say that a subset S of a real vector space is *n-convex* if the intersection of S with any straight line has at most n connected components.

Lemma 5.1. *Let $S_1 \supseteq S_2 \supseteq \cdots S_k \supseteq \cdots$ be a nested family of compact n-convex sets. Then $\bigcap_{k\in\mathbb{N}} S_k$ is an n-convex set.*

Proof. Choose a line l that intersects $\bigcap_{k\in\mathbb{N}} S_k$. Suppose the intersection of l with $\cap(\bigcap_{k\in\mathbb{N}} S_k)$ has more than n connected components. Let $x_1, \ldots, x_{n+1}$ be points on l arranged in linear order such that each x_i is in a different connected component of $l \cap (\bigcap_{k\in\mathbb{N}} S_k)$. Between any pair (x_i, x_{i+1}), there must be a point y_i that is not in $l \cap (\bigcap_{k\in\mathbb{N}} S_k)$. For each y_i, there is some k_i such that $y_i \notin S_{k_i}$. By letting k be the maximum k_i, we see that $y_i \notin S_k$ for all $i \in \{1, \ldots, n\}$. But each x_i must be in S_k so we see that S_k must have at least $n + 1$ connected components which is a contradiction. □

Theorem 5.1. *If $A, B \in M_2(\mathbb{C})$ and $0 \notin F(B)$, then $R(A, B)$ is 2-convex.*

Proof. We may assume without loss of generality that A and B are in congruential canonical form, so that $B = \mathrm{diag}(1, e^{i\theta})$ for some $\theta \in [0, 2\pi)$ and $A = (a_{ij})_{1\leqslant i,j\leqslant 2}$. In what follows we will use the notation for $E = E(A)$, $E_\lambda = E_\lambda(A)$, and $\tau_\lambda = \tau_\lambda(B)$ established in Eqs. (3.1), (3.2), and (3.3).

Case 1. Suppose that E is a proper ellipse. Choose any line l in $\mathbb{C}$ that intersects $R(A, B)$. Assume that $l = \{\alpha t + \beta : t \in \mathbb{R}\}$ in which $\alpha, \beta \in \mathbb{C}$. Since $R(A, B) = \bigcup_{0\leqslant\lambda\leqslant 1} \tau_\lambda E_\lambda$, the intersection of l with $R(A, B)$ is non-empty if and only if $l \cap \tau_\lambda E_\lambda \neq \varnothing$ for some λ. Let

$$l_\lambda = \tau_\lambda^{-1} l = (\lambda e^{i\theta} + (1 - \lambda))l.$$

Then l intersects $\tau_\lambda E_\lambda$ if and only if l_λ intersects E_λ. Equivalently, $l_\lambda - \lambda a_{22} - (1 - \lambda)a_{11}$ intersects $\sqrt{\lambda(1 - \lambda)}E$.

Let S be a real linear transformation such that $S(E)$ is the unit circle. Then the condition above is equivalent to saying that the distance between the line $S(l_\lambda - \lambda a_{22} - (1 - \lambda)a_{11})$ and the origin is less than $\sqrt{\lambda(1 - \lambda)}$ for some $0 \leqslant \lambda \leqslant 1$. Note that

$$S(l_\lambda - \lambda a_{22} - (1 - \lambda)a_{11}) = \{p(\lambda)t + q(\lambda) : t \in \mathbb{R}\}$$

in which

$$p(\lambda) = \lambda S(e^{i\theta}\alpha) + (1 - \lambda)S(\alpha) \quad \text{and}$$

$$q(\lambda) = \lambda(S(e^{i\theta}\beta) - S(a_{22})) + (1 - \lambda)(S(\beta) - S(a_{11})).$$

Since $\mathbb{C}$ is a real inner product space with inner product $\langle z_1, z_2\rangle = \mathrm{Re}\,(z_1\overline{z_2})$, we conclude that distance between the line $\{p(\lambda)t + q(\lambda) : t \in \mathbb{R}\}$ and the origin is less than $\sqrt{\lambda(1 - \lambda)}$ for some $\lambda \in (0, 1)$ if and only if the following fourth degree polynomial inequality is satisfied by that λ.

$$|\mathrm{Re}\,(ip(\lambda)\overline{q(\lambda)})|^2 + (\lambda^2 - \lambda)|p(\lambda)|^2 \leqslant 0. \tag{5.1}$$

In other words, the line l intersects $\tau_\lambda E_\lambda$ if and only (5.1) is satisfied by λ.

Suppose that the roots $\{r_1, r_2, r_3, r_4\}$ of the polynomial in Eq. (5.1) are real, and assume that $r_1 \leqslant r_2 \leqslant r_3 \leqslant r_4$. Since the polynomial on the left side of (5.1) is always nonnegative for $\lambda < 0$ and $\lambda > 1$, the solutions of the inequality can have at most two connected components, namely the intervals $[r_1, r_2]$ and $[r_3, r_4]$. If the roots are not all real, then we see by the same reasoning that the set of solutions to (5.1) must have fewer than two connected components.

We have shown that the set of $\lambda \in [0, 1]$ such that l intersects $\tau_\lambda E_\lambda$ can have at most two connected components. We will now show that this implies that $l \cap R(A, B)$ can have at most two connected components. Let I be a closed interval of λ such that $l \cap \tau_\lambda E_\lambda \neq \varnothing$ for all $\lambda \in I$. We will prove that subset of l which intersects $\bigcup_{\lambda \in I} \tau_\lambda E_\lambda$ is connected. Write $l = \{\alpha t + \beta : t \in \mathbb{R}\}$ and define $t_{\max}(\lambda) = \max\{t : \alpha t + \beta \in \tau_\lambda E_\lambda\}$ and $t_{\min}(\lambda) = \min\{t : \alpha t + \beta \in \tau_\lambda E_\lambda\}$. Since the intersection points of the line l with the ellipse $\tau_\lambda E_\lambda$ are the solutions of a quadratic equation, we see that both $t_{\max}$ and $t_{\min}$ are continuous functions of λ. Since I is a closed interval in $[0, 1]$, $t_{\max}$ attains a maximum on I which we will call M and t_{min} attains a minimum on I which we will can m. Note that for any $m \leqslant t \leqslant M$, we may use the continuity of $t_{\max}$ and $t_{\min}$ to find a λ such that $t_{\min}(\lambda) \leqslant t \leqslant t_{\max}(\lambda)$. It follows that $\alpha t + \beta \in \operatorname{conv} \tau_\lambda E_\lambda$ and therefore $\alpha t + \beta \in R(A, B)$. Thus the intersection of l with $\bigcup_{\lambda \in I} \operatorname{conv}(\tau_\lambda E_\lambda)$ is convex and therefore connected.

The set of λ for which l intersects $\tau_\lambda E_\lambda$ has at most two connected components, which we will call I_1 and I_2. Note that

$$l \cap R(A, B) = l \cap \left(\bigcup_{0 \leqslant \lambda \leqslant 1} \tau_\lambda E_\lambda \right) = \left(l \cap \left(\bigcup_{\lambda \in I_1} \tau_\lambda E_\lambda \right) \right) \cup \left(l \cap \left(\bigcup_{\lambda \in I_2} \tau_\lambda E_\lambda \right) \right)$$

which implies that $l \cap R(A, B)$ is the union of two connected sets.

Case 2. If E is a line segment, then it is possible to construct a family of ellipses $\{\tilde{E}_k\}_{k \in \mathbb{N}}$, each centered at the origin, such that $\operatorname{conv} \tilde{E}_{k+1} \subseteq \operatorname{conv} \tilde{E}_k$ and $E = \bigcap_{k \in \mathbb{N}} \operatorname{conv} \tilde{E}_k$. Each ellipse $\tilde{E}_k$ can be written as $\{b_k e^{i\omega} + c_k e^{-i\omega} : 0 \leqslant \omega \leqslant 2\pi\}$ for some pair $b_k, c_k \in \mathbb{C}$ by Lemma 3.1. Let

$$A_k = \begin{bmatrix} a_{11} & b_k \\ c_k & a_{22} \end{bmatrix}.$$

By construction, $R(A_{k+1}, B) \subseteq R(A_k, B)$ and $R(A, B) = \bigcap_{k \in \mathbb{N}} R(A_k, B)$. Since each $R(A_k, B)$ does satisfy the conditions of case 1 of this proof, $R(A_k, B)$ is 2-convex for each $k \in \mathbb{N}$. Therefore, $R(A, B)$ is the intersection of a nested family of 2-convex sets and so by Lemma 5.1, $R(A, B)$ is 2-convex. □

Based on computer generated images of ratio fields of values (see e.g. Figs. 13 and 14 in Appendix), we propose the following conjecture.

Conjecture 5.1. *Let $A, B \in M_n(\mathbb{C})$ and $0 \notin F(B)$. Then $R(A, B)$ is n-convex.*

Remark 2. It is possible to construct pairs of n-by-n matrices A and B such that $R(A, B)$ is n-convex, but not $(n-1)$-convex. To see this, choose a circle $C \subseteq \mathbb{C}$ such that $0 \in C$. Choose a convex n-sided polygon $P \subseteq \mathbb{C}$ such that every side of P intersects C but every vertex of P is outside of C. Let us also assume that $0 \notin P$. Let $B \in M_n(\mathbb{C})$ be the diagonal matrix with diagonal entries equal to the n vertices of P. Let I_n be the n-by-n identity matrix. Then $R(I_n, B) = 1/P$. Note that the set $1/C$ is a line that passes through $R(I_n, B)$ exactly n times.

Remark 3. For 3-by-3 diagonal matrices, A and B, the ratio field of values will be 3-convex. By Corollary 15 in [9], the boundary of $R(A, B)$ consists of precisely three sets that are either circular arcs or line segments. Each of these boundary sets is contained in a set of the form $\{x + iy \in \mathbb{C} : p(x, y) = 0\}$ where $p(x, y)$ is a second degree polynomial for a circular arc or a first degree polynomial for a line segment. Thus the boundary of $R(A, B)$ is a subset of the set of roots of a polynomial of degree at most six. Of course, this implies that a line can cross the boundary in at most six places. If the intersection of the line with $R(A, B)$ has an isolated point, then that isolated point must be either a place where the line is tangent to one of the circular arcs in the boundary, or it is a place where two of the boundary arcs intersect. In either case, the isolated point will be a root with multiplicity at least two in the boundary polynomial. This implies the intersection of a line with $R(A, B)$ can have at most three connected components since each component will contain at least two roots of the boundary polynomial counting multiplicity.

The current best known upper bound for Conjecture 5.1 is as follows.

Theorem 5.2. *Let $A, B \in M_n(\mathbb{C})$ and $0 \notin F(B)$. Then $R(A, B)$ is m-convex, where $m = 2n(n-1)$.*

Proof. By Theorem 3 in [2] the boundary of $R(A, B)$ is a subset of a set of the form $\{x + iy : x, y \in \mathbb{R} \text{ and } p(x, y) = 0\}$ where p is a polynomial of degree at most $2n(n-1)$. Now consider a line passing through $R(A, B)$. If the line has equation $x_0 + iy_0 + t(x_1 + iy_1)$, then the intersection of the line with $\partial R(A, B)$ is contained in the set of roots of a polynomial in t of degree at most $2n(n-1)$. Therefore, the intersection of the line with $R(A, B)$ has at most $2n(n-1)$ connected components. □

6. (k, m)-Field of values

In addition to the ratio field of values, there are other possible generalizations of the field of values involving more than one matrix. For any collection of matrices $A_1, A_2, \ldots, A_k \in M_n(\mathbb{C})$ and a second collection of matrices $B_1, B_2, \ldots, B_m \in M_n(\mathbb{C})$ with the property that $0 \notin F(B_i)$ for each $i \in \{1, \ldots, m\}$ we can define the (k, m)**-field of values** as the set

$$F_{(k,m)}(A_1, \ldots, A_k; B_1, \ldots, B_m) = \left\{ \frac{(x^*A_1x)\,(x^*A_2x)\ \cdots\ (x^*A_kx)}{(x^*B_1x)\,(x^*B_2x)\ \cdots\ (x^*B_mx)} : x \in \mathbb{C}^n,\ ||x|| = 1 \right\}.$$

As with ratio fields of value, we can ask whether (k, m)-fields are always simply connected. Unlike, the standard field of values and the ratio field of values, there are examples of (k, m)-fields of values that are not simply connected.

Example 6.1. Let $A_1 = A_2 = \begin{bmatrix} i & 0 \\ 0 & 1 \end{bmatrix}$ and $B_1 = B_2 = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$. Let $x = (x_1, x_2) \in \mathbb{C}^2$ with $||x|| = 1$. Note that

$$\frac{(x^*A_1x)(x^*A_2x)}{(x^*B_1x)(x^*B_2x)} = \left(\frac{|x_1|^2 i + |x_2|^2}{|x_1|^2 + |x_2|^2 i} \right)^2 = \left(\frac{\lambda i + (1-\lambda)}{\lambda + (1-\lambda)i} \right)^2,$$

where $\lambda = |x_1|^2$. The expression

$$f(\lambda) = \frac{\lambda i + (1-\lambda)}{\lambda + (1-\lambda)i}$$

is a linear fractional transformation with $f(0) = -i$, $f\left(\frac{1}{2}\right) = 1$ and $f(1) = i$. Since linear fractional transformations map lines to lines or circles, it follows that the image of the line segment $[0, 1]$ under f is the half circle $\{e^{i\theta} : -\pi/2 \leqslant \theta \leqslant \pi/2\}$. The $(2, 2)$-field of values $F_{(2,2)}(A_1, A_2; B_1, B_2)$ is the set $(f([0, 1]))^2 = \{e^{2i\theta} : -\pi/2 \leqslant \theta \leqslant \pi/2\}$ which is the unit circle and is not simply connected.

Example 6.2. Let $A_1 = A_2 = \begin{bmatrix} 1+i & 0 \\ 0 & 1 \end{bmatrix}$ and let $B = \begin{bmatrix} 2i & 0 \\ 0 & 1 \end{bmatrix}$. Let $x = (x_1, x_2) \in \mathbb{C}^2$ with $||x|| = 1$ and let $\lambda = |x_1|^2$. Then

$$\frac{(x^*A_1x)(x^*A_2x)}{x^*Bx} = \frac{(|x_1|^2(1+i) + |x_2|^2)^2}{2i|x_1|^2 + |x_2|^2} = \frac{(\lambda(1+i) + 1 - \lambda)^2}{2i\lambda + 1 - \lambda}$$
$$= \frac{2i\lambda - \lambda^2 + 1}{2i\lambda - \lambda + 1} = 1 + \frac{\lambda(1-\lambda)}{(2i-1)\lambda + 1}.$$

Let

$$f(\lambda) = \frac{\lambda(1-\lambda)}{(2i-1)\lambda+1},$$

then $F_{(2,1)}(A_1, A_2; B) = \{f(\lambda) + 1 : \lambda \in [0, 1]\}$. Note that $f(0) = f(1) = 0$, and $f(\lambda)$ is one-to-one on $(0, 1)$ since the argument of $((2i-1)\lambda + 1)^{-1}$ is a one-to-one function. We conclude that $F_{(2,1)}(A_1, A_2; B)$ is a simple closed curve in $\mathbb{C}$ and therefore not simply connected.

Example 6.3. Let $A_1 = A_2 = A_3 = \begin{bmatrix} e^{2\pi i/3} & 0 \\ 0 & 1 \end{bmatrix}$. If $x = (x_1, x_2) \in \mathbb{C}^2$ satisfies $||x|| = 1$, then

$$(x^*A_1x)(x^*A_2x)(x^*A_3x) = (x^*A_1x)^3 = |x_1|^2e^{2\pi i/3} + |x_2|^2.$$

Let $\lambda = |x_1|^2$ and

$$f(\lambda) = (\lambda e^{2\pi i/3} + (1-\lambda)).$$

Then $F_{(3,0)}(A_1, A_2, A_3) = (f([0, 1]))^3$. Since the function f is linear, the set $f([0, 1])$ is a line segment that does not pass through the origin and the argument of $f(\lambda)$ is a one-to-one function on $[0, 1]$. Furthermore, each point in $f((0, 1))$ has an argument in $(0, 2\pi/3)$. Therefore, the function $(f(\lambda))^3$ is one-to-one on the interval $(0, 1)$. Since $(f(0))^3 = (f(1))^3$, the curve $F_{(3,0)}(A_1, A_2, A_3) = (f([0, 1]))^3$ is a simple closed curve and is not simply connected.

Remark 4. Examples 6.2 and 6.3 prove that in general we cannot expect the (2, 1)- and (3, 0)-field of values to be simply convex. In fact, since the map $z \mapsto 1/z$ is a homeomorphism from $\mathbb{C}\backslash\{0\}$ onto itself, it follows that we may construct (1, 2) and (0, 3) ratio fields of values that are not simply connected by taking the reciprocals of Examples 6.2 and 6.3, respectively. This observation leads directly to the following theorem.

Theorem 6.4. *Let m and k be nonnegative integers such that $m + k \geqslant 3$. Then there is a (k, m)-field of values $F_{(k,m)}(A_1, \ldots, A_k; B_1, \ldots, B_m)$ with $A_1, \ldots A_k$ and $B_1, \ldots, B_m$ in $M_2(\mathbb{C})$ that is not simply connected.*

Proof. If $k + m = 3$, then Example 6.2, Example 6.3, and their reciprocals cover all possible cases. If $k+m > 3$, then we may choose nonnegative integers $k' \leqslant k$ and $m' \leqslant m$ such that $k'+m' = 3$. Then we may construct a (k', m')-field of values $F_{(k',m')}(A_1, \ldots, A_{k'}; B_1, \ldots, B_{m'})$ that is not simply connected. If we let A_i and B_j be the 2-by-2 identity matrix for all $i > k'$ and $j > m'$, then $F_{(k,m)}(A_1, \ldots A_k) = F_{(k',m')}(A_1, \ldots, A_{k'}; B_1, \ldots, B_{m'})$. Thus we have constructed a (k, m)-field of values that is not simply connected. □

Appendix A. Sample illustrative pictures of $R(A, B)$

In Figs. 1–14,[1] blue background shows the entire ratio field, green lines indicate a small sample of the ellipses E_λ, yellow shows the master curve and red lines are segments connecting the eigenvalues of $B^{-1}A$. Some segments may have been added by Matlab.

[1] For interpretation of the references to color in this figure, the reader is referred to the web version of this article.

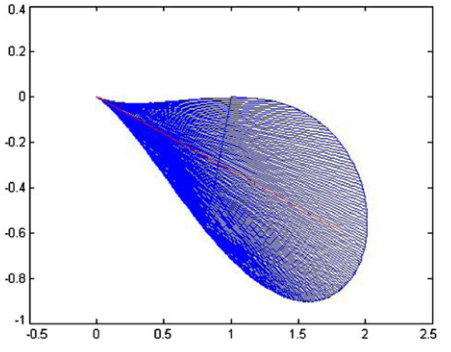

Fig. 1. The ratio field with a singular Hermitian matrix in the numerator:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/5} \end{bmatrix}.$$

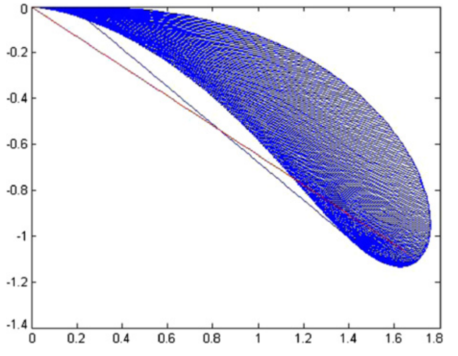

Fig. 2. Unitary rotation of the numerator in figure 1 yields a completely different shape.

$R(U^*AU, B)$ with A, B as in figure 1 and $U = \begin{bmatrix} \cos(\frac{11\pi}{13}) & -\sin(\frac{11\pi}{13}) \\ \sin(\frac{11\pi}{13}) & \cos(\frac{11\pi}{13}) \end{bmatrix}$.

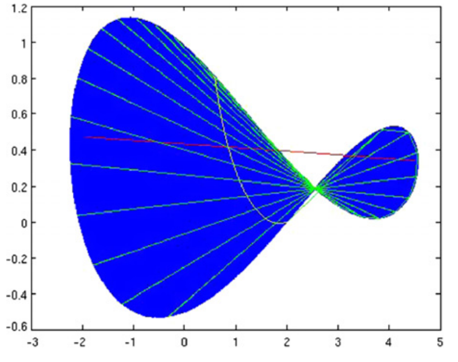

Fig. 3. A ratio field containing a cut point exhibiting master curve and line segments in place of ellipses: $R(A, B)$ for

$$A = \begin{bmatrix} 2 & 3+i \\ 3+i & i \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/5} \end{bmatrix}.$$

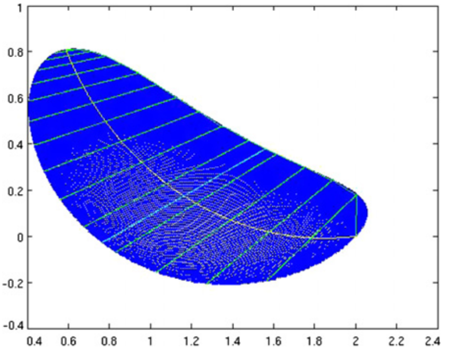

Fig. 4. A ratio field similar to figure 3 but with shorter line segments, so no cut point is attained, $R(A, B)$ for:

$$A = \begin{bmatrix} 2 & .3 + .3i \\ .3 + .3i & i \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/5} \end{bmatrix}.$$

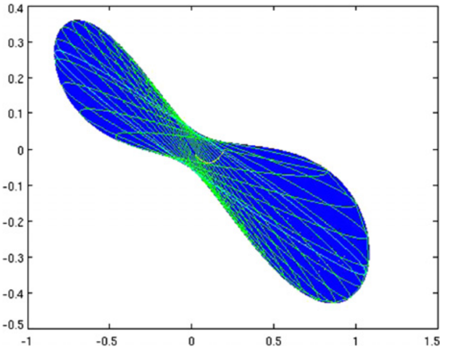

Fig. 5. A ratio field similar to figure 3 but with ellipses instead of line segments, so there is no cut point, $R(A, B)$ for:

$$A = \begin{bmatrix} .2 & 1 \\ .9 & .03 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/5} \end{bmatrix}.$$

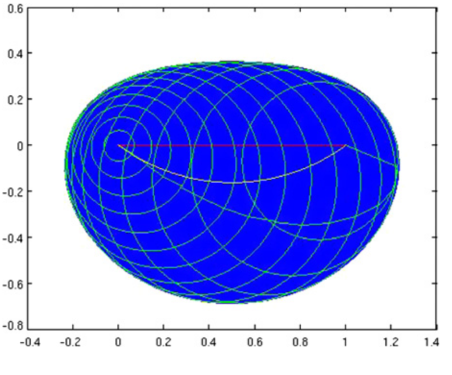

Fig. 6. For this pair A, B, the ratio field of values is convex.

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/5} \end{bmatrix}.$$

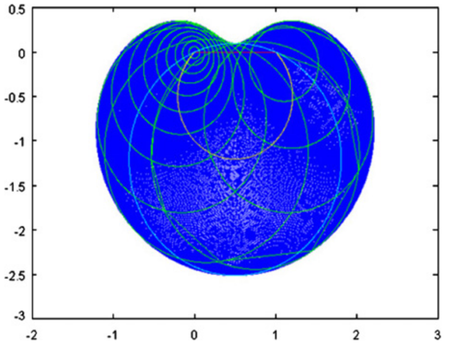

Fig. 7. Changing the second canonical angle from figure 6 makes $R(A, B)$ non-convex:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 \\ 0 & e^{3i\pi/4} \end{bmatrix}.$$

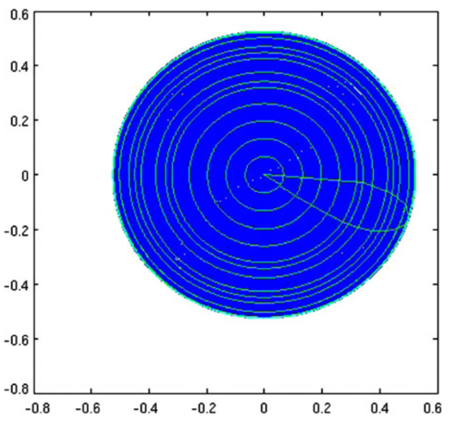

Fig. 8. A ratio field which is a convex disc with trivial master curve, $R(A, B)$ for:

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/5} \end{bmatrix}.$$

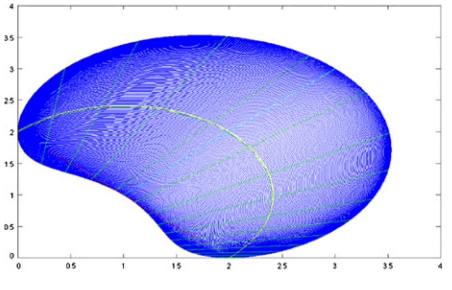

Fig. 9. In the next four figures, A is a normal matrix withparameter k. Here $k = 1$.

$$A = \begin{bmatrix} 2i & k \\ -k & 2i \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}.$$

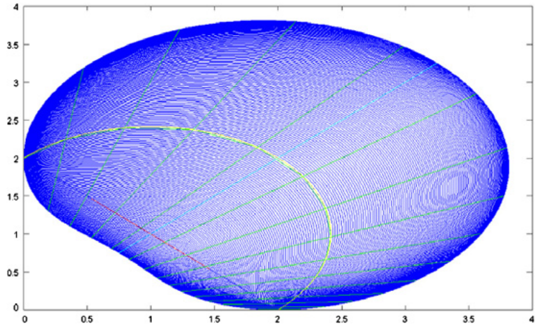

Fig. 10. Here $k = 1.25$

$$A = \begin{bmatrix} 2i & k \\ -k & 2i \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}.$$

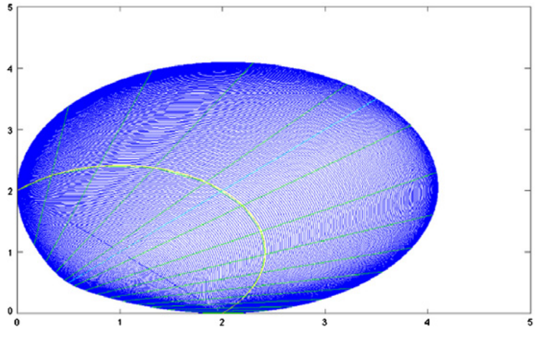

Fig. 11. Here $k = 1.5$.

$$A = \begin{bmatrix} 2i & k \\ -k & 2i \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}.$$

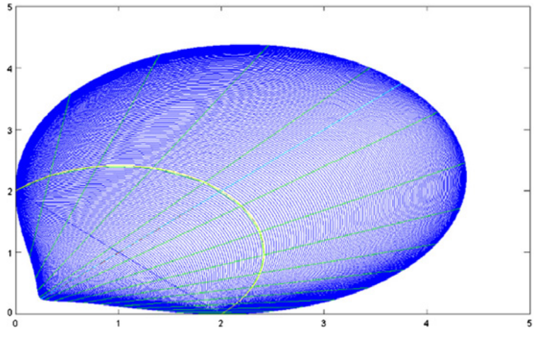

Fig. 12. Here $k = 1.75$.

$$A = \begin{bmatrix} 2i & k \\ -k & 2i \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}.$$

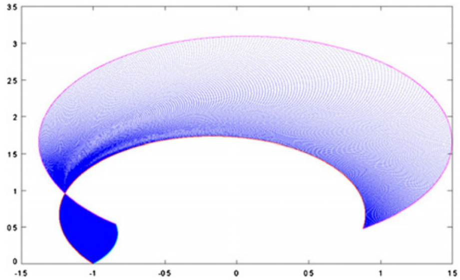

Fig. 13. A 3-by-3 ratio field with a cut point, $R(A, B)$ for:

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & e^{2.5i} & 0 \\ 0 & 0 & e^{\frac{i}{2}} \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{2.5i} & 0 \\ 0 & 0 & e^{-\frac{i}{2}} \end{bmatrix}.$$

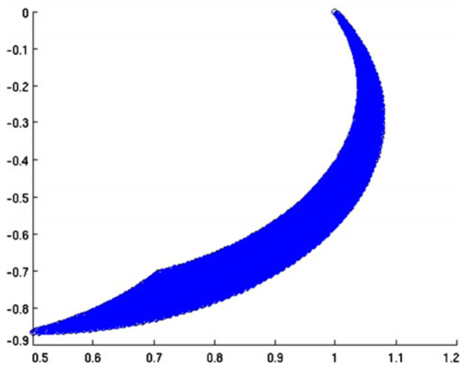

Fig. 14. A 3-by-3 ratio field which is 3-convex:

$$R(A, B) \text{ for } A = I_3 \text{ and } B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{\pi/3} & 0 \\ 0 & 0 & e^{\pi/4} \end{bmatrix}.$$

References

[1] Mao-Ting Chien, Hiroshi Nakazato, The numerical range of linear pencils of 2-by-2 matrices, Linear Algebra Appl. 341 (2002) 69–100. Special issue dedicated to Professor T. Ando.

[2] Mao-Ting Chien, Hiroshi Nakazato, Panayiotis Psarrakos, Point equation of the boundary of the numerical range of a matrix polynomial, Linear Algebra Appl. 347 (2002) 205–217.

[3] Roger A. Horn, Charles R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1985.

[4] Roger A. Horn, Charles R. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge, 1994. Corrected reprint of the 1991 original.

[5] Charles R. Johnson, Susana Furtado, A generalization of Sylvester's law of inertia, Linear Algebra Appl. 338 (2001) 287–290.

[6] Charles R. Johnson, Lev Krukier, General resolution of a convergence question of L. Krukier, Numer. Linear Algebra Appl. 16 (2009) 949–950.

[7] Chi-Kwong Li, Leiba Rodman, Numerical range of matrix polynomials, SIAM J. Matrix Anal. Appl. 15 (4) (1994) 1256–1265.

[8] Chi-Kwong Li, Nam-Kiu Tsing, Frank Uhlig, Numerical range of an operator on an indefinite inner product space, Electron. J. Linear Algebra 1 (1996) 1–17.

[9] Hiroshi Nakazato, Panayiotis Psarrakos, On the shape of numerical range of matrix polynomials, Linear Algebra Appl. 338 (2001) 105–123.

[10] Panayiotis J. Psarrakos, Numerical range of linear pencils, Linear Algebra Appl. 317 (1–3) (2000) 127–141.

[11] Helmut Wielandt, On the eigenvalues of $A + B$ and AB, J. Res. Nat. Bur. Standards Sect. B 77B (1973) 61–63.

[12] J.B. Wilker, When is the inverse of an ellipse convex?, Util. Math. 17 (1980) 45–50.