Topics in Matrix Analysis

Roger A. Horn, Charles R. Johnson

Book DOI: http://dx.doi.org/10.1017/CBO9780511840371

Online ISBN: 9780511840371

Hardback ISBN: 9780521305877

Paperback ISBN: 9780521467131

Chapter

Chapter 1 - The field of values pp. 1-88

Chapter DOI: http://dx.doi.org/10.1017/CBO9780511840371.002

Cambridge University Press

Chapter 1

The field of values

1.0 Introduction

Like the spectrum (or set of eigenvalues) $\sigma(\cdot)$, the field of values $F(\cdot)$ is a set of complex numbers naturally associated with a given n-by-n matrix A:

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

The spectrum of a matrix is a discrete point set; while the field of values can be a continuum, it is always a compact convex set. Like the spectrum, the field of values is a set that can be used to learn something about the matrix, and it can often give information that the spectrum alone cannot give. The eigenvalues of Hermitian and normal matrices have especially pleasant properties, and the field of values captures certain aspects of this nice structure for general matrices.

1.0.1 Subadditivity and eigenvalues of sums

If only the eigenvalues $\sigma(A)$ and $\sigma(B)$ are known about two n-by-n matrices A and B, remarkably little can be said about $\sigma(A+B)$, the eigenvalues of the sum. Of course, $\text{tr}(A+B) = \text{tr}\, A + \text{tr}\, B$, so the sum of all the eigenvalues of $A+B$ is the sum of all the eigenvalues of A plus the sum of all the eigenvalues of B. But beyond this, nothing can be said about the eigenvalues of $A+B$ without more information about A and B. For example, even if all the eigenvalues of two n-by-n matrices A and B are known and fixed, the spectral radius of $A+B$ [the largest absolute value of an eigenvalue of $A+B$, denoted by $\rho(A+B)$] can be arbitrarily large (see Problem 1). On the other hand, if A and B are normal, then much can be said about the eigenvalues of $A + B$; for example, $\rho(A + B) \leq \rho(A) + \rho(B)$ in this case. Sums of matrices do arise in practice, and two relevant properties of the field of values $F(\cdot)$ are:

(a) The field of values is subadditive: $F(A + B) \subset F(A) + F(B)$, where the set sum has the natural definition of sums of all possible pairs, one from each; and
(b) The eigenvalues of a matrix lie inside its field of values: $\sigma(A) \subset F(A)$.

Combining these two properties yields the inclusions

$$\sigma(A + B) \subset F(A + B) \subset F(A) + F(B)$$

so if the two fields of values $F(A)$ and $F(B)$ are known, something can be said about the spectrum of the sum.

1.0.2 An application from the numerical solution of partial differential equations

Suppose that $A = [a_{ij}] \in M_n(\mathbb{R})$ satisfies

(a) A is tridiagonal ($a_{ij} = 0$ for $|i - j| > 1$), and

(b) $a_{i,i+1}a_{i+1,i} < 0$ for $i = 1, \ldots, n-1$.

Matrices of this type arise in the numerical solution of partial differential equations and in the analysis of dynamical systems arising in mathematical biology. In both cases, knowledge about the real parts of the eigenvalues of A is important. It turns out that rather good information about the eigenvalues of such a matrix can be obtained easily using the field of values $F(\cdot)$.

1.0.2.1 Fact: For any eigenvalue λ of a matrix A of the type indicated, we have

$$\min_{1 \leq i \leq n} a_{ii} \leq \operatorname{Re} \lambda \leq \max_{1 \leq i \leq n} a_{ii}$$

A proof of this fact is fairly simple using some properties of the field of values to be developed in Section (1.2). First, choose a diagonal matrix D

with positive diagonal entries such that $D^{-1}AD \equiv \hat{A} = [\hat{a}_{ij}]$ satisfies $\hat{a}_{ji} = -\hat{a}_{ij}$ for $j \neq i$. The matrix $D \equiv \operatorname{diag}(d_1,\ldots, d_n)$ defined by

$$d_1 = 1, \text{ and } d_i = \left|\frac{a_{i,i-1}}{a_{i-1,i}}\right|^{\frac{1}{2}} d_{i-1},\ d_i > 0,\ i = 2,\ldots, n$$

will do. Since $\hat{A}$ and A are similar, their eigenvalues are the same. We then have

$$\begin{aligned}\operatorname{Re}\sigma(A) = \operatorname{Re}\sigma(\hat{A}) \subset \operatorname{Re} F(\hat{A}) &= F(\tfrac{1}{2}(\hat{A} + \hat{A}^T)) \\ &= F(\operatorname{diag}(a_{11},\ldots, a_{nn})) \\ &= \text{Convex hull of } \{a_{11},\ldots, a_{nn}\} = [\min_i a_{ii}, \max_i a_{ii}]\end{aligned}$$

The first inclusion follows from the spectral containment property (1.2.6), the next equality follows from the projection property (1.2.5), the next equality follows from the special form achieved for $\hat{A}$, and the last equality follows from the normality property (1.2.9) and the fact that the eigenvalues of a diagonal matrix are its diagonal entries. Since the real part of each eigenvalue $\lambda \in \sigma(A)$ is a convex combination of the main diagonal entries a_{ii}, $i = 1,\ldots, n$, the asserted inequalities are clear and the proof is complete.

1.0.3 Stability analysis

In an analysis of the stability of an equilibrium in a dynamical system governed by a system of differential equations, it is important to know if the real part of every eigenvalue of a certain matrix A is negative. Such a matrix is called *stable*. In order to avoid juggling negative signs, we often work with *positive stable* matrices (all eigenvalues have positive real parts). Obviously, A is positive stable if and only if $-A$ is stable. An important sufficient condition for a matrix to be positive stable is the following fact.

1.0.3.1 Fact: Let $A \in M_n$. If $A + A^*$ is positive definite, then A is positive stable.

This is another application of properties of the field of values $F(\cdot)$ to be developed in Section (1.2). By the spectral containment property (1.2.6), $\operatorname{Re}\sigma(A) \subset \operatorname{Re} F(A)$, and, by the projection property (1.2.5), $\operatorname{Re} F(A) =$

$F(\frac{1}{2}(A + A^*))$. But, since $A + A^*$ is positive definite, so is $\frac{1}{2}(A + A^*)$, and hence, by the normality property (1.2.9), $F(\frac{1}{2}(A + A^*))$ is contained in the positive real axis. Thus, each eigenvalue of A has a positive real part, and A is positive stable.

Actually, more is true. If $A + A^*$ is positive definite, and if $P \in M_n$ is any positive definite matrix, then PA is positive stable because

$$(P^{\frac{1}{2}})^{-1}[PA]P^{\frac{1}{2}} = P^{\frac{1}{2}}AP^{\frac{1}{2}}, \text{ and}$$

$$P^{\frac{1}{2}}AP^{\frac{1}{2}} + (P^{\frac{1}{2}}AP^{\frac{1}{2}})^* = P^{\frac{1}{2}}(A + A^*)P^{\frac{1}{2}}$$

where $P^{\frac{1}{2}}$ is the unique (Hermitian) positive definite square root of P. Since congruence preserves positive definiteness, the eigenvalues of PA have positive real parts for the same reason as A. Lyapunov's theorem (2.2.1) shows that all positive stable matrices arise in this way.

1.0.4 An approximation problem

Suppose we wish to approximate a given matrix $A \in M_n$ by a complex multiple of a Hermitian matrix of rank at most one, as closely as possible in the Frobenius norm $\|\cdot\|_2$. This is the problem

$$\text{minimize } \|A - cxx^*\|_2^2 \text{ for } x \in \mathbb{C}^n \text{ with } x^*x = 1 \text{ and } c \in \mathbb{C} \qquad (1.0.4.1)$$

Since the inner product $[A,B] \equiv \operatorname{tr} AB^*$ generates the Frobenius norm, we have

$$\begin{aligned}\|A - cxx^*\|_2^2 &= [A - cxx^*, A - cxx^*] \\ &= \|A\|_2^2 - 2 \operatorname{Re} \bar{c}\,[A, xx^*] + |c|^2\end{aligned}$$

which, for a given unit vector x, is minimized by $c = [A, xx^*]$. Substitution of this value into (1.0.4.1) transforms our problem into

$$\text{minimize } (\|A\|_2^2 - |[A, xx^*]|^2) \text{ for } x \in \mathbb{C}^n \text{ with } x^*x = 1$$

or, equivalently,

$$\text{maximize } |[A, xx^*]| \text{ for } x \in \mathbb{C}^n \text{ with } x^*x = 1$$

A vector x_0 that solves the latter problem (we are maximizing a continuous function on a compact set) will yield a rank one solution $[A,x_0x_0^*]x_0x_0^*$ to our original problem. Since $[A,xx^*] = \operatorname{tr} Axx^* = x^*Ax$, we are led naturally to finding a unit vector x such that the point x^*Ax in the field of values $F(A)$ has maximum distance from the origin. The absolute value of such a point is called the *numerical radius* of A [often denoted by $r(A)$] by analogy with the *spectral radius* [often denoted by $\rho(A)$], which is the absolute value of a point in the spectrum $\sigma(A)$ that is at maximum distance from the origin.

Problems

1. Consider the real matrices

$$A = \begin{bmatrix} 1-\alpha & 1 \\ \alpha(1-\alpha)-1 & \alpha \end{bmatrix} \text{ and } B = \begin{bmatrix} 1+\alpha & 1 \\ -\alpha(1+\alpha)-1 & -\alpha \end{bmatrix}$$

Show that $\sigma(A)$ and $\sigma(B)$ are independent of the value of $\alpha \in \mathbb{R}$. What are they? What is $\sigma(A+B)$? Show that $\rho(A+B)$ is unbounded as $\alpha \to \infty$.

2. In contrast to Problem 1, show that if $A, B \in M_n$ are normal, then $\rho(A+B) \leq \rho(A) + \rho(B)$.

3. Show that "<" in (1.0.2(b)) may be replaced by "$\leq$," the main diagonal entries a_{ii} may be complex, and Fact (1.0.2.1) still holds if a_{ii} is replaced by $\operatorname{Re} a_{ii}$.

4. Show that the problem of approximating a given $A \in M_n$ by a positive semidefinite rank one matrix with spectral radius one can be solved if one can find a unit vector x such that the point x^*Ax in $F(A)$ is furthest to the right in the complex plane, that is, $\operatorname{Re} x^*Ax$ is maximized.

1.1 Definitions

In this section we define the field of values and certain related objects.

1.1.1 Definition. The *field of values* of $A \in M_n$ is

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

Thus, $F(\cdot)$ is a function from M_n into subsets of the complex plane.

$F(A)$ is just the normalized locus of the Hermitian form associated with A. The field of values is often called the *numerical range*, especially in the context of its analog for operators on infinite dimensional spaces.

Exercise. Show that $F(I) = \{1\}$ and $F(\alpha I) = \{\alpha\}$ for all $\alpha \in \mathbb{C}$. Show that $F\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is the closed unit interval $[0,1]$, and $F\begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$ is the closed unit disc $\{z \in \mathbb{C}: |z| \leq 1\}$.

The field of values $F(A)$ may also be thought of as the image of the surface of the Euclidean unit ball in $\mathbb{C}^n$ (a compact set) under the continuous transformation $x \to x^*Ax$. As such, $F(A)$ is a compact (and hence bounded) set in $\mathbb{C}$. An unbounded analog of $F(\cdot)$ is also of interest.

1.1.2 Definition. The *angular field of values* is

$$F'(A) \equiv \{x^*Ax: x \in \mathbb{C}^n, x \neq 0\}$$

Exercise. Show that $F'(A)$ is determined geometrically by $F(A)$; every open ray from the origin that intersects $F(A)$ in a point other than the origin is in $F'(A)$, and $0 \in F'(A)$ if and only if $0 \in F(A)$. Draw a typical picture of an $F(A)$ and $F'(A)$ assuming that $0 \notin F(A)$.

It will become clear that $F'(A)$ is an angular sector of the complex plane that is anchored at the origin (possibly the entire complex plane). The angular opening of this sector is of interest.

1.1.3 Definition. The *field angle* $\Theta \equiv \Theta(A) \equiv \Theta(F'(A)) \equiv \Theta(F(A))$ of $A \in M_n$ is defined as follows:

(a) If 0 is an interior point of $F(A)$, then $\Theta(A) \equiv 2\pi$.

(b) If 0 is on the boundary of $F(A)$ and there is a (unique) tangent to the boundary of $F(A)$ at 0, then $\Theta(A) \equiv \pi$.

(c) If $F(A)$ is contained in a line through the origin, $\Theta(A) \equiv 0$.

(d) Otherwise, consider the two different support lines of $F(A)$ that go through the origin, and let $\Theta(A)$ be the angle subtended by these two lines at the origin. If $0 \notin F(A)$, these support lines will be uniquely determined; if 0 is on the boundary of $F(A)$, choose the two support lines that give the minimum angle.

We shall see that $F(A)$ is a compact convex set for every $A \in M_n$, so this informal definition of the field angle makes sense. The field angle is just the angular opening of the smallest angular sector that includes $F(A)$, that is, the angular opening of the sector $F'(A)$.

Finally, the size of the bounded set $F(A)$ is of interest. We measure its size in terms of the radius of the smallest circle centered at the origin that contains $F(A)$.

1.1.4 **Definition.** The *numerical radius* of $A \in M_n$ is

$$r(A) \equiv \max\{|z| : z \in F(A)\}$$

The numerical radius is a vector norm on matrices that is not a matrix norm (see Section (5.7) of [HJ]).

Problems

1. Show that among the vectors entering into the definition of $F(A)$, only vectors with real nonnegative first coordinate need be considered.

2. Show that both $F(A)$ and $F'(A)$ are simply connected for any $A \in M_n$.

3. Show that for each $0 \leq \theta \leq \pi$, there is an $A \in M_2$ with $\Theta(A) = \theta$. Is $\Theta(A) = 3\pi/2$ possible?

4. Why is the "max" in (1.1.4) attained?

5. Show that the following alternative definition of $F(A)$ is equivalent to the one given:

$$F(A) \equiv \{x^*Ax/x^*x: \ x \in \mathbb{C}^n \text{ and } x \neq 0\}$$

Thus, $F(\cdot)$ is a normalized version of $F'(\cdot)$.

6. Determine $F\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}$, $F\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$, and $F\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$.

7. If $A \in M_n$ and $\alpha \in F(A)$, show that there is a unitary matrix $U \in M_n$ such that α is the 1,1 entry of U^*AU.

8. Determine as many different possible types of sets as you can that can be an $F'(A)$.

9. Show that $F(A^*) = \overline{F(A)}$ and $F'(A^*) = \overline{F'(A)}$ for all $A \in M_n$.

10. Show that all of the main diagonal entries and eigenvalues of a given $A \in M_n$ are in its field of values $F(A)$.

1.2 Basic properties of the field of values

As a function from M_n into subsets of $\mathbb{C}$, the field of values $F(\cdot)$ has many useful functional properties, most of which are easily established. We catalog many of these properties here for reference and later use. The important property of convexity is left for discussion in the next section.

The sum or product of two subsets of $\mathbb{C}$, or of a subset of $\mathbb{C}$ and a scalar, has the usual algebraic meaning. For example, if $S, T \subset \mathbb{C}$, then $S + T \equiv \{s + t: s \in S, t \in T\}$.

1.2.1 Property: *Compactness*. For all $A \in M_n$,

$$F(A) \text{ is a compact subset of } \mathbb{C}$$

Proof: The set $F(A)$ is the range of the continuous function $x \to x^*Ax$ over the domain $\{x: x \in \mathbb{C}^n, x^*x = 1\}$, the surface of the Euclidean unit ball, which is a compact set. Since the continuous image of a compact set is compact, it follows that $F(A)$ is compact. $\square$

1.2.2 Property: *Convexity*. For all $A \in M_n$,

$$F(A) \text{ is a convex subset of } \mathbb{C}$$

The next section of this chapter is reserved for a proof of this fundamental fact, known as the *Toeplitz-Hausdorff theorem*. At this point, it is clear that $F(A)$ must be a *connected* set since it is the continuous image of a connected set.

Exercise. If A is a diagonal matrix, show that $F(A)$ is the convex hull of the diagonal entries (the eigenvalues) of A.

The field of values of a matrix is changed in a simple way by adding a scalar multiple of the identity to it or by multiplying it by a scalar.

1.2.3 **Property**: *Translation*. For all $A \in M_n$ and $\alpha \in \mathbb{C}$,

$$F(A + \alpha I) = F(A) + \alpha$$

Proof: We have $F(A + \alpha I) = \{x^*(A + \alpha I)x: x^*x = 1\} = \{x^*Ax + \alpha x^*x: x^*x = 1\} = \{x^*Ax + \alpha: x^*x = 1\} = \{x^*Ax: x^*x = 1\} + \alpha = F(A) + \alpha$. $\square$

1.2.4 **Property**: *Scalar multiplication*. For all $A \in M_n$ and $\alpha \in \mathbb{C}$,

$$F(\alpha A) = \alpha F(A)$$

Exercise. Prove property (1.2.4) by the same method used in the proof of (1.2.3).

For $A \in M_n$, $\mathrm{H}(A) \equiv \frac{1}{2}(A + A^*)$ denotes the *Hermitian part of* A and $\mathrm{S}(A) \equiv \frac{1}{2}(A - A^*)$ denotes the *skew-Hermitian part of* A; notice that $A = \mathrm{H}(A) + \mathrm{S}(A)$ and that $\mathrm{H}(A)$ and $i\mathrm{S}(A)$ are both Hermitian. Just as taking the real part of a complex number projects it onto the real axis, taking the Hermitian part of a matrix projects its field of values onto the real axis. This simple fact helps in locating the field of values, since, as we shall see, it is relatively easy to deal with the field of values of a Hermitian matrix. For a set $S \subset \mathbb{C}$, we interpret Re S as $\{\mathrm{Re}\ s: s \in S\}$, the projection of S onto the real axis.

1.2.5 **Property**: *Projection*. For all $A \in M_n$,

$$F(\mathrm{H}(A)) = \mathrm{Re}\ F(A)$$

Proof: We calculate $x^*\mathrm{H}(A)x = x^*\frac{1}{2}(A + A^*)x = \frac{1}{2}(x^*Ax + x^*A^*x) = \frac{1}{2}(x^*Ax + (x^*Ax)^*) = \frac{1}{2}(x^*Ax + \overline{x^*Ax}) = \mathrm{Re}\ x^*Ax$. Thus, each point in $F(\mathrm{H}(A))$ is of the form Re z for some $z \in F(A)$ and vice versa. $\square$

We denote the open upper half-plane of $\mathbb{C}$ by $UHP \equiv \{z \in \mathbb{C}: \mathrm{Im}\ z > 0\}$, the open left half-plane of $\mathbb{C}$ by $LHP \equiv \{z \in \mathbb{C}: \mathrm{Re}\ z < 0\}$, the open right half-plane of $\mathbb{C}$ by $RHP \equiv \{z \in \mathbb{C}: \mathrm{Re}\ z > 0\}$, and the closed right half-plane of $\mathbb{C}$ by $RHP_0 \equiv \{z \in \mathbb{C}: \mathrm{Re}\ z \geq 0\}$. The projection property gives a simple indication of when $F(A) \subset RHP$ or RHP_0 in terms of positive definiteness or positive semidefiniteness.

1.2.5a Property: *Positive definite indicator function.* Let $A \in M_n$. Then $F(A) \subset RHP$ if and only if $A + A^*$ is positive definite

1.2.5b Property: *Positive semidefinite indicator function.* Let $A \in M_n$. Then

$$F(A) \subset RHP_0 \text{ if and only if } A + A^* \text{ is positive semidefinite}$$

Exercise. Prove (1.2.5a) and (1.2.5b) (the proofs are essentially the same) using (1.2.5) and the definition of positive definite and semidefinite (see Chapter 7 of [HJ]).

The point set of eigenvalues of $A \in M_n$ is denoted by $\sigma(A)$, the *spectrum* of A. A very important property of the field of values is that it includes the eigenvalues of A.

1.2.6 Property: *Spectral containment.* For all $A \in M_n$,

$$\sigma(A) \subset F(A)$$

Proof: Suppose that $\lambda \in \sigma(A)$. Then there exists some nonzero $x \in \mathbb{C}^n$, which we may take to be a unit vector, for which $Ax = \lambda x$ and hence $\lambda = \lambda x^* x = x^*(\lambda x) = x^* A x \in F(A)$. □

Exercise. Use the spectral containment property (1.2.6) to show that the eigenvalues of a positive definite matrix are positive real numbers.

Exercise. Use the spectral containment property (1.2.6) to show that the eigenvalues of $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ are imaginary.

The following property underlies the fact that the numerical radius is a vector norm on matrices and is an important reason why the field of values is so useful.

1.2.7 Property: *Subadditivity.* For all $A, B \in M_n$,

$$F(A + B) \subset F(A) + F(B)$$

Proof: $F(A + B) = \{x^*(A + B)x \colon x \in \mathbb{C}^n,\ x^* x = 1\} = \{x^* Ax + x^* Bx \colon x \in \mathbb{C}^n,$

$x^*x = 1\} \subset \{x^*Ax: x \in \mathbb{C}^n, x^*x = 1\} + \{y^*By: y \in \mathbb{C}^n, y^*y = 1\} = F(A) + F(B)$. □

Exercise. Use (1.2.7) to show that the numerical radius $r(\cdot)$ satisfies the triangle inequality on M_n.

Another important property of the field of values is its invariance under unitary similarity.

1.2.8 Property: *Unitary similarity invariance.* For all $A, U \in M_n$ with U unitary,

$$F(U^*AU) = F(A)$$

Proof: Since a unitary transformation leaves invariant the surface of the Euclidean unit ball, the complex numbers that comprise the sets $F(U^*AU)$ and $F(A)$ are the same. If $x \in \mathbb{C}^n$ and $x^*x = 1$, we have $x^*(U^*AU)x = y^*Ay \in F(A)$, where $y = Ux$, so $y^*y = x^*U^*Ux = x^*x = 1$. Thus, $F(U^*AU) \subset F(A)$. The reverse containment is obtained similarly. □

The unitary similarity invariance property allows us to determine the field of values of a normal matrix. Recall that, for a set S contained in a real or complex vector space, $\mathrm{Co}(S)$ denotes the convex hull of S, which is the set of all convex combinations of finitely many points of S. Alternatively, $\mathrm{Co}(S)$ can be characterized as the intersection of all convex sets containing S, so it is the "smallest" closed convex set containing S.

1.2.9 Property: *Normality.* If $A \in M_n$ is normal, then

$$F(A) = \mathrm{Co}(\sigma(A))$$

Proof: If A is normal, then $A = U^*\Lambda U$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal and U is unitary. By the unitary similarity invariance property (1.2.8), $F(A) = F(\Lambda)$ and, since

$$x^*\Lambda x = \sum_{i=1}^{n} \bar{x}_i x_i \lambda_i = \sum_{i=1}^{n} |x_i|^2 \lambda_i$$

$F(\Lambda)$ is just the set of all convex combinations of the diagonal entries of Λ ($x^*x = 1$ implies $\Sigma_i |x_i|^2 = 1$ and $|x_i|^2 \geq 0$). Since the diagonal entries of Λ are the eigenvalues of A, this means that $F(A) = \mathrm{Co}(\sigma(A))$. $\square$

Exercise. Show that if H is Hermitian, $F(H)$ is a closed real line segment whose endpoints are the largest and smallest eigenvalues of A.

Exercise. Show that the field of values of a normal matrix is always a polygon whose vertices are eigenvalues of A. If $A \in M_n$, how many sides may $F(A)$ have? If A is unitary, show that $F(A)$ is a polygon inscribed in the unit circle.

Exercise. Show that $\mathrm{Co}(\sigma(A)) \subset F(A)$ for all $A \in M_n$.

Exercise. If $A, B \in M_n$, show that

$$\sigma(A + B) \subset F(A) + F(B)$$

If A and B are normal, show that

$$\sigma(A + B) \subset \mathrm{Co}(\sigma(A)) + \mathrm{Co}(\sigma(B))$$

The next two properties have to do with fields of values of matrices that are built up from or extracted from other matrices in certain ways. Recall that for $A \in M_{n_1}$ and $B \in M_{n_2}$, the *direct sum* of A and B is the matrix

$$A \oplus B \equiv \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \in M_{n_1+n_2}$$

If $J \subset \{1, 2, \ldots, n\}$ is an index set and if $A \in M_n$, then $A(J)$ denotes the principal submatrix of A contained in the rows and columns indicated by J.

1.2.10 Property: *Direct sums.* For all $A \in M_{n_1}$ and $B \in M_{n_2}$,

$$F(A \oplus B) = \mathrm{Co}(F(A) \cup F(B))$$

Proof: Note that $A \oplus B \in M_{n_1+n_2}$. Partition any given unit vector $z \in \mathbb{C}^{n_1+n_2}$ as $z = \begin{bmatrix} x \\ y \end{bmatrix}$, where $x \in \mathbb{C}^{n_1}$ and $y \in \mathbb{C}^{n_2}$. Then $z^*(A \oplus B)z = x^*Ax + y^*By$. If $y^*y = 1$ then $x = 0$ and $z^*(A \oplus B)z = y^*By \in F(B)$, so $F(A \oplus B) \supset F(B)$. By a similar argument when $x^*x = 1$, $F(A \oplus B) \supset F(A)$ and hence

$F(A \oplus B) \supset F(A) \cup F(B)$. But since $F(A \oplus B)$ is convex, it follows that $F(A \oplus B) \supset \mathrm{Co}(F(A) \cup F(B))$ (see Problem 21).

To prove the reverse containment, let $z = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{C}^{n_1+n_2}$ be a unit vector again. If $x^*x = 0$, then $y^*y = 1$ and $z^*(A \oplus B)z = y^*By \in F(B) \subset \mathrm{Co}(F(A) \cup F(B))$. The argument is analogous if $y^*y = 0$. Now suppose that both x and y are nonzero and write

$$z^*(A \oplus B)z = x^*Ax + y^*By = x^*x\left[\frac{x^*Ax}{x^*x}\right] + y^*y\left[\frac{y^*By}{y^*y}\right]$$

Since $x^*x + y^*y = z^*z = 1$, this last expression is a convex combination of

$$\frac{x^*Ax}{x^*x} \in F(A) \quad \text{and} \quad \frac{y^*By}{y^*y} \in F(B)$$

and we have $F(A \oplus B) \subset \mathrm{Co}(F(A) \cup F(B))$. □

1.2.11 **Property**: *Submatrix inclusion.* For all $A \in M_n$ and index sets $J \subset \{1,\dots, n\}$,

$$F(A(J)) \subset F(A)$$

Proof: Suppose that $J = \{j_1,\dots, j_k\}$, with $1 \le j_1 < j_2 < \cdots < j_k \le n$, and suppose $x \in \mathbb{C}^k$ satisfies $x^*x = 1$. We may insert zero entries into appropriate locations in x to produce a vector $\hat{x} \in \mathbb{C}^n$ such that $\hat{x}_{j_i} = x_i$, $i = 1, 2,\dots, k$, and $\hat{x}_j = 0$ for all other indices j. A calculation then shows that $x^*A(J)x = \hat{x}^*A\hat{x}$ and $\hat{x}^*\hat{x} = 1$, which verifies the asserted inclusion. □

1.2.12 **Property**: *Congruence and the angular field of values.* Let $A \in M_n$ and suppose that $C \in M_n$ is nonsingular. Then

$$F'(C^*AC) = F'(A)$$

Proof: Let $x \in \mathbb{C}^n$ be a nonzero vector, so that $x^*C^*ACx = y^*Ay$, where $y \equiv Cx \neq 0$. Thus, $F'(C^*AC) \subset F'(A)$. In the same way, one shows that $F'(A) \subset F'(C^*AC)$ since $A = (C^{-1})^*C^*AC(C^{-1})$. □

Problems

1. For what $A \in M_n$ does $F(A)$ consist of a single point? Could $F(A)$ consist of k distinct points for finite $k > 1$?

2. State and prove results corresponding to (1.2.5) and (1.2.5a,b) about projecting the field of values onto the imaginary axis and about when $F(A)$ is in the upper half-plane.

3. Show that $F(A) + F(B)$ is not the same as $F(A + B)$ in general. Why not?

4. If $A, B \in M_n$, is $F(AB) \subset F(A)F(B)$? Prove or give a counterexample.

5. If $A, B \in M_n$, is $F'(AB) \subset F'(A)F'(B)$? Is $\Theta(AB) \leq \Theta(A) + \Theta(B)$? Prove or give a counterexample.

6. If $A, B \in M_n$ are normal with $\sigma(A) = \{\alpha_1,\ldots, \alpha_n\}$ and $\sigma(B) = \{\beta_1,\ldots, \beta_n\}$, show that $\sigma(A + B) \subset \mathrm{Co}(\{\alpha_i + \beta_j\colon i, j = 1,\ldots, n\})$. If $0 \leq \alpha \leq 1$ and if $U, V \in M_n$ are unitary, show that $\rho(\alpha U + (1 - \alpha) V) \leq \alpha\rho(U) + (1 - \alpha)\rho(V) = 1$, where $\rho(\cdot)$ denotes the spectral radius.

7. If $A, B \in M_n$ are Hermitian with ordered eigenvalues $\alpha_1 \leq \cdots \leq \alpha_n$ and $\beta_1 \leq \cdots \leq \beta_n$, respectively, use a field of values argument with the subadditivity property (1.2.7) to show that $\alpha_1 + \beta_1 \leq \gamma_1$ and $\gamma_n \leq \alpha_n + \beta_n$, where $\gamma_1 \leq \cdots \leq \gamma_n$ are the ordered eigenvalues of $C = A + B$. What can you say if equality holds in either of the inequalities? Compare with the conclusions and proof of Weyl's theorem (4.3.1) in [HJ].

8. Which convex subsets of $\mathbb{C}$ are fields of values? Show that the class of convex subsets of $\mathbb{C}$ that are fields of values is closed under the operation of taking convex hulls of finite unions of the sets. Show that any convex polygon and any disc are in this class.

9. According to property (1.2.8), if two matrices are unitarily similar, they have the same field of values. Although the converse is true when $n = 2$ [see Problem 18 in Section (1.3)], it is not true in general. Construct two matrices that have the same size and the same field of values, but are not unitarily similar. A complete characterization of all matrices of a given size with a given field of values is unknown.

10. According to (1.2.9), if $A \in M_n$ is normal, then its field of values is the convex hull of its spectrum. Show that the converse is not true by consid-

ering the matrix $A \equiv \mathrm{diag}(1, i, -1, -i) \oplus \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \in M_6$. Show that $F(A) = \mathrm{Co}(\sigma(A))$, but that A is not normal. Construct a counterexample of the form $A = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3) \oplus \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \in M_5$. Why doesn't this kind of example work for M_4? Is there some other kind of counterexample in M_4? See (1.6.9).

11. Let z be a complex number with modulus 1, and let

$$A \equiv \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ z & 0 & 0 \end{bmatrix}$$

Show that $F(A)$ is the closed equilateral triangle whose vertices are the cube roots of z. More generally, what is $F(A)$ if $A = [a_{ij}] \in M_n$ has all $a_{i,i+1} = 1$, $a_{n1} = z$, and all other $a_{ij} = 0$?

12. If $A \in M_n$ and $F(A)$ is a real line segment, show that A is Hermitian.

13. Give an example of a matrix A and an index set J for which equality occurs in (1.2.11). Can you characterize the cases of equality?

14. What is the geometric relationship between $F'(C^*AC)$ and $F'(A)$ if: (a) $C \in M_n$ is singular, or (b) if C is n-by-k and rank $C = k$?

15. What is the relationship between $F(C^*AC)$ and $F(A)$ when $C \in M_n$ is nonsingular but not unitary? Compare with Sylvester's law of inertia, Theorem (4.5.8) in [HJ].

16. Properties (1.2.8) and (1.2.11) are special cases of a more general property. Let $A \in M_n$, $k \leq n$, and $P \in M_{n,k}$ be given. If $P^*P = I \in M_k$, then P is called an *isometry* and P^*AP is called an *isometric projection of* A. Notice that an isometry $P \in M_{n,k}$ is unitary if and only if $k = n$. If A' is a principal submatrix of A, show how to construct an isometry P such that $A' = P^*AP$. Prove the following statement and explain how it includes both (1.2.8) and (1.2.11):

1.2.13 Property: *Isometric projection.* For all $A \in M_n$ and $P \in M_{n,k}$ with $k \leq n$ and $P^*P = I$, $F(P^*AP) \subset F(A)$, and $F(P^*AP) = F(A)$ when $k = n$.

17. It is natural to inquire whether there are any nonunitary cases in which the containment in (1.2.13) is an equality. Let $A \in M_n$ be given, let $P \in M_{n,k}$ be a given isometry with $k < n$, and suppose the column space of P contains

all the columns of both A and A^*. Show that $F'(A) = F'(P^*AP) \cup \{0\}$ and $F(A) = \text{Co}(F(P^*AP) \cup \{0\})$. If, in addition, $0 \in F(P^*AP)$, show that $F(A) = F(P^*AP)$; consider $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $P = \begin{bmatrix} 1 \\ 1 \end{bmatrix}/\sqrt{2}$ to show that if $0 \notin F(P^*AP)$, then the containment $F(P^*AP) \subset F(A)$ can be strict.

18. Let $A \in M_n$ and let $P \in M_{n,k}$ be an isometry with $k \leq n$. Consider $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $P = \begin{bmatrix} 1 \\ 1 \end{bmatrix}/\sqrt{2}$ to show that one can have $F(P^*AP) = F(A)$ even if the column space of P does not contain the columns of A and A^*. Suppose $U \in M_n$ is unitary and $UAU^* = \Delta$ is upper triangular with $\Delta = \Delta_1 \oplus \cdots \oplus \Delta_p$ in which each Δ_i is upper triangular and some of the matrices Δ_i are diagonal. Describe how to construct an isometry P from selected columns of U so that $F(A) = F(P^*AP)$. Apply your construction to $A = I$, to $A = \text{diag}(1,2,3) \in M_3$, and to a normal $A \in M_n$ and discuss.

19. Suppose that $A \in M_n$ is given and that $G \in M_n$ is positive definite. Show that $\sigma(A) \subset F'(GA)$ and, therefore, that $\sigma(A) \subset \cap\{F'(GA)\colon\ G \text{ is positive definite}\}$.

20. Consider the two matrices

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

Precisely determine $F(A)$, $F(B)$, and $F(\alpha A + (1-\alpha)B)$, $0 \leq \alpha \leq 1$.

21. Let S be a given subset of a complex vector space, and let S_1 and S_2 be given subsets of S. Show that $S \supset \text{Co}(S_1 \cup S_2)$ if S is convex, but that this statement need not be correct if S is not convex. Notice that this general principle was used in the proof of the direct sum property (1.2.10).

22. Let $A \in M_n$ be nonsingular. Show that $F(A) \subset RHP$ if and only if $F(A^{-1}) \subset RHP$, or, equivalently, $\text{H}(A)$ is positive definite if and only if $\text{H}(A^{-1})$ is positive definite.

23. Let $A \in M_n$, and suppose λ is an eigenvalue of A.

(a) If λ is real, show that $\lambda \in F(\text{H}(A))$ and hence $|\lambda| \leq \rho(\text{H}(A)) = |||\text{H}(A)|||_2$, where $\rho(\cdot)$ and $|||\cdot|||_2$ denote the spectral radius and spectral norm, respectively.

(b) If λ is purely imaginary, show that $\lambda \in F(\mathrm{S}(A))$ and hence $|\lambda| \leq \rho(\mathrm{S}(A)) = |||\,\mathrm{S}(A)\,|||_2$.

24. Suppose all the entries of a given $A \in M_n$ are nonnegative. Then the nonnegative real number $\rho(A)$ is always an eigenvalue of A [HJ, Theorem (8.3.1)]. Use Problem 23 (a) to show that $\rho(A) \leq \rho(\mathrm{H}(A)) = \frac{1}{2}\rho(A + A^T)$ whenever A is a square matrix with nonnegative entries. See Problem 23 (n) in Section (1.5) for a better inequality involving the numerical radius of A.

25. We have seen that the field of values of $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is a closed disc of radius $\frac{1}{2}$ centered at the origin; this result has a useful generalization.

(a) Let $B \in M_{m,n}$ be given, let $A = \begin{bmatrix} 0 & B \\ 0 & 0 \end{bmatrix} \in M_{m+n}$, and let $\sigma_1(B)$ denote the largest singular value (the spectral norm) of B. Show that $F(A) = \{x^*By\colon x \in \mathbb{C}^m,\ y \in \mathbb{C}^n,\ \|x\|_2^2 + \|y\|_2^2 = 1\}$, and conclude that $F(A)$ is a closed disc of radius $\sigma_1(B)/2$ centered at the origin. In particular, $r(A) = \sigma_1(B)/2$.

(b) Consider $A = \begin{bmatrix} 0 & B \\ 0 & 0 \end{bmatrix} \in M_3$ with $B \equiv \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \in M_2$. Show that $\sigma_1(B) = 2$. Consider e^TAe with $e = [1, 1, 1]^T$ to show that $r(A) \geq 4/3 > \sigma_1(B)/2$. Does this contradict (a)?

26. Let $A \in M_n$ be given. Show that the following are equivalent:

(a) $r(A) \leq 1$.

(b) $\rho(\mathrm{H}(e^{i\theta}A)) \leq 1$ for all $\theta \in \mathbb{R}$.

(c) $\lambda_{max}(\mathrm{H}(e^{i\theta}A)) \leq 1$ for all $\theta \in \mathbb{R}$.

(d) $|||\,\mathrm{H}(e^{i\theta}A)\,|||_2 \leq 1$ for all $\theta \in \mathbb{R}$.

Convexity

In this section we prove the fundamental convexity property (1.2.2) of the field of values and discuss several important consequences of that convexity. We shall make use of several basic properties exhibited in the previous section. Our proof contains several useful observations and consists of three parts:

1. Reduction of the problem to the 2-by-2 case;
2. Use of various basic properties to transform the general 2-by-2 case to 2-by-2 matrices of special form; and

3. Demonstration of convexity of the field of values for the special 2-by-2 form.

See Problems 7 and 10 for two different proofs that do not involve reduction to the 2-by-2 case. There are other proofs in the literature that are based upon more advanced concepts from other branches of mathematics.

Reduction to the 2-by-2 case

In order to show that a given set $S \subset \mathbb{C}$ is convex, it is sufficient to show that $\alpha s + (1-\alpha)t \in S$ whenever $0 \leq \alpha \leq 1$ and $s, t \in S$. Thus, for a given $A \in M_n$, $F(A)$ is convex if $\alpha x^* Ax + (1-\alpha)y^* Ay \in F(A)$ whenever $0 \leq \alpha \leq 1$ and $x, y \in \mathbb{C}^n$ satisfy $x^*x = y^*y = 1$. It suffices to prove this only in the 2-by-2 case because we need to consider only convex combinations associated with *pairs* of vectors. For each given pair of vectors $x, y \in \mathbb{C}^n$, there is a unitary matrix U and vectors $v, w \in \mathbb{C}^n$ such that $x = Uv$, $y = Uw$, and all entries of v and w after the first two are equal to zero (see Problem 1). Using this transformation, we have

$$\begin{aligned}
\alpha x^* Ax + (1-\alpha)y^* Ay &= \alpha v^* U^* A U v + (1-\alpha)w^* U^* A U w \\
&= \alpha v^* B v + (1-\alpha) w^* B w \\
&= \alpha \xi^* B(\{1,2\})\xi + (1-\alpha)\eta^* B(\{1,2\})\eta
\end{aligned}$$

where $B \equiv U^* A U$, $B(\{1,2\})$ is the upper left 2-by-2 principal submatrix of B, and $\xi, \eta \in \mathbb{C}^2$ consist of the first two entries of v and w, respectively. Thus, it suffices to show that the field of values of any 2-by-2 matrix is convex. This reduction is possible because of the unitary similarity invariance property (1.2.8) of the field of values.

Sufficiency of a special 2-by-2 form

We prove next that in order to show that $F(A)$ is convex for every matrix $A \in M_2$, it suffices to demonstrate that $F\begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}$ is convex for any $a, b \geq 0$. The following observation is useful.

1.3.1 Lemma. For each $A \in M_2$, there is a unitary $U \in M_2$ such that the two main diagonal entries of $U^* A U$ are equal.

Proof: We may suppose without loss of generality that $\operatorname{tr} A = 0$ since we may replace A by $A - (\frac{1}{2}\operatorname{tr} A)I$. We need to show that A is unitarily similar to a matrix whose diagonal entries are both equal to zero. To show this, it is sufficient to show that there is a nonzero vector $w \in \mathbb{C}^2$ such that $w^*Aw = 0$. Such a vector may be normalized and used as the first column of a unitary matrix W, and a calculation reveals that the 1,1 entry of W^*AW is zero; the 2,2 entry of W^*AW must also be zero since the trace is zero. Construct the vector w as follows: Since A has eigenvalues $\pm\alpha$ for some complex number α, let x be a normalized eigenvector associated with $-\alpha$ and let y be a normalized eigenvector associated with $+\alpha$. If $\alpha = 0$, just take $w = x$. If $\alpha \neq 0$, x and y are independent and the vector $w = e^{i\theta}x + y$ is nonzero for all $\theta \in \mathbb{R}$. A calculation shows that $w^*Aw = \alpha(e^{-i\theta}x^*y - e^{i\theta}y^*x) = 2i\alpha\operatorname{Im}(e^{-i\theta}x^*y)$. Now choose θ so that $e^{-i\theta}x^*y$ is real. $\square$

We now use Lemma (1.3.1) together with several of the properties given in the previous section to reduce the question of convexity in the 2-by-2 case to consideration of the stated special form. If $A \in M_2$ is given, apply the translation property (1.2.3) to conclude that $F(A)$ is convex if and only if $F(A + \alpha I)$ is convex. If we choose $\alpha = -\frac{1}{2}\operatorname{tr} A$, we may suppose without loss of generality that our matrix has trace 0. According to (1.3.1) and the unitary similarity invariance property (1.2.8), we may further suppose that both main diagonal entries of our matrix are 0.

Thus, we may assume that the given matrix has the form $\begin{bmatrix} 0 & c \\ d & 0 \end{bmatrix}$ for some $c, d \in \mathbb{C}$. Now we can use the unitary similarity invariance property (1.2.8) and a diagonal unitary matrix to show that we may consider

$$\begin{bmatrix} 1 & 0 \\ 0 & e^{-i\theta} \end{bmatrix} \begin{bmatrix} 0 & c \\ d & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{bmatrix} = \begin{bmatrix} 0 & ce^{i\theta} \\ de^{-i\theta} & 0 \end{bmatrix}$$

for any $\theta \in \mathbb{R}$. If $c = |c|e^{i\theta_1}$ and $d = |d|e^{i\theta_2}$, and if we choose $\theta = \frac{1}{2}(\theta_2 - \theta_1)$, the latter matrix becomes $e^{i\varphi}\begin{bmatrix} 0 & |c| \\ |d| & 0 \end{bmatrix}$ with $\varphi = \frac{1}{2}(\theta_1 + \theta_2)$.

Thus, it suffices to consider a matrix of the form $e^{i\varphi}\begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}$ with $\varphi \in \mathbb{R}$ and $a, b \geq 0$. Finally, by the scalar multiplication property (1.2.4), we need to consider only the special form

$$\begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}, \qquad a, b \geq 0 \tag{1.3.2}$$

That is, we have shown that the field of values of every 2-by-2 complex matrix is convex if the field of values of every matrix of the special form (1.3.2) is convex.

Convexity of the field of values of the special 2-by-2 form

1.3.3 Lemma. If $A \in M_2$ has the form (1.3.2), then $F(A)$ is an ellipse (with its interior) centered at the origin. Its minor axis is along the imaginary axis and has length $|a-b|$. Its major axis is along the real axis and has length $a+b$. Its foci are at $\pm\sqrt{ab}$, which are the eigenvalues of A.

Proof: Without loss of generality, we assume $a \geq b \geq 0$. Since $z^*Az = (e^{i\theta}z)^*A(e^{i\theta}z)$ for any $\theta \in \mathbb{R}$, to determine $F(A)$, it suffices to consider z^*Az for unit vectors z whose first component is real and nonnegative. Thus, we consider the 2-vector $z = [t, e^{i\theta}(1-t^2)^{\frac{1}{2}}]^T$ for $0 \leq t \leq 1$ and $0 \leq \theta \leq 2\pi$. A calculation shows that

$$z^*Az = t(1-t^2)^{\frac{1}{2}}[(a+b)\cos\theta + i(a-b)\sin\theta]$$

As θ varies from 0 to 2π, the point $(a+b)\cos\theta + i(a-b)\sin\theta$ traces out a possibly degenerate ellipse $\mathcal{E}$ centered at the origin; the major axis extends from $-(a+b)$ to $(a+b)$ on the real axis and the minor axis extends from $i(b-a)$ to $i(a-b)$ on the imaginary axis in the complex plane. As t varies from 0 to 1, the factor $t(1-t^2)^{\frac{1}{2}}$ varies from 0 to $\frac{1}{2}$ and back to 0, ensuring that every point in the interior of the ellipse $\frac{1}{2}\mathcal{E}$ is attained and verifying that $F(A)$ is the asserted ellipse with its interior, which is convex. The two foci of the ellipse $\frac{1}{2}\mathcal{E}$ are located on the major axis at distance $[\frac{1}{4}(a+b)^2 - \frac{1}{4}(a-b)^2]^{\frac{1}{2}} = \pm\sqrt{ab}$ from the center. This completes the argument to prove the convexity property (1.2.2). $\square$

There are many important consequences of convexity of the field of values. One immediate consequence is that Lemma (1.3.1) holds for matrices of any size, not just for 2-by-2 matrices.

1.3.4 Theorem. For each $A \in M_n$ there is a unitary matrix $U \in M_n$ such that all the diagonal entries of U^*AU have the same value $\mathrm{tr}(A)/n$.

Proof: Without loss of generality, we may suppose that $\mathrm{tr}\, A = 0$, since we may replace A by $A - [\mathrm{tr}(A)/n]I$. We proceed by induction to show that A is unitarily similar to a matrix with all zero main diagonal entries. We know from Lemma (1.3.1) that this is true for $n = 2$, so let $n \geq 3$ and suppose that the assertion has been proved for all matrices of all orders less than n. We have

$$0 = \frac{1}{n}\operatorname{tr} A = \frac{1}{n}\lambda_1 + \frac{1}{n}\lambda_2 + \cdots + \frac{1}{n}\lambda_n = 0$$

and this is a convex combination of the eigenvalues λ_i of A. Since each λ_i is in $F(A)$, and since $F(A)$ is convex, we conclude that $0 \in F(A)$. If $x \in \mathbb{C}^n$ is a unit vector such that $x^*Ax = 0$, let $W = [x\ w_2\ \ldots\ w_n] \in M_n$ be a unitary matrix whose first column is x. One computes that

$$W^*AW = \begin{bmatrix} 0 & z^* \\ \zeta & \hat{A} \end{bmatrix}, \quad z, \zeta \in \mathbb{C}^{n-1}, \ \hat{A} \in M_{n-1}$$

But $0 = \operatorname{tr} A = \operatorname{tr} W^*AW = \operatorname{tr} \hat{A} = 0$, and so by the induction hypothesis there is some unitary $\hat{V} \in M_{n-1}$ such that all the main diagonal entries of $\hat{V}^* \hat{A} \hat{V}$ are zero. Define the unitary direct sum

$$V = \begin{bmatrix} 1 & 0 \\ 0 & \hat{V} \end{bmatrix} \in M_n$$

and compute

$$(WV)^*A(WV) = V^*W^*AWV = \begin{bmatrix} 0 & z^*\hat{V} \\ \hat{V}\zeta & \hat{V}^*\hat{A}\hat{V} \end{bmatrix}$$

which has a zero main diagonal by construction. □

A different proof of (1.3.4) using compactness of the set of unitary matrices is given in Problem 3 of Section (2.2) of [HJ].

Another important, and very useful, consequence of convexity of the field of values is the following *rotation property* of a matrix whose field of values does not contain the point 0.

1.3.5 Theorem. Let $A \in M_n$ be given. There exists a real number θ such that the Hermitian matrix $H(e^{i\theta}A) = \frac{1}{2}[e^{i\theta}A + e^{-i\theta}A^*]$ is positive definite if and only if $0 \notin F(A)$.

Proof: If $H(e^{i\theta}A)$ is positive definite for some $\theta \in \mathbb{R}$, then $F(e^{i\theta}A) \subset RHP$ by

(1.2.5a), so $0 \notin F(e^{i\theta}A)$ and hence $0 \notin F(A)$ by (1.2.4). Conversely, suppose $0 \notin F(A)$. By the separating hyperplane theorem (see Appendix B of [HJ]), there is a line L in the plane such that each of the two nonintersecting compact convex sets $\{0\}$ and $F(A)$ lies entirely within exactly one of the two open half-planes determined by L. The coordinate axes may now be rotated so that the line L is carried into a vertical line in the right half-plane with $F(A)$ strictly to the right of it, that is, for some $\theta \in \mathbb{R}$, $F(e^{i\theta}A) = e^{i\theta}F(A) \subset RHP$, so $H(e^{i\theta}A)$ is positive definite by (1.2.5a). □

Some useful information can be extracted from a careful examination of the steps we have taken to transform a given matrix $A \in M_2$ to the special form (1.3.2). The first step was a translation $A \rightarrow A - (\frac{1}{2}\text{tr } A)I \equiv A_0$ to achieve tr $A_0 = 0$. The second step was a unitary similarity $A_0 \rightarrow UA_0U^* \equiv A_1$ to make both diagonal entries of A_1 zero. The third step was another unitary similarity $A_1 \rightarrow VA_1V^* \equiv A_2$ to put A_2 into the form

$$A_2 = e^{i\varphi}\begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix} \text{ with } a, b \geq 0 \text{ and } \varphi \in \mathbb{R}$$

The last step was a unitary rotation $A_2 \rightarrow e^{-i\varphi}A_2 \equiv A_3$ to achieve the special form (1.3.2). Since the field of values of A_3 is an ellipse (possibly degenerate, that is, a point or line segment) centered at the origin with its major axis along the real axis and its foci at $\pm\sqrt{ab}$, the eigenvalues of A_3, the field of values of A_2 is also an ellipse centered at the origin, but its major axis is tilted at an angle φ to the real axis. A line through the two eigenvalues of A_2, $\pm e^{i\varphi}\sqrt{ab}$, which are the foci of the ellipse, contains the major axis of the ellipse; if $ab = 0$, the ellipse is a circle (possibly degenerate), so any diameter is a major axis. Since A_1 and A_0 are achieved from A_2 by successive unitary similarities, each of which leaves the eigenvalues and field of values invariant, we have $F(A_0) = F(A_1) = F(A_2)$. Finally,

$$F(A) = F(A_0 + [\tfrac{1}{2}\text{tr } A]I) = F(A_0) + \tfrac{1}{2}\text{tr } A = F(A_2) + \tfrac{1}{2}\text{tr } A$$

a shift that moves both eigenvalues by $\frac{1}{2}$tr A, so we conclude that the field of values of any matrix $A \in M_2$ is an ellipse (possibly degenerate) with center at the point $\frac{1}{2}$tr A. The major axis of this ellipse lies on a line through the two eigenvalues of A, which are the foci of the ellipse; if the two eigenvalues coincide, the ellipse is a circle or a point.

According to Lemma (1.3.3), the ellipse $F(A)$ is degenerate if and only if $A_3 = \begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}$ has $a = b$. Notice that $A_3^* A_3 = \begin{bmatrix} b^2 & 0 \\ 0 & a^2 \end{bmatrix}$ and $A_3 A_3^* = \begin{bmatrix} a^2 & 0 \\ 0 & b^2 \end{bmatrix}$, so $a = b$ if and only if A_3 is normal. But A can be recovered from A_3 by a nonzero scalar multiplication, two unitary similarities, and a translation, each of which preserves both normality and nonnormality. Thus, A_3 is normal if and only if A is normal, and we conclude that for $A \in M_2$, the ellipse $F(A)$ is degenerate if and only if A is normal.

The eigenvalues of A_3 are located at the foci on the major axis of $F(A_3)$ at a distance of $\sqrt{ab}$ from the center, and the length of the semimajor axis is $\frac{1}{2}(a+b)$ by (1.3.3). Thus, $\frac{1}{2}(a+b) - \sqrt{ab} = \frac{1}{2}(\sqrt{a} - \sqrt{b})^2 \geq 0$ with equality if and only if $a = b$, that is, if and only if A is normal. We conclude that the eigenvalues of a nonnormal $A \in M_2$ always lie in the interior of $F(A)$.

For $A \in M_2$, the parameters of the ellipse $F(A)$ (even if degenerate) can be computed easily using (1.3.3) if one observes that a and b are the singular values of $A_3 = \begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}$, that is, the square roots of the eigenvalues of $A_3^* A_3$, and the singular values of A_3 are invariant under pre- or post-multiplication by any unitary matrix. Thus, the singular values $\sigma_1 \geq \sigma_2 \geq 0$ of $A_0 = A - (\frac{1}{2}\text{tr } A)I$ are the same as those of A_3. The length of the major axis of $F(A)$ is $a + b = \sigma_1 + \sigma_2$, the length of the minor axis is $|a - b| = \sigma_1 - \sigma_2$, and the distance of the foci from the center is $[(a+b)^2 - (a-b)^2]^{\frac{1}{2}}/4 = \sqrt{ab} = \sqrt{\sigma_1\sigma_2} = |\det A_3|^{\frac{1}{2}} = |\det A_0|^{\frac{1}{2}}$. Moreover, $\sigma_1^2 + \sigma_2^2 = \text{tr } A_0^* A_0$ (the sum of the squares of the moduli of the entries of A_0), so $\sigma_1 \pm \sigma_2 = [\sigma_1^2 + \sigma_2^2 \pm 2\sigma_1\sigma_2]^{\frac{1}{2}} = [\text{tr } A_0^* A_0 \pm 2|\det A_0|]^{\frac{1}{2}}$. We summarize these observations for convenient reference in the following theorem.

1.3.6 Theorem. Let $A \in M_2$ be given, and set $A_0 \equiv A - (\frac{1}{2}\text{tr } A)I$. Then

(a) The field of values $F(A)$ is a closed ellipse (with interior, possibly degenerate).

(b) The center of the ellipse $F(A)$ is at the point $\frac{1}{2}\text{tr } A$. The length of the major axis is $[\text{tr } A_0^* A_0 + 2|\det A_0|]^{\frac{1}{2}}$; the length of the minor axis is $[\text{tr } A_0^* A_0 - 2|\det A_0|]^{\frac{1}{2}}$; the distance of the foci from the center is $|\det A_0|^{\frac{1}{2}}$. The major axis lies on a line passing through the two eigenvalues of A, which are the foci of $F(A)$; these two eigenvalues coincide if and only if the ellipse is a circle (possibly a point).

(c) $F(A)$ is a closed line segment if and only if A is normal; it is a single point if and only if A is a scalar matrix.

(d) $F(A)$ is a nondegenerate ellipse (with interior) if and only if A is not normal, and in this event the eigenvalues of A are interior points of $F(A)$.

Problems

1. Let $x, y \in \mathbb{C}^n$ be two given vectors. Show how to construct vectors $v, w \in \mathbb{C}^n$ and a unitary matrix $U \in M_n$ such that $x = Uv$, $y = Uw$, and all entries of v and w after the first two are zero.

2. Verify all the calculations in the proof of (1.3.1).

3. Sketch the field of values of a matrix of the form (1.3.2), with $a \geq 0$, $b \geq 0$.

4. Use (1.3.3) to show that the field of values of $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ is a closed ellipse (with interior) with foci at 0 and 1, major axis of length $\sqrt{2}$, and minor axis of length 1. Verify these assertions using Theorem (1.3.6).

5. Show that if $A \in M_n(\mathbb{R})$ then $F(A)$ is symmetric with respect to the real axis.

6. If $x_1,\ldots, x_k \in \mathbb{C}^n$ are given orthonormal vectors, let $P \equiv [x_1 \ldots x_k] \in M_{n,k}$ and observe that $P^*P = I \in M_k$. If $A \in M_n$, show that $F(P^*AP) \subset F(A)$ and that $x_i^*Ax_i \in F(P^*AP)$ for $i = 1,\ldots, k$. Use this fact for $k = 2$ to give an alternate reduction of the question of convexity of $F(A)$ to the 2-by-2 case.

7. Let $A \in M_n$ be given. Provide details for the following proof of the convexity of $F(A)$ that does not use a reduction to the 2-by-2 case. There is nothing to prove if $F(A)$ is a single point. Pick any two distinct points in $F(A)$. Using (1.2.3) and (1.2.4), there is no loss of generality to assume that these two points are 0, $a \in \mathbb{R}$, $a > 0$. We must show that the line segment joining 0 and a lies in $F(A)$. Write A as $A = H + iK$, in which $H = \mathrm{H}(A)$ and $K = -i\mathrm{S}(A) \equiv -i\frac{1}{2}(A - A^*)$, so that H and K are Hermitian. Assume further, without loss of generality after using a unitary similarity, that K is diagonal. Let $x, y \in \mathbb{C}^n$ satisfy $x^*Ax = 0$, $x^*x = 1$, $y^*Ay = a$, and $y^*y = 1$, and let $x_j = |x_j|e^{i\theta_j}$, $y_j = |y_j|e^{i\varphi_j}$, $j = 1,\ldots, n$. Note that $x^*Hx = x^*Kx = y^*Ky = 0$ and $y^*Hy = a$. Define $z(t) \in \mathbb{C}^n$, $0 \leq t \leq 1$, by

$$z_j(t) = \begin{cases} |x_j| e^{i(1-3t)\theta_j}, & 0 \le t \le 1/3 \\ \left[(2 - 3t)|x_j|^2 + (3t-1)|y_j|^2\right]^{\frac{1}{2}}, & 1/3 < t < 2/3 \\ |y_j| e^{i(3t-2)\varphi_j}, & 2/3 \le t \le 1 \end{cases}$$

Verify that $z^*(t)z(t) = 1$ and $z^*(t)Kz(t) = 0, 0 \le t \le 1$, and note that $z^*(t)Az(t) = z^*(t)Hz(t)$ is real and equal to 0 at $t = 0$, is equal to a at $t = 1$, and is a continuous function of t for $0 \le t \le 1$. Conclude that the line segment joining 0 and a lies in $F(A)$, which, therefore, is convex.

8. If $A \in M_n$ is such that $0 \in F(A)$, show that the Euclidean unit sphere in the definition of $F(A)$ may be replaced by the Euclidean unit ball, that is, show that $F(A) = \{x^*Ax : x \in \mathbb{C}^n, x^*x \le 1\}$. What if $0 \notin F(A)$?

9. Let $J_n(0) \in M_n$ be the n-by-n nilpotent Jordan block

$$J_n(0) \equiv \begin{bmatrix} 0 & 1 & & 0 \\ & 0 & 1 & \\ & & 0 & \ddots \\ & & & \ddots & 1 \\ 0 & & & & 0 \end{bmatrix}$$

Show that $F(J_n(0))$ is a disc centered at the origin with radius $\rho(\mathrm{H}(J_n(0)))$. Use this to show that $F(J_n(0))$ is strictly contained in the unit disc. If $D \in M_n$ is diagonal, show that $F(DJ_n(0))$ is also a disc centered at the origin with radius $\rho(\mathrm{H}(DJ_n(0))) = \rho(\mathrm{H}(|D|J_n(0)))$. See Problem 29 in Section (1.5) for a stronger result.

10. Let $A \in M_n$ be given. Provide details for the following proof of the convexity of $F(A)$ that does not use an explicit reduction to the 2-by-2 case. If $F(A)$ is not a single point, pick any two distinct points $a, b \in F(A)$, and let c be any given point on the open line segment between a and b. We may assume that $c = 0$, $a, b \in \mathbb{R}$, and $a < 0 < b$. Let $x, y \in \mathbb{C}^n$ be unit vectors such that $x^*Ax = a$, $y^*Ay = b$, so x and y are independent. Consider $z(t,\theta) \equiv e^{i\theta}x + ty$, where t and θ are real parameters to be determined. Show that $f(t,\theta) \equiv z(t,\theta)^*Az(t,\theta) = bt^2 + \alpha(\theta)t + a$, where $\alpha(\theta) \equiv e^{-i\theta}x^*Ay + e^{i\theta}y^*Ax$. If $y^*Ax - \overline{x^TA\bar{y}} = re^{i\varphi}$ with $r \ge 0$ and $\varphi \in \mathbb{R}$, then $\alpha(\theta)$ is real for $\theta = -\varphi$, and $f(t_0,-\varphi) = 0$ for $t_0 = [-\alpha(-\varphi) + [\alpha(-\varphi)^2 - 4ab]^{\frac{1}{2}}]/2b \in \mathbb{R}$. Then $z(t_0,-\varphi) \ne 0$ and $0 = z^*Az \in F(A)$ for $z \equiv z(t_0,-\varphi)/\|z(t_0,-\varphi)\|_2$.

11. Let $A \in M_2$ be given. Show that A is normal (and hence $F(A)$ is a line segment) if and only if $A_0 \equiv A - (\frac{1}{2}\mathrm{tr}\,A)I$ is a scalar multiple of a unitary

matrix, that is, $A_0^* A_0 = cI$ for some $c \geq 0$.

12. Give an example of a normal matrix $A \in M_n$, $n \geq 3$, such that $F(A)$ is not a line segment. Why can't this happen in M_2?

13. Let $A \in M_n$ be given. If $F(A)$ is a line segment or a point, show that A must be normal.

14. Consider the upper triangular matrix $A = \begin{bmatrix} \lambda_1 & \beta \\ 0 & \lambda_2 \end{bmatrix} \in M_2$. Show that the length of the major axis of the ellipse $F(A)$ is $[|\lambda_1 - \lambda_2|^2 + |\beta|^2]^{\frac{1}{2}}$ and the length of the minor axis is $|\beta|$. Where is the center? Where are the foci? In particular, conclude that the eigenvalues of A are interior points of $F(A)$ if and only if $\beta \neq 0$.

15. Let $A \in M_n$ be given with $0 \notin F(A)$. Provide the details for the following proof that $F(A)$ lies in an open half-plane determined by some line through the origin, and explain how this result may be used as an alternative to the argument involving the separating hyperplane theorem in the proof of (1.3.5). For $z \in F(A)$, consider the function $g(z) \equiv z/|z| = \exp(i \arg z)$. Since $g(\cdot)$ is a continuous function on the compact connected set $F(A)$, its range $g(F(A)) \equiv R$ is a compact connected subset of the unit circle. Thus, R is a closed arc whose length must be strictly less than π since $F(A)$ is convex and $0 \notin F(A)$.

16. Let $A \in M_n$ be given. Ignoring for the moment that we know the field of values $F(A)$ is convex, $F(A)$ is obviously nonempty, closed, and bounded, so its complement $F(A)^c$ has an unbounded component and possibly some bounded components. The *outer boundary* of $F(A)$ is the intersection of $F(A)$ with the closure of the unbounded component of $F(A)^c$. Provide details for the following proof (due to Toeplitz) that the outer boundary of $F(A)$ is a convex curve. For any given $\theta \in [0,2\pi]$, let $e^{i\theta}A = H + iK$ with Hermitian $H, K \in M_n$. Let $\lambda_n(H)$ denote the algebraically largest eigenvalue of H, let $S_\theta \equiv \{x \in \mathbb{C}^n\colon x \neq 0,\ Hx = \lambda_n(H)x\}$, and suppose $\dim(S_\theta) = k \geq 1$. Then the intersection of $F(e^{i\theta}A)$ with the vertical line $\operatorname{Re} z = \lambda_n(H)$ is the set $\lambda_n(H) + i\{x^* Kx\colon x \in S_\theta,\ \|x\|_2 = 1\}$, which is a single point if $k = 1$ and can be a finite interval if $k > 1$ (this is the convexity property of the field of values for a Hermitian matrix, which follows simply from the spectral theorem). Conclude, by varying θ, that the outer boundary of $F(A)$ is a convex curve, which may contain straight line segments. Why doesn't this prove that $F(A)$ is convex?

17. (a) Let $A \in M_2$ be given. Use Schur's unitary triangularization theorem (Theorem (2.3.1) in [HJ]) and a diagonal unitary similarity to show that A is unitarily similar to an upper triangular matrix of the form

$$\begin{bmatrix} \lambda_1 & \alpha(A) \\ 0 & \lambda_2 \end{bmatrix}$$

where $\alpha(A) \geq 0$. Show that $\operatorname{tr} A^*A = |\lambda_1|^2 + |\lambda_2|^2 + \alpha(A)^2$ and that $\alpha(A)$ is a unitary similarity invariant, that is, $\alpha(A) = \alpha(UAU^*)$ for any unitary $U \in M_2$.
(b) If $A, B \in M_2$ have the same eigenvalues, show that A is unitarily similar to B if and only if $\operatorname{tr} A^*A = \operatorname{tr} B^*B$.
(c) Show that $A, B \in M_2$ have the same eigenvalues if and only if $\operatorname{tr} A = \operatorname{tr} B$ and $\operatorname{tr} A^2 = \operatorname{tr} B^2$.
(d) Conclude that two given matrices $A, B \in M_2$ are unitarily similar if and only if $\operatorname{tr} A = \operatorname{tr} B$, $\operatorname{tr} A^2 = \operatorname{tr} B^2$, and $\operatorname{tr} A^*A = \operatorname{tr} B^*B$.

18. Use Theorem (1.3.6) and the preceding problem to show that two given 2-by-2 complex or real matrices are unitarily similar if and only if their fields of values are identical. Consider $A_t = \operatorname{diag}(0,1,t)$ for $t \in [0,1]$ to show that nonsimilar 3-by-3 Hermitian matrices can have the same fields of values.

19. Let $A = [a_{ij}] \in M_2$ and suppose that $F(A) \subset UHP_0 = \{z \in \mathbb{C}\colon \operatorname{Im} z \geq 0\}$. Use Theorem (1.3.6) to show that if either (a) a_{11} and a_{22} are real, or (b) $\operatorname{tr} A$ is real, then A is Hermitian.

Notes and Further Readings. The convexity of the field of values (1.2.2) was first discussed in O. Toeplitz, Das algebraische Analogon zu einem Satze von Fejér, *Math. Zeit.* 2 (1918), 187-197, and F. Hausdorff, Das Wertvorrat einer Bilinearform, *Math. Zeit.* 3 (1919), 314-316. Toeplitz showed that the outer boundary of $F(A)$ is a convex curve, but left open the question of whether the interior of this curve is completely filled out with points of $F(A)$; see Problem 16 for Toeplitz's elegant proof. He also proved the inequality $|||A|||_2 \leq 2r(A)$ between the spectral norm and the numerical radius; see Problem 21 in Section (5.7) of [HJ]. In a paper dated six months after Toeplitz's (the respective dates of their papers were May 22 and November 28, 1918), Hausdorff rose to the challenge and gave a short proof, similar to the argument outlined in Problem 7, that $F(A)$ is actually a convex set. There are many other proofs, besides the elementary one given in this section, such as that of W. Donoghue, On the Numerical Range of a Bounded Operator, *Mich. Math. J.* 4 (1957), 261-263. The result of Theorem (1.3.4) was first noted by W. V. Parker in Sets of Complex Numbers Associated with a Matrix, *Duke Math. J.* 15 (1948), 711-715. The modification of Hausdorff's original convexity proof outlined in Problem 7 was given by Donald Robinson; the proof outlined in Problem 10 is due to Roy Mathias.

The fact that the field of values of a square complex matrix is convex has an immediate extension to the infinite-dimensional case. If T is a bounded linear operator on a complex Hilbert space $\mathcal{H}$ with inner product $<\cdot,\cdot>$, then its field of values (often called the *numerical range*) is $F(T) \equiv \{<Tx,x>: x \in \mathcal{H} \text{ and } <x,x> = 1\}$. One can show that $F(T)$ is convex by reducing to the two-dimensional case, just as we did in the proof in this section.

1.4 Axiomatization

It is natural to ask (for both practical and aesthetic reasons) whether the list of properties of $F(A)$ given in Section (1.2) is, in some sense, complete. Since special cases and corollary properties may be of interest, it may be that no finite list is truly complete; but a mathematically precise version of the completeness question is whether or not, among the properties given thus far, there is a subset that characterizes the field of values. If so, then further properties, and possibly some already noted, would be corollary to a set of characterizing properties, and the mathematical utility of the field of values would be captured by these properties. This does not mean that it is not useful to write down properties beyond a characterizing set. Some of the most applicable properties do follow, if tediously, from others.

1.4.1 Example. Spectral containment (1.2.6) follows from compactness (1.2.1), translation (1.2.3), scalar multiplication (1.2.4), unitary invariance (1.2.8), and submatrix inclusion (1.2.11) in the sense that any set-valued function on M_n that has these five properties also satisfies (1.2.6). If $A \in M_n$ and if $\beta \in \sigma(A)$, then for some unitary $U \in M_n$ the matrix U^*AU is upper triangular with β in the 1,1 position. Then by (1.2.8) and (1.2.11) it is enough to show that $\beta \in F([\beta])$, and (because of (1.2.3)) it suffices to show that $0 \in F([0])$; here we think of $[\beta]$ and $[0]$ as members of M_1 However, because of (1.2.4), $F([0]) = \alpha F([0])$ for any α, and there are only two nonempty subsets of the complex plane possessing this property: $\{0\}$ and the entire plane. The latter is precluded by (1.2.1), and hence (1.2.6) follows.

Exercise. Show that (1.2.8) and (1.2.10) together imply (1.2.9).

The main result of this section is that there is a subset of the properties already mentioned that characterizes $F(\cdot)$ as a function from M_n into subsets of $\mathbb{C}$.

1.4.2 Theorem. Properties (1.2.1–4 and 5b) characterize the field of values. That is, the usual field of values $F(\cdot)$ is the only complex set-valued function on M_n such that

(a) $F(A)$ is compact (1.2.1) and convex (1.2.2) for all $A \in M_n$;

(b) $F(A + \alpha I) = F(A) + \alpha$ (1.2.3) and $F(\alpha A) = \alpha F(A)$ (1.2.4) for all $\alpha \in \mathbb{C}$ and all $A \in M_n$; and

(c) $F(A)$ is a subset of the closed right half-plane if and only if $A + A^*$ is positive semidefinite (1.2.5b).

Proof: Suppose $F_1(\cdot)$ and $F_2(\cdot)$ are two given complex set-valued functions on M_n that satisfy the five cited functional properties. Let $A \in M_n$ be given. We first show that $F_1(A) \subset F_2(A)$. Suppose, to the contrary, that $\beta \in F_1(A)$ and $\beta \notin F_2(A)$ for some complex number β. Then because of (1.2.1) and (1.2.2), there is a straight line L in the complex plane that has the point β strictly on one side of it and the set $F_2(A)$ on the other side (by the separating hyperplane theorem for convex sets; see Appendix B of [HJ]). The plane may be rotated and translated so that the imaginary axis coincides with L and β lies in the open left half-plane. That is, there exist complex numbers $\alpha_1 \neq 0$ and α_2 such that $\mathrm{Re}(\alpha_1\beta + \alpha_2) < 0$ while $\alpha_1 F_2(A) + \alpha_2$ is contained in the closed right half-plane. However, $\alpha_1 F_2(A) + \alpha_2 = F_2(\alpha_1 A + \alpha_2 I)$ because of (1.2.3) and (1.2.4), so $\mathrm{Re}\,\beta' < 0$ while $F_2(A')$ lies in the closed right half-plane, where $A' \equiv \alpha_1 A + \alpha_2 I$ and $\beta' \equiv \alpha_1\beta + \alpha_2 \in F_1(A')$. Then, by (1.2.5b), $A' + A'^*$ is positive semidefinite. This, however, contradicts the fact that $F_1(A')$ is not contained in the closed right half-plane, and we conclude that $F_1(A) \subset F_2(A)$.

Reversing the roles of $F_1(\cdot)$ and $F_2(\cdot)$ shows that $F_2(A) \subset F_1(A)$ as well. Thus, $F_1(A) = F_2(A)$ for all $A \in M_n$. Since the usual field of values $F(\cdot)$ satisfies the five stated properties (1.2.1–4 and 5b), we obtain the desired conclusion by taking $F_1(A) \equiv F(A)$. $\square$

Problems

1. Show that no four of the five properties cited in Theorem (1.4.2) are sufficient to characterize $F(\cdot)$; that is, each subset of four of the five properties in the theorem is satisfied by some complex set-valued function other than $F(\cdot)$.

2. Show that (1.2.2-4 and 5a) also characterize $F(\cdot)$, and that no subset of these four properties is sufficient to characterize $F(\cdot)$.

3. Determine other characterizing sets of properties. Can you find one that does not contain (1.2.2)?

4. Show that the complex set-valued function $F(\cdot) \equiv \mathrm{Co}(\sigma(\cdot))$ is characterized by the four properties (1.2.1-4) together with the fifth property "$F(A)$ is contained in the closed right half-plane if and only if all the eigenvalues of A have nonnegative real parts" (A is *positive semistable*).

5. Give some other complex set-valued functions on M_n that satisfy (1.2.7).

Further Reading. This section is based upon C. R. Johnson, Functional Characterization of the Field of Values and the Convex Hull of the Spectrum, *Proc. Amer. Math. Soc.* 61 (1976), 201-204.

1.5 Location of the field of values

Thus far we have said little about where the field of values $F(A)$ sits in the complex plane, although it is clear that knowledge of its location could be useful for applications such as those mentioned in (1.0.2-4). In this section, we give a Geršgorin-type inclusion region for $F(A)$ and some observations that facilitate its numerical determination.

Because the eigenvalues of a matrix depend continuously upon its entries and because the eigenvalues of a diagonal matrix are the diagonal entries, it is not too surprising that there is a spectral location result such as Geršgorin's theorem (6.1.1) in [HJ]. We argue by analogy that because the set $F(A)$ depends continuously upon the entries of A and because the field of values of a diagonal matrix is the convex hull of the diagonal entries, there ought to be some sort of inclusion region for $F(A)$ that, like the Geršgorin discs, depends in a simple way on the entries of A. We next present such an inclusion region.

1.5.1 Definition. Let $A = [a_{ij}] \in M_n$, let the deleted absolute row and column sums of A be denoted by

$$R_i'(A) = \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}|, \qquad i = 1,\ldots, n$$

and

$$C_j'(A) = \sum_{\substack{i=1\\ i\neq j}}^{n} |a_{ij}|, \qquad j = 1,\ldots, n$$

respectively, and let

$$g_i(A) = \tfrac{1}{2}[R_i'(A) + C_i'(A)],\ i = 1,\ldots, n$$

be the average of the ith deleted absolute row and column sums of A. Define the complex set-valued function $G_F(\cdot)$ on M_n by

$$G_F(A) \equiv \mathrm{Co}\left[\bigcup_{i=1}^{n} \{z\colon |z - a_{ii}| \le g_i(A)\}\right]$$

Recall Geršgorin's theorem about the spectrum $\sigma(A)$ of $A \in M_n$, which says that

$$\sigma(A) \subset G(A) \equiv \bigcup_{i=1}^{n} \{z\colon |z - a_{ii}| \le R_i'(A)\}$$

and that

$$\sigma(A) \subset G(A^T) \equiv \bigcup_{j=1}^{n} \{z\colon |z - a_{jj}| \le C_j'(A)\}$$

Note that the Geršgorin regions $G(A)$ and $G(A^T)$ are unions of circular discs with centers at the diagonal entries of A, while the set function $G_F(A)$ is the convex hull of some circular discs with the same centers, but with radii that are the average of the radii leading to $G(A)$ and $G(A^T)$, respectively. The Geršgorin regions $G(A)$ and $G(A^T)$ need not be convex, but $G_F(A)$ is convex by definition.

Exercise. Show that the set-valued function $G_F(\cdot)$ satisfies the properties (1.2.1-4) of the field of values. Which additional properties of the field of values function $F(\cdot)$ does $G_F(\cdot)$ share?

Exercise. Show that $R_i'(A) = C_i'(A^T)$ and use the triangle inequality to show that $R_i'(\mathrm{H}(A)) = R_i'(\frac{1}{2}[A + A^*]) \leq g_i(A)$ for all $i = 1, \ldots, n$.

Our Geršgorin-type inclusion result for the field of values is the following:

1.5.2 Theorem. For any $A \in M_n$, $F(A) \subset G_F(A)$.

Proof: The demonstration has three steps. Let RHP denote the open right half-plane $\{z \in \mathbb{C}\colon \operatorname{Re} z > 0\}$. We first show that if $G_F(A) \subset RHP$, then $F(A) \subset RHP$. If $G_F(A) \subset RHP$, then $\operatorname{Re} a_{ii} > g_i(A)$. Let $\mathrm{H}(A) = \frac{1}{2}(A + A^*) \equiv B = [b_{ij}]$. Since $R_i'(A^*) = C_i'(A)$ and $R_i'(B) \leq g_i(A)$ (by the preceding exercise), it follows that $b_{ii} = \operatorname{Re} a_{ii} > g_i(A) \geq R_i'(B)$. In particular, $G(B) \subset RHP$. Since $\sigma(B) \subset G(B)$ by Geršgorin's theorem, we have $\sigma(B) \subset RHP$. But since B is Hermitian, $F(\mathrm{H}(A)) = F(B) = \mathrm{Co}(\sigma(B)) \subset RHP$ and hence $F(A) \subset RHP$ by (1.2.5).

We next show that if $0 \notin G_F(A)$, then $0 \notin F(A)$. Suppose $0 \notin G_F(A)$. Since $G_F(A)$ is convex, there is some $\theta \in [0, 2\pi)$ such that $G_F(e^{i\theta}A) = e^{i\theta}G_F(A) \subset RHP$. As we showed in the first step, this means that $F(e^{i\theta}A) \subset RHP$, and, since $F(A) = e^{-i\theta}F(e^{i\theta}A)$, it follows that $0 \notin F(A)$.

Finally, if $\alpha \notin G_F(A)$, then $0 \notin G_F(A - \alpha I)$ since the set function $G_F(\cdot)$ satisfies the translation property. By what we have just shown, it follows that $0 \notin F(A - \alpha I)$ and hence $\alpha \notin F(A)$, so $F(A) \subset G_F(A)$. □

A simple bound for the numerical radius $r(A)$ (1.1.4) follows directly from Theorem (1.5.2).

1.5.3 Corollary. For all $A \in M_n$,

$$r(A) \leq \max_{1 \leq i \leq n} \tfrac{1}{2} \sum_{j=1}^{n} (|a_{ij}| + |a_{ji}|)$$

It follows immediately from (1.5.3) that

$$r(A) \leq \tfrac{1}{2}\left[\max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}| + \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ji}|\right]$$

and the right-hand side of this inequality is just the average of the maximum absolute row and column sum matrix norms, $|||A|||_1$ and $|||A|||_\infty$ (see Section (5.6) of [HJ]).

1.5.4 Corollary. For all $A \in M_n$, $r(A) \leq \tfrac{1}{2}(|||A|||_1 + |||A|||_\infty)$.

A norm on matrices is called *spectrally dominant* if, for all $A \in M_n$, it is an upper bound for the spectral radius $\rho(A)$. It is apparent from the spectral containment property (1.2.6) that the numerical radius $r(\cdot)$ is spectrally dominant.

1.5.5 Corollary. For all $A \in M_n$, $\rho(A) \leq r(A) \leq \tfrac{1}{2}(|||A|||_1 + |||A|||_\infty)$.

We next discuss a procedure for determining and plotting $F(A)$ numerically. Because the set $F(A)$ is convex and compact, it suffices to determine the boundary of $F(A)$, which we denote by $\partial F(A)$. The general strategy is to calculate many well-spaced points on $\partial F(A)$ and support lines of $F(A)$ at these points. The convex hull of these boundary points is then a convex polygonal approximation to $F(A)$ that is contained in $F(A)$, while the intersection of the half-spaces determined by the support lines is a convex polygonal approximation to $F(A)$ that contains $F(A)$. The area of the region between these two convex polygonal approximations may be thought of as a measure of how well either one approximates $F(A)$. Furthermore, if the boundary points and support lines are produced as one traverses $\partial F(A)$ in one direction, it is easy to plot these two approximating polygons. A pictorial summary of this general scheme is given in Figure (1.5.5.1). The points q_i are at intersections of consecutive support lines and, therefore, are the vertices of the external approximating polygon.

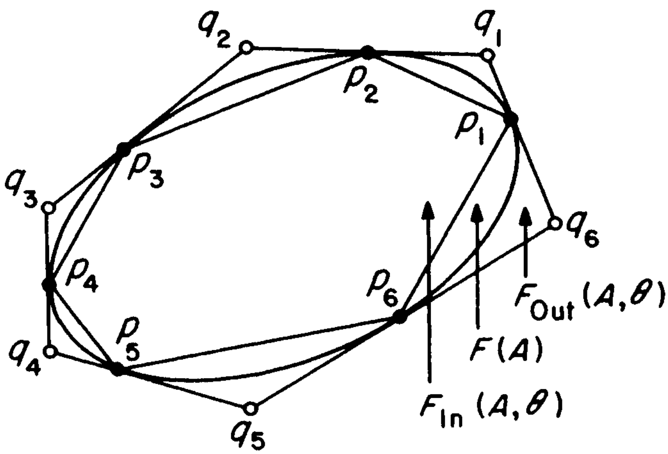

Figure 1.5.5.1

The purpose of the next few observations is to show how to produce boundary points and support lines around $\partial F(A)$.

From (1.2.4) it follows that

$$e^{-i\theta}F(e^{i\theta}A) = F(A) \tag{1.5.6}$$

for every $A \in M_n$ and all $\theta \in [0,2\pi)$. Furthermore, we have the following lemma.

1.5.7 Lemma. If $x \in \mathbb{C}^n$, $x^*x = 1$, and $A \in M_n$, the following three conditions are equivalent:

(a) $\operatorname{Re} x^*Ax = \max\{\operatorname{Re} \alpha\colon \alpha \in F(A)\}$

(b) $x^*\mathrm{H}(A)x = \max\{r\colon r \in F(\mathrm{H}(A))\}$

(c) $\mathrm{H}(A)x = \lambda_{max}(\mathrm{H}(A))x$

where $\lambda_{max}(B)$ denotes the algebraically largest eigenvalue of the Hermitian matrix B.

Proof: The equivalence of (a) and (b) follows from the calculation Re $x^*Ax = \frac{1}{2}(x^*Ax + x^*A^*x) = x^*\mathrm{H}(A)x$ and the projection property (1.2.5). If $\{y_1,\ldots, y_n\}$ is an orthonormal set of eigenvectors of the Hermitian matrix $\mathrm{H}(A)$ and if $\mathrm{H}(A)y_j = \lambda_j y_j$, then x may be written as

$$x = \sum_{j=1}^{n} c_j y_j, \text{ with } \sum_{j=1}^{n} \bar{c}_j c_j = 1$$

since $x^*x = 1$. Thus,

$$x^*\mathrm{H}(A)x = \sum_{j=1}^{n} \bar{c}_j c_j \lambda_j$$

from which the equivalence of (b) and (c) is immediately deduced. □

It follows from the lemma that

$$\max\{\mathrm{Re}\,\alpha : \alpha \in F(A)\} = \max\{r : r \in F(\mathrm{H}(A))\} = \lambda_{max}(\mathrm{H}(A)) \quad (1.5.8)$$

This means that the furthest point to the right in $F(\mathrm{H}(A))$ is the real part of the furthest point to the right in $F(A)$, which is $\lambda_{max}(\mathrm{H}(A))$. A unit vector yielding any one of these values also yields the others.

Lemma (1.5.7) shows that, if we compute $\lambda_{max}(\mathrm{H}(A))$ and an associated unit eigenvector x, we obtain a boundary point x^*Ax of $F(A)$ and a support line $\{\lambda_{max}(\mathrm{H}(A)) + ti : t \in \mathbb{R}\}$ of the convex set $F(A)$ at this boundary point; see Figure (1.5.8.1).

Using (1.5.6), however, one can obtain as many such boundary points and support lines as desired by rotating $F(A)$ and carrying out the required eigenvalue-eigenvector calculation. For an angle $\theta \in [0, 2\pi)$, we define

$$\lambda_\theta \equiv \lambda_{max}(\mathrm{H}(e^{i\theta}A)) \quad (1.5.9)$$

and let $x_\theta \in \mathbb{C}^n$ be an associated unit eigenvector

$$\mathrm{H}(e^{i\theta}A)x_\theta = \lambda_\theta x_\theta, \qquad x_\theta^* x_\theta = 1 \quad (1.5.10)$$

We denote

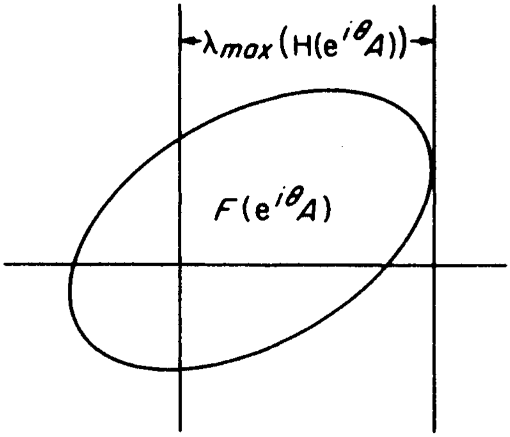

Figure 1.5.8.1

$$L_\theta \equiv \text{the line } \{e^{-i\theta}(\lambda_\theta + ti)\colon t \in \mathbb{R}\}$$

and denote the half-plane determined by the line L_θ by

$$H_\theta \equiv \text{the half-plane } e^{-i\theta}\{z\colon \text{Re } z \leq \lambda_\theta\}$$

Based upon (1.5.6), (1.5.7), and the preceding discussion, we then have the following:

1.5.11 Theorem. For each $A \in M_n$ and each $\theta \in [0,2\pi)$, the complex number $p_\theta \equiv x_\theta^* A x_\theta$ is a boundary point of $F(A)$. The line L_θ is a support line for $F(A)$, with $p_\theta \in L_\theta \cap F(A)$ and $F(A) \subset H_\theta$ for all $\theta \in [0,2\pi)$.

Because $F(A)$ is convex, it is geometrically clear that each extreme point of $F(A)$ occurs as a p_θ and that for any $\alpha \notin F(A)$ there is an L_θ separating $F(A)$ and α, that is, $\alpha \notin H_\theta$. Thus, we may represent $F(A)$ in the following way:

1.5.12 Theorem. For all $A \in M_n$,

$$F(A) = \mathrm{Co}(\{p_\theta: 0 \le \theta < 2\pi\}) = \bigcap_{0 \le \theta < 2\pi} H_\theta$$

Since it is not possible to compute infinitely many points p_θ and lines L_θ, we must be content with a discrete analog of (1.5.12) with equalities replaced by set containments. Let Θ denote a set of angular mesh points, $\Theta = \{\theta_1, \theta_2, \ldots, \theta_k\}$, where $0 \le \theta_1 < \theta_2 < \cdots < \theta_k < 2\pi$.

1.5.13 Definition. Let $A \in M_n$ be given, let a finite set of angular mesh points $\Theta = \{0 \le \theta_1 < \cdots < \theta_k < 2\pi\}$ be given, let $\{p_{\theta_i}\}$ be the associated set of boundary points of $F(A)$ given by (1.5.11), and let $\{H_{\theta_i}\}$ be the half-spaces associated with the support lines L_{θ_i} for $F(A)$ at the points p_{θ_i}. Then we define

$$F_{In}(A,\Theta) \equiv \mathrm{Co}(\{p_{\theta_1}, \ldots, p_{\theta_k}\}), \text{ and}$$

$$F_{Out}(A,\Theta) \equiv H_{\theta_1} \cap \cdots \cap H_{\theta_k}$$

These are the constructive inner and outer approximating sets for $F(A)$, as illustrated in Figure (1.5.5.1).

1.5.14 Theorem. For every $A \in M_n$ and every angular mesh Θ,

$$F_{In}(A,\Theta) \subset F(A) \subset F_{Out}(A,\Theta)$$

The set $F_{Out}(A,\Theta)$ is most useful as an outer estimate if the angular mesh points θ_j are sufficiently numerous and well spaced that the set $\cap\{H_{\theta_i}: 1 \le i \le k\}$ is bounded (which we assume henceforth). In this case, it is also simply determined. Let q_{θ_i} denote the (finite) intersection point of L_{θ_i} and $L_{\theta_{i+1}}$, where $i = 1, \ldots, k$ and $i = k+1$ is identified with $i = 1$. The existence of these intersection points is equivalent to the assumption that $F_{Out}(A,\Theta)$ is bounded, in which case we have the following simple alternate representation of $F_{Out}(A,\Theta)$:

$$F_{Out}(A,\Theta) = \bigcap_{1 \le i \le k} H_{\theta_i} = \mathrm{Co}(\{q_{\theta_1}, \ldots, q_{\theta_k}\}) \tag{1.5.15}$$

Because of the ordering of the angular mesh points θ_j, the points p_{θ_j} and q_{θ_j}

occur consecutively around $\partial F(A)$ for $j = 1,\ldots, k$, and $\partial F_{In}(A,\Theta)$ is just the union of the k line segments $[p_{\theta_1}, p_{\theta_2}],\ldots, [p_{\theta_{k-1}}, p_{\theta_k}], [p_{\theta_k}, p_{\theta_1}]$ while $\partial F_{Out}(A,\Theta)$ consists of the k line segments $[q_{\theta_1}, q_{\theta_2}], \ldots, [q_{\theta_{k-1}}, q_{\theta_k}], [q_{\theta_k}, q_{\theta_1}]$. Thus, each approximating set is easily plotted, and the difference of their areas (or some other measure of their set difference), which is easily calculated (see Problem 10), may be taken as a measure of the closeness of the approximation. If the approximation is not sufficiently close, a finer angular mesh may be used; because of (1.5.12), such approximations can be made arbitrarily close to $F(A)$. It is interesting to note that $F(A)$ (which contains $\sigma(A)$ for all $A \in M_n$), may be approximated arbitrarily closely with only a series of Hermitian eigenvalue-eigenvector computations.

These procedures allow us to calculate the numerical radius $r(A)$ as well. The following result is an immediate consequence of Theorem (1.5.14).

1.5.16 Corollary. For each $A \in M_n$ and every angular mesh Θ,

$$\max_{1 \le i \le k} |p_{\theta_i}| \le r(A) \le \max_{1 \le i \le k} |q_{\theta_i}|$$

Recall from (1.2.11) that $F[A(i')] \subset F(A)$ for $i = 1,\ldots, n$, where $A(i')$ denotes the principal submatrix of $A \in M_n$ formed by deleting row and column i. It follows from (1.2.2) that

$$\mathrm{Co}\left[\bigcup_{i=1}^{n} F[A(i')]\right] \subset F(A)$$

A natural question to ask is: How much of the right-hand side does the left-hand side fill up? The answer is "all of it in the limit," as the dimension goes to infinity. In order to describe this fact conveniently, we define Area(S) for a convex subset S of the complex plane to be the conventional area unless S is a line segment (possibly a point), in which case Area(S) is understood to be the length of the line segment (possibly zero). Furthermore, in the following area ratio we take 0/0 to be 1.

1.5.17 Theorem. For $A \in M_n$ with $n \ge 2$, let $A(i') \in M_{n-1}$ denote the principal submatrix of A obtained by deleting row and column i from A. There exists a sequence of constants $c_2, c_3, \ldots \in [0,1]$ such that for any $A \in M_n$,

$$\frac{\text{Area}\left[\text{Co}\left[\bigcup_{i=1}^{n} F[A(i')]\right]\right]}{\text{Area}[F(A)]} \geq c_n, \qquad n = 2, 3, \ldots$$

and $\lim c_n = 1$ as $n \to \infty$. The constants

$$c_n \equiv \frac{n-2}{7n - 2 - 6[n(n-2)]^{\frac{1}{2}}} \tag{1.5.18a}$$

satisfy these conditions, and so do the constants

$$c_n' \equiv \frac{2n-5}{2n+7} \tag{1.5.18b}$$

Problems

1. Show by example that there is no general containment relation between any two of the sets $\text{Co}(G(A))$, $\text{Co}(G(A^T))$, and $G_F(A)$.

2. Show that neither $r(A) \leq |||A|||_1$ nor $r(A) \leq |||A|||_\infty$ holds in general. Comment in light of (1.5.4).

3. Show that $A = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$ is a case of equality in Theorem (1.5.2). What are other cases of equality?

4. Show that $F(A) \subset \cap\{G_F(U^*AU): U \in M_n \text{ is unitary}\}$. Is this containment an equality for all $A \in M_n$? Prove or give a counterexample.

5. Show that (1.5.2) is the best possible inclusion region based solely on the $3n$ pieces of information a_{ii}, $R_i'(A)$, and $C_i'(A)$, $i = 1, \ldots, n$. Let $s(x,y)$ be a given function on $[0,\infty) \times [0,\infty)$, and define

$$G_s(A) = \text{Co}\left[\bigcup_{i=1}^{n} \{z : |z - a_{ii}| \leq s(R_i'(A), C_i'(A))\}\right]$$

If $F(A) \subset G_s(A)$ for all $A \in M_n$, show that $s(x,y) \geq \frac{1}{2}(x+y)$.

6. Let $A = [a_{ij}] \in M_n$ with $a_{ij} = 1$ if exactly one of i, j equals 1, and $a_{ij} = 0$ otherwise. Let $B = [b_{ij}] \in M_n$ with $b_{ij} = 0$ if $i = j$ and $b_{ij} = 1$ if $i \neq j$. Compare $G_F(A)$ and $G_F(B)$ with $F(A)$ and $F(B)$ to give examples of how imprecisely $G_F(\cdot)$ can estimate $F(\cdot)$.

7. Suppose the diagonal entries of $A \in M_n$ are positive. If A is row diagonally dominant *or* column diagonally dominant, then A is positive stable. Show that DA is positive stable for all positive diagonal D. Show that if A is row diagonally dominant *and* column diagonally dominant, then PA is positive stable for all positive definite matrices P. See (1.0.3).

8. A real square matrix with nonnegative entries and all row and column sums equal to $+1$ is said to be *doubly stochastic* (see Section (8.7) of [HJ]). Let $A \in M_n$ be doubly stochastic.

(a) Use (1.5.2) to show that $F(A)$ is contained in the unit disc. Is there any smaller disc that contains $F(A)$ for all doubly stochastic $A \in M_n$?
(b) Let $P \in M_n$ be a permutation matrix. Observe that P is doubly stochastic and describe $F(P)$ explicitly in terms of the eigenvalues of P; where are they?
(c) Now use Birkhoff's theorem (8.7.1) in [HJ], (1.2.4), and (1.2.7) to give another proof that $F(A)$ is contained in the unit disc.
(d) From what you have proved about $F(A)$, what can you say about the location of the eigenvalues of a doubly stochastic matrix? How does this compare with what you already know from the Perron-Frobenius theorem (8.3.1) in [HJ]? See Problems 18-21 for a sharper statement about $\sigma(A)$.

9. Show that the points q_{θ_j} in (1.5.15) are given by

$$q_{\theta_j} = e^{-i\theta_j}\left[\lambda_{\theta_j} + i\,\frac{\lambda_{\theta_j}\cos\delta_j - \lambda_{\theta_{j+1}}}{\sin\delta_j}\right]$$

where $\delta_j \equiv \theta_{j+1} - \theta_j$.

10. If P is a k-sided polygon in the complex plane with vertices $c_1, \ldots, c_k$ given in counterclockwise order, show that the area of P may be written as

$$\text{Area}\,(P) = \tfrac{1}{2}\text{Im}\left[\bar{c}_1 c_2 + \bar{c}_2 c_3 + \cdots + \bar{c}_{k-1}\, c_k + \bar{c}_k c_1\right]$$

Use this identity to formulate the accuracy estimate

$$\text{Area}(F_{Out}(A,\Theta)) - \text{Area}\,(F_{In}(A,\Theta))$$

in terms of the points p_{θ_i} and q_{θ_i}. What other measures of accuracy might be calculated easily?

11. If $A \in M_n$, show that the numerical radius of A is given by

$$r(A) = \max_{0 \le \theta < 2\pi} |p_\theta| = \max_{0 \le \theta < 2\pi} |\lambda_\theta| = \max_{0 \le \theta < 2\pi} \lambda_\theta$$

12. If $A \in M_n$, show that $F(A)$ is inscribed in a rectangle with vertical sides parallel to the imaginary axis going through the smallest and largest eigenvalues of $\mathrm{H}(A)$, respectively, and with horizontal sides parallel to the real axis going through i times the smallest and largest eigenvalues, respectively, of $-i\mathrm{S}(A)$. In particular, conclude that $\max\{\mathrm{Re}\,\lambda\colon \lambda \in \sigma(A)\} \le \max\{\lambda\colon \lambda \in \sigma(\mathrm{H}(A))\}$.

13. For $A = \begin{bmatrix} -1 & 2 \\ 0 & 1 \end{bmatrix}$ and angular mesh $\Theta = \{0, \pi/3, 2\pi/3, \pi, 4\pi/3, 5\pi/3\}$, explicitly determine and sketch $F_{In}(A,\Theta)$ and $F_{Out}(A,\Theta)$.

14. If $A \in M_n$ is normal, describe a simpler procedure for plotting $F(A)$.

15. Let $A \in M_n$ have constant main diagonal, and let $A_i \in M_{n-1}$ denote the principal submatrix formed by deleting row and column i from A. Show that

$$\mathrm{Area}\left[\mathrm{Co}\left[\bigcup_{i=1}^{n} F(A_i)\right]\right] / \mathrm{Area}(F(A)) \ge \frac{n-2}{n}$$

that is, a possible choice of the constants c_n in Theorem (1.5.17) in this special case is $c_n = (n-2)/n$.

16. Let $A \in M_n$ be a given Hermitian matrix. If A is positive definite, show that every principal submatrix $A_i \in M_{n-1}$ (formed by deleting row i and column i from A) is also positive definite. Show by example that the simple converse of this statement is false: Each A_i can be positive definite and A indefinite. Now let α denote the smallest and β the largest of all the eigenvalues of the A_i, $i = 1,\ldots, n$, and let $\{c_k\}_{k=1}^{\infty}$ be any sequence of constants satisfying the conditions in Theorem (1.5.17). Show that if all A_i are positive definite and $(\alpha/\beta) > 1 - c_n$, then A is positive definite.

17. Consider the constants c_n and c_n' defined in (1.5.18a,b). Show that the sequence $\{c_n'\}$ increases monotonically to 1, $c_n \ge c_n'$ for $n = 2, 3,\ldots$, $\lim c_n/c_n' = 1$ as $n \to \infty$, and the approximation $c_n \sim c_n'$ as $n \to \infty$ is quite good since $c_n - c_n' = 9/(2n^3) + O(1/n^4)$.

The following four problems improve Problem 8's simple estimate that the field of values of a doubly stochastic matrix is contained in the unit disc.

18. Let $m \geq 2$ and let $Q = [q_{ij}] \in M_m$ be a *cyclic permutation matrix*, that is, $q_{i,i+1} = 1$ for $i = 1,\ldots, m-1$, $q_{m,1} = 1$, and all other $q_{ij} = 0$. What does this have to do with a cyclic permutation of the integers $\{1,\ldots, m\}$? Show that $F(Q)$ is the convex hull of the mth roots of unity $e^{2\pi ik/m}$, $k = 1,\ldots, m$, which is contained in the angular wedge L_m in the complex plane with vertex at $z = 1$ and passing through $e^{\pm 2\pi i/m}$:

$$L_m \equiv \left\{z = 1 + r e^{i\theta}: r \geq 0, \ |\theta - \pi| \leq \frac{m-2}{2m}\pi\right\} \tag{1.5.19}$$

Sketch L_m. Notice that the directed graph of Q contains a *simple cycle of length* m, that is, a directed path that begins and ends at the same node, which occurs exactly twice in the path, and no other node occurs more than once in the path; see (6.2.12) in [HJ].

19. For any $B \in M_n$, define $\mu(B) \equiv$ the length of a maximum-length simple cycle in the directed graph of B. Notice that $0 \leq \mu(B) \leq n$, and $\mu(B) = n$ if and only if B is irreducible. Now let $B_1, B_2,\ldots, B_k \in M_n$ be nonnegative matrices, and let $\alpha_1, \alpha_2,\ldots, \alpha_k > 0$. Show that $\mu(\alpha_1 B_1 + \cdots + \alpha_k B_k) \geq \max\{\mu(B_1),\ldots, \mu(B_k)\}$.

20. Let $P \in M_n$ be a given permutation matrix. Since any permutation of the integers $\{1,\ldots, n\}$ is a product of cyclic permutations, explain why P is permutation similar (hence, unitarily similar) to a direct sum of cyclic permutation matrices $Q_1,\ldots, Q_k$, $Q_i \in M_{n_i}$, $1 \leq n_i \leq n$, $n_1 + \cdots + n_k = n$, where $\mu(P) = \max\{n_1,\ldots, n_k\}$. Use (1.2.10) to show that $F(P)$ is contained in the wedge $L_{\mu(P)}$ defined by (1.5.19).

21. Let $A \in M_n$ be a given doubly stochastic matrix. Use Birkhoff's theorem (8.7.1) in [HJ] to write A as a finite positive convex combination of permutation matrices, that is,

$$A = \sum_i \alpha_i P_i, \quad \text{all } \alpha_i > 0, \quad \sum_i \alpha_i = 1$$

Although there may be many ways to express A as such a sum, show that $\mu(A) \geq \max_i \mu(P_i)$ always, and hence we have the uniform upper bound $F(P_i) \subset L_{\mu(A)} \subset L_n$ for all i. Use (1.2.4) and (1.2.7) to show that $F(A) \subset$

$L_{\mu(A)} \subset L_n$. Conclude that $F(A)$ is contained in the part of the wedge $L_{\mu(A)}$ (and L_n) that is contained in the unit disc. Sketch this set. What does this say when $n = 2$? When $\mu(A) = 2$? When A is tridiagonal? From what you have proved about $F(A)$, what can you say about $\sigma(A)$? Do not use the fact that a doubly stochastic tridiagonal matrix must be symmetric.

22. Let $A \in M_n(\mathbb{R})$ have nonnegative entries and spectral radius 1. Let $\mu(A)$ denote the length of the maximum-length simple cycle in the directed graph of A. It is a theorem of Kellogg and Stephens that $\sigma(A)$ is contained in the intersection of the unit disc and the angular wedge $L_{\mu(A)}$ defined by (1.5.19).

(a) Compare this statement with the result of Problem 21.

(b) If $\mu(A) = 2$, use the Kellogg-Stephens theorem to show that all the eigenvalues of A are real, and deduce that all the eigenvalues of a nonnegative tridiagonal matrix are real. See Problem 5 of Section (4.1) of [HJ] for a different proof of this fact.

(c) Use the Kellogg-Stephens theorem to show that all the eigenvalues of a general n-by-n nonnegative matrix with spectral radius 1 are contained in the intersection of the unit disc and the wedge L_n. Sketch this set. A precise, but rather complicated, description of the set of all possible eigenvalues of a nonnegative matrix with spectral radius 1 has been given by Dmitriev, Dynkin, and Karpelevich.

23. Verify the following facts about the *numerical radius* function $r(\cdot)$: $M_n \rightarrow \mathbb{R}_+$ defined by $r(A) = \max\{|z| : z \in F(A)\}$. We write $\rho(A)$ for the spectral radius, $|||A|||_2$ for the spectral norm, $\|A\|_2$ for the Frobenius norm, and $|||A|||_1$, $|||A|||_\infty$ for the maximum column sum and maximum row sum matrix norms, respectively. The following statements hold for all $A, B \in M_n$. See Section (5.7) of [HJ] and its problems for background information about the numerical radius and its relationship to other norms.

(a) $r(\cdot)$ is a norm on M_n but is not a matrix norm.

(b) $4r(\cdot)$ is submultiplicative with respect to the ordinary matrix product, that is, $4r(AB) \leq 4r(A)4r(B)$, and 4 is the least positive constant with this property. Thus, $4r(\cdot)$ is a matrix norm on M_n.

(c) $2r(\cdot)$ is submultiplicative with respect to the Hadamard product, that is, $2r(A \circ B) \leq 2r(A)2r(B)$, and 2 is the least positive constant with this property, but if either A or B is normal then $r(A \circ B) \leq r(A)r(B)$; see Corollaries (1.7.24-25).

(d) $r(A^m) \leq [r(A)]^m$ for all $m = 1, 2, \ldots$; this is the *power inequality* for the numerical radius. Give an example to show that it is *not* always true that $r(A^{k+m}) \le r(A^k)r(A^m)$.

(e) $\rho(A) \le r(A)$.

(f) $r(A) = r(A^*)$.

(g) $\frac{1}{2}|||A|||_2 \le r(A) \le |||A|||_2$ and both bounds are sharp.

(h) $c_n\|A\|_2 \le r(A) \le \|A\|_2$ with $c_n = (4n)^{-\frac{1}{2}}$, and the upper, but not the lower, bound is sharp. A sharp lower bound is given by $c_n = (2n)^{-\frac{1}{2}}$ if n is even and $c_n = (2n-1)^{-\frac{1}{2}}$ if n is odd.

(i) $r(A) \le |||A|||_2 \le (|||A|||\ |||A^*|||)^{\frac{1}{2}}$ for any matrix norm $|||\cdot|||$ on M_n. In particular, $r(A) \le |||A|||_2 \le (|||A|||_1 |||A|||_\infty)^{\frac{1}{2}} \le \frac{1}{2}(|||A|||_1 + |||A|||_\infty)$, an inequality discovered by I. Schur; see Problem 21 in Section (5.6) of [HJ].

(j) $r(\hat{A}) \le r(A)$ for any principal submatrix $\hat{A}$ of A.

(k) $r(A_1 \oplus A_2) = \max\{r(A_1), r(A_2)\}$ for any $A_1 \in M_k$ and $A_2 \in M_m$.

(l) $r(A) \le r(|A|)$, where $A = [a_{ij}]$ and $|A| \equiv [|a_{ij}|]$.

(m) If $A \ge 0$, that is, all the entries of A are real and nonnegative, then $\rho(A) \le r(A) = r(\mathrm{H}(A)) = \rho(\mathrm{H}(A))$, where $\mathrm{H}(A) \equiv \frac{1}{2}(A + A^T)$ is the symmetric part of A; moreover, $r(A) = \max\{x^TAx: x \in \mathbb{R}^n, x \ge 0, x^Tx = 1\}$.

(n) $r(A) \le r(|A|) = \frac{1}{2}\rho(|A| + |A|^T)$, where $A = [a_{ij}]$ and $|A| \equiv [|a_{ij}|]$.

(o) $|||A^m|||_2 \le 2\, r(A)^m$ for all $m = 1, 2, \ldots$.

24. Let $A \in M_n$ be given. In the preceding problem we saw that $\rho(A) \le r(A) \le |||A|||_2$, so $\rho(A) = |||A|||_2$ implies $r(A) = |||A|||_2$. A. Wintner proved in 1929 that the converse is also true: $r(A) = |||A|||_2$ if and only if $\rho(A) = |||A|||_2$. Prove this. A matrix $A \in M_n$ is said to be *radial* (some authors use the term *radialoid*) if $\rho(A) = |||A|||_2$, or, equivalently, if $r(A) = |||A|||_2$.

25. Provide details for the following sketch for a proof of a theorem of Ptak: Let $A \in M_n$ be given with $|||A|||_2 \le 1$. Then $\rho(A) = 1$ if and only if $|||A^n|||_2 = 1$. *Proof:* $\mathcal{S} \equiv \{x \in \mathbb{C}^n: \|x\|_2 = \|Ax\|_2\}$ is a subspace of $\mathbb{C}^n$ since it is the nullspace of the positive semidefinite matrix $I - A^*A$. If $|||A^n|||_2 = 1$, let $x_0 \in \mathbb{C}^n$ be a unit vector such that $\|A^nx_0\|_2 = \|x_0\|_2$. Then $\|x_0\|_2 = \|A^nx_0\|_2 \le \|A^{n-1}x_0\|_2 \le \cdots \le \|Ax_0\|_2 \le \|x_0\|_2$, so all these inequalities are equalities and $\mathcal{S}_0 \equiv \mathrm{Span}\{x_0, Ax_0, \ldots, A^{n-1}x_0\} \subset \mathcal{S}$, $k \equiv \dim \mathcal{S}_0 \ge 1$, and $A\mathcal{S}_0 \subset \mathcal{S}_0$

by the Cayley-Hamilton theorem. Let $U=[U_1 \quad U_2] \in M_n$ be a unitary matrix such that the columns of $U_1 \in M_{n,k}$ form an orthonormal basis of S_0. Then U^*AU is a block triangular matrix of the form

$$U^*AU = \begin{bmatrix} U_1^*AU_1 & * \\ 0 & * \end{bmatrix}$$

where $U_1^*AU_1$ is a Euclidean isometry on S_0 (a linear transformation from S_0 to S_0 that preserves the Euclidean length of every vector) and hence is unitary (see (2.1.4-5) in [HJ]). Thus, $1 = \rho(U_1^*AU_1) \le \rho(U^*AU) = \rho(A) \le |||A|||_2 \le 1$.

26. Modify the proof of the preceding problem to show that if $A \in M_n$ has $|||A|||_2 \le 1$, then $\rho(A) = 1$ if and only if $|||A^m|||_2 = 1$ for some positive integer m greater than or equal to the degree of the minimal polynomial of A.

27. Let $A \in M_n$ be given. Combine the results of the preceding three problems and verify that the following conditions are equivalent:

(a) A is radial.
(b) $\rho(A) = |||A|||_2$.
(c) $r(A) = |||A|||_2$.
(d) $|||A^n|||_2 = (|||A|||_2)^n$.
(e) $|||A^m|||_2 = (|||A|||_2)^m$ for some integer m not less than the degree of the minimal polynomial of A.
(f) $|||A^k|||_2 = (|||A|||_2)^k$ for all $k = 1, 2, \ldots$.

In addition, one more equivalent condition follows from Problem 37(f) in Section (1.6); see Problem 38 in Section (1.6).

(g) A is unitarily similar to $|||A|||_2(U \oplus B)$, where $U \in M_k$ is unitary, $1 \le k \le n$, and $B \in M_{n-k}$ has $\rho(B) < 1$ and $|||B|||_2 \le 1$.

28. (a) If $A \in M_n$ is normal, show that A is radial. If A is radial and $n = 2$, show that A is normal. If $n \ge 3$, exhibit a radial matrix in M_n that is not normal.
(b) Show that if $A \in M_n$ is radial, then every positive integer power of A is radial, but it is possible for some positive integer power of a matrix to be radial without the matrix being radial.

29. Let $J_n(0)$ denote the n-by-n nilpotent Jordan block considered in Problem 9 in Section (1.3), where it was shown that $F(J_n(0))$ is a closed disc centered at the origin with radius $r(J_n(0)) = \rho(\mathrm{H}(J_n(0))$. If k is a positive integer such that $2k \leq n$, show that $2^{-1/k} \leq r(J_n(0)) < 1$. Describe $F(J_n(0))$ as $n \to \infty$.

30. A given $A \in M_{m,n}$ is called a *contraction* if its spectral norm satisfies $|||A|||_2 \leq 1$; it is a *strict contraction* if $|||A|||_2 < 1$. T. Ando has proved two useful representations for the unit ball of the numerical radius norm in M_n:

(a) $r(A) \leq 1$ if and only if

$$A = (I - Z)^{\frac{1}{2}} C (I + Z)^{\frac{1}{2}} \tag{1.5.20}$$

for some contractions $C, Z \in M_n$ with Z Hermitian; the indicated square roots are the unique positive semidefinite square roots of $I \pm Z$.

(b) $r(A) \leq 1$ if and only if

$$A = 2(I - C^* C)^{\frac{1}{2}} C \tag{1.5.21}$$

for some contraction $C \in M_n$.

Verify half of Ando's theorem—show that both representations (1.5.20, 21) give matrices A for which $r(A) \leq 1$. See Problem 44 in Section (3.1) for an identity related to (1.5.21).

31. Use Ando's representation (1.5.20) to show the following for $A \in M_n$:

(a) $r(A^m) \leq [r(A)]^m$ for all $m = 1, 2, \ldots$. This is the famous *power inequality* for the numerical radius; see Problem 23(d).

(b) $r(A) \leq 1$ if and only if there is a Hermitian $Z \in M_n$ such that the 2-by-2 block matrix

$$\begin{bmatrix} I - Z & A \\ A^* & I + Z \end{bmatrix}$$

is positive semidefinite.

See Lemma (3.5.12).

32. Let $p(t)$ be a polynomial, all of whose coefficients are nonnegative. Show that $r(p(A)) \leq p(r(A))$ for all $A \in M_n$, and that the same inequality holds if the numerical radius $r(\cdot)$ is replaced by the spectral radius $\rho(\cdot)$ or by any vector norm $\|\cdot\|$ on M_n that satisfies the power inequality $\|A^k\| \leq$

$\|A\|^k$ for $k = 1, 2,\ldots$ for all $A \in M_n$.

Notes and Further Readings. Most of this section is based on A Geršgorin Inclusion Set for the Field of Values of a Finite Matrix, *Proc. Amer. Math. Soc.* 41 (1973), 57-60, and Numerical Determination of the Field of Values of a General Complex Matrix, *SIAM J. Numer. Anal.* 15 (1978), 595-602, both by C. R. Johnson. See also C. R. Johnson, An Inclusion Region for the Field of Values of a Doubly Stochastic Matrix Based on Its Graph, *Aequationes Mathematicae* 17 (1978), 305-310. For a proof of Theorem (1.5.17), see C. R. Johnson, Numerical Ranges of Principal Submatrices, *Linear Algebra Applic.* 37 (1981), 11-22; the best possible value for the lower bound c_n in each dimension n is not yet known. For a proof of the theorem mentioned at the beginning of Problem 22, see R. B. Kellogg and A. B. Stephens, Complex Eigenvalues of a Non-Negative Matrix with a Specified Graph, *Linear Algebra Appl.* 20 (1978), 179-187. For the precise description of the set of all possible eigenvalues of n-by-n nonnegative matrices mentioned at the end of Problem 22, see N. Dmitriev and E. Dynkin, Eigenvalues of a Stochastic Matrix, *Izv. Akad. Nauk SSSR Ser. Mat.* 10 (1946), 167-184, and F. I. Karpelevich, On the Eigenvalues of a Matrix with Non-Negative Elements, *Izv. Akad. Nauk SSSR Ser. Mat.* 15 (1951), 361-383. A discussion of the bounds in Problem 23 (h) is in C. R. Johnson and C.-K. Li, Inequalities Relating Unitarily Invariant Norms and the Numerical Range, *Linear Multilinear Algebra* 23 (1988), 183-191. The theorem mentioned in Problem 24 is in A. Wintner, Zur Theorie der beschränkten Bilinearformen, *Math. Zeit.* 30 (1929), 228-282. Ptak's theorem mentioned in Problem 25 was first published in 1960; the proof we have outlined, and many interesting related results, is in V. Ptak, Lyapunov Equations and Gram Matrices, *Linear Algebra Applic.* 49 (1983), 33-55. For surveys of results about the numerical radius and field of values, generalizations, applications to finite difference methods for hyperbolic partial differential equations, and an extensive bibliography, see M. Goldberg, On Certain Finite Dimensional Numerical Ranges and Numerical Radii, *Linear Multilinear Algebra* 7 (1979), 329-342, as well as M. Goldberg and E. Tadmor, On the Numerical Radius and its Applications, *Linear Algebra Applic.* 42 (1982), 263-284. The representations (1.5.20, 21) for the unit ball of the numerical radius norm in M_n are in T. Ando, Structure of Operators with Numerical Radius One, *Acta. Sci. Math.* (*Szeged*) 34 (1973), 11-15.

1.6 Geometry

In Section (1.3) we saw that if $A \in M_2$, then the field of values $F(A)$ is a (possibly degenerate) ellipse with interior and that the ellipse may be determined in several ways from parameters associated with the entries of the matrix A.

Exercise. If $A = \begin{bmatrix} \lambda_1 & \beta \\ 0 & \lambda_2 \end{bmatrix} \in M_2$, show that $F(A)$ is: (a) a point if and only if $\lambda_2 = \lambda_1$ and $\beta = 0$; (b) a line segment joining λ_1 and λ_2 if and only if $\beta = 0$; (c) a circular disc of radius $\frac{1}{2}|\beta|$ if and only if $\lambda_2 = \lambda_1$; or (d) an ellipse (with interior) with foci at λ_1 and λ_2 otherwise, and λ_1 and λ_2 are interior points of the ellipse in this case.

In higher dimensions, however, a considerably richer variety of shapes is possible for the field of values. Any convex polygon is the field of values of a matrix of sufficiently high dimension; by (1.2.9), one can use a normal matrix whose eigenvalues are the vertices of the polygon. Thus, any bounded convex set can be approximated as closely as desired by the field of values of some matrix, but the dimension of the matrix may have to be large. Further interesting shapes can be pieced together using the direct sum property (1.2.10); if two sets occur, then so does their convex hull, but a direct sum of higher dimension may be required to realize it. Except for $n = 1$ and 2, it is not known which compact convex sets in $\mathbb{C}$ occur as fields of values for matrices of a fixed finite dimension n. Our purpose in this section is to present a few basic facts about the shape of $F(A)$ and its relation to $\sigma(A)$. These results elaborate upon some of the basic properties in (1.2).

The relative interior of a set in $\mathbb{C}$ is just its interior relative to the smallest dimensional space in which it sits. A compact convex set in the complex plane is either a point (with no relative interior), a nontrivial closed line segment (whose relative interior is an open line segment), or a set with a two-dimensional interior in the usual sense. What can be said about interior points of $F(A)$? If $A \in M_n$, the point $\frac{1}{n}\operatorname{tr} A$ always lies in $F(A)$ since it is a convex combination of points in $F(A)$ (the eigenvalues as well as the main diagonal entries of A), but somewhat more can be said. The key observation is that if a convex set C is contained in the closed upper half-plane and if a strict convex combination of given points in C is real, then all of these points must be real.

1.6.1 Theorem Let $A = [a_{ij}] \in M_n$ be given, let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of A, and let $\mu_1,\ldots,\mu_n$ be given positive real numbers such that $\mu_1+\cdots+\mu_n=1$. The following are equivalent:

(a) A is not a scalar multiple of the identity.
(b) $\mu_1 a_{11}+\cdots+\mu_n a_{nn}$ is in the relative interior of $F(A)$.
(c) $\mu_1\lambda_1+\cdots+\mu_n\lambda_n$ is in the relative interior of $F(A)$.

In particular, we have the following dichotomy: Either A is a scalar multiple of the identity or the point $\frac{1}{n}\operatorname{tr} A$ lies in the relative interior of $F(A)$.

Proof: If A is a scalar multiple of the identity, then $F(A)$ has no relative interior, so both (b) and (c) imply (a).

Let $\zeta\equiv\mu_1 a_{11}+\cdots+\mu_n a_{nn}$, observe that $\zeta\in F(A)$, and suppose that ζ is not in the relative interior of $F(A)$. Let $B=[b_{ij}]\equiv A-\zeta I$ and observe that $\mu_1 b_{11}+\cdots+\mu_n b_{nn}=0$, $0\in F(B)$, and 0 is not in the relative interior of $F(B)$. Since $F(B)$ is convex, it can be rotated about the boundary point 0 so that, for some $\theta\in\mathbb{R}$, the field of values of $C=[c_{ij}]\equiv e^{i\theta}B$ lies in the closed upper half-plane $UHP\equiv\{z\in\mathbb{C}\colon \operatorname{Im} z\geq 0\}$. Then $0\in F(C)\subset UHP$, 0 is not in the relative interior of $F(C)$, and $\mu_1 c_{11}+\cdots+\mu_n c_{nn}=0$. Since each $c_{ii}\in F(C)$, $\operatorname{Im} c_{ii}\geq 0$ for all $i=1,\ldots,n$. But $\mu_1\operatorname{Im} c_{11}+\cdots+\mu_n\operatorname{Im} c_{nn}=0$ and each $\mu_i>0$, so all the main diagonal entries of C are real. For any given indices $i,j\in\{1,\ldots,n\}$, let Γ_{ij} denote the 2-by-2 principal submatrix of C obtained as the intersections of rows and columns i and j, and let λ_1,λ_2 denote its eigenvalues. Since $F(\Gamma_{ij})\subset F(C)\subset UHP$, we have $\lambda_1,\lambda_2\in UHP$. But $\operatorname{Im}(\lambda_1)+\operatorname{Im}(\lambda_2)=\operatorname{Im}(\lambda_1+\lambda_2)=\operatorname{Im}(\operatorname{tr}\Gamma_{ij})=0$, so λ_1 and λ_2 are both real and neither is an interior point of $F(\Gamma_{ij})\subset UHP$. It follows from Theorem (1.3.6(d)) that Γ_{ij} is normal; it is actually Hermitian since its eigenvalues are real. Since i and j are arbitrary indices, it follows that C is Hermitian and hence $F(C)$ is a real interval, which must have 0 as an endpoint since 0 is not in the relative interior of $F(C)$. Thus, every point in $F(C)$, in particular, every c_{ii} and every λ_i, has the same sign. But $\mu_1 c_{11}+\cdots+\mu_n c_{nn}=0$ and all $\mu_i>0$, so all $c_{ii}=0$. Since $\lambda_1+\cdots+\lambda_n=c_{11}+\cdots+c_{nn}$, it follows that all $\lambda_i=0$ as well, and hence $C=0$. We conclude that $A=\zeta I$ and hence (a) implies (b).

To show that (a) implies (c), suppose $\zeta\equiv\mu_1\lambda_1+\cdots+\mu_n\lambda_n$ is not in the relative interior of $F(A)$. Choose a unitary $U\in M_n$ such that $\mathcal{A}\equiv UAU^*$ is upper triangular and has main diagonal entries $\lambda_1,\ldots,\lambda_n$. Then $F(A)=F(\mathcal{A})$, and a strict convex combination of the main diagonal entries of $\mathcal{A}$ is not in the relative interior of its field of values. By the equivalence of (a) and (b) we conclude that $\mathcal{A}=\zeta I$ and hence $A=U^*\mathcal{A}U=\zeta I$. $\square$

We next investigate the smoothness of the boundary of the field of values $\partial F(A)$ and its relationship with $\sigma(A)$. Intuitively, a "sharp point" of a convex set S is an extreme point at which the boundary takes an abrupt turn, a boundary point where there are nonunique tangents, or a "corner."

1.6.2 Definition. Let $A \in M_n$. A point $\alpha \in \partial F(A)$ is called a *sharp point* of $F(A)$ if there are angles θ_1 and θ_2 with $0 \leq \theta_1 < \theta_2 < 2\pi$ for which

$$\operatorname{Re} e^{i\theta}\alpha = \max \{\operatorname{Re} \beta : \beta \in F(e^{i\theta}A)\} \text{ for all } \theta \in (\theta_1, \theta_2)$$

1.6.3 Theorem. Let $A \in M_n$. If α is a sharp point of $F(A)$, then α is an eigenvalue of A.

Proof: If α is a sharp point of $F(A)$, we know from (1.5.7) that there is a unit vector $x \in \mathbb{C}^n$ such that

$$x^* \mathrm{H}(e^{i\theta}A)x = \lambda_{max}(\mathrm{H}(e^{i\theta}A)) \qquad \text{for all } \theta \in (\theta_1, \theta_2)$$

from which it follows that (see (1.5.9))

$$\mathrm{H}(e^{i\theta}A)x = \lambda_\theta x \qquad \text{for all } \theta \in (\theta_1, \theta_2) \tag{1.6.3a}$$

The vector x is independent of θ. Differentiation of (1.6.3a) with respect to θ yields

$$\mathrm{H}(i\, e^{i\theta}A)\, x = \lambda_\theta' x$$

which is equivalent to

$$S(e^{i\theta}A)\, x = -i\lambda_\theta'\, x \tag{1.6.3b}$$

Adding (1.6.3a) and (1.6.3b) gives $e^{i\theta}Ax = (\lambda_\theta - i\lambda_\theta')x$, or

$$Ax = e^{-i\theta}(\lambda_\theta - i\lambda_\theta')x$$

Interpreted as an evaluation at any θ in the indicated interval, this means that

$$\alpha = x^* Ax = e^{-i\theta}(\lambda_\theta - i\lambda_\theta')$$

is an eigenvalue of A. $\square$

Since each $A \in M_n$ has a finite number of eigenvalues, Theorem (1.6.3) implies that if a set is the field of values of a finite matrix, it can have only a finite number of sharp points. Moreover, if the only extreme points of $F(A)$ are sharp points, then $F(A)$ is the convex hull of some eigenvalues of A.

1.6.4 Corollary. Let $A \in M_n$ be given. Then $F(A)$ has at most n sharp points, and $F(A)$ is a convex polygon if and only if $F(A) = \mathrm{Co}(\sigma(A))$.

Although every sharp point on the boundary of $F(A)$ is an eigenvalue, not every eigenvalue on the boundary of $F(A)$ is a sharp point. Nevertheless, every such point does have a special characteristic.

1.6.5 Definition. A point $\lambda \in \sigma(A)$ is a *normal eigenvalue* for the matrix $A \in M_n$ if

(a) Every eigenvector of A corresponding to λ is orthogonal to every eigenvector of A corresponding to each eigenvalue different from λ, and

(b) The geometric multiplicity of the eigenvalue λ (the dimension of the corresponding eigenspace of A) is equal to the algebraic multiplicity of λ (as a root of the characteristic polynomial of A).

Exercise. Show that every eigenvalue of a normal matrix is a normal eigenvalue.

Exercise. If $A \in M_n$ has as many as $n-1$ normal eigenvalues, counting multiplicity, show that A is a normal matrix.

Exercise. Show that λ is a normal eigenvalue of a matrix $A \in M_n$ if and only if it is a normal eigenvalue of UAU^* for every unitary matrix $U \in M_n$, that is, the property of being a normal eigenvalue is a unitary similarity invariant.

Our main observation here is that every eigenvalue on the boundary of the field of values is a normal eigenvalue.

1.6.6 Theorem. If $A \in M_n$ and if $\alpha \in \partial F(A) \cap \sigma(A)$, then α is a normal eigenvalue of A. If m is the multiplicity of α, then A is unitarily similar to $\alpha I \oplus B$, with $I \in M_m$, $B \in M_{n-m}$, and $\alpha \notin \sigma(B)$.

Proof: If the algebraic multiplicity of α is m, then A is unitarily similar to an upper triangular matrix T (this is Schur's theorem; see (2.3.1) in [HJ]) whose first m main diagonal entries are equal to α and whose remaining diagonal entries (all different from α) are the other eigenvalues of A. Suppose there were a nonzero entry off the main diagonal in one of the first m rows of T. Then T would have a 2-by-2 principal submatrix T_0 of the form $T_0 = \begin{bmatrix} \alpha & \gamma \\ 0 & \beta \end{bmatrix}$, $\gamma \neq 0$. Since $F(T_0)$ is either a circular disc about α with radius $\frac{1}{2}|\gamma|$ or a nondegenerate ellipse (with interior) with foci at α and β, the point α must be in the interior of $F(T_0)$. But $F(T_0) \subset F(T) = F(A)$ by (1.2.8) and (1.2.11), which means that α is in the interior of $F(A)$. This contradiction shows that there are no nonzero off-diagonal entries in the first m rows of T, and hence $T = \alpha I \oplus B$ with $I \in M_m$ and $B \in M_{n-m}$. The remaining assertions are easily verified. □

Exercise. Complete the details of the proof of Theorem (1.6.6).

We have already noted that the converse of the normality property (1.2.9) does not hold, but we are now in a position to understand fully the relationship between the two conditions on $A \in M_n$: A is normal, and $F(A) = \mathrm{Co}(\sigma(A))$.

1.6.7 Corollary. If $\sigma(A) \subset \partial F(A)$, then A is normal.

Proof: If all the eigenvalues of A are normal eigenvalues, then there is an orthonormal basis of eigenvectors of A, which must therefore be normal. □

Exercise. Show that a given matrix $A \in M_n$ is normal if and only if A is unitarily similar to a direct sum $A_1 \oplus A_2 \oplus \cdots \oplus A_k$ with $\sigma(A_i) \subset \partial F(A_i)$, $i = 1,\ldots, k$.

Two further corollary facts complete our understanding of the extent to which there is a converse to (1.2.9).

1.6.8 Theorem. Let $A \in M_n$. Then $F(A) = \mathrm{Co}(\sigma(A))$ if and only if *either* A is normal *or* A is unitarily similar to a matrix of the form $\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$, where A_1 is normal and $F(A_2) \subset F(A_1)$.

1.6.9 Corollary. If $A \in M_n$ and $n \leq 4$, then A is normal if and only if

$F(A) = \mathrm{Co}(\sigma(A))$.

Exercise. Supply a proof for (1.6.8) using (1.6.6).
Exercise. Supply a proof for (1.6.9) by considering the geometrical possibilities allowed in (1.6.8).

We wish to discuss one additional connection between normality and the field of values. If $A \in M_{m,n}$ is given and if $B = \begin{bmatrix} A & * \\ * & * \end{bmatrix}$ is a larger matrix that contains A in its upper-left corner, we say that B is a *dilation* of A; by the submatrix inclusion property (1.2.11), $F(A) \subset F(B)$ whenever $B \in M_N$ is a dilation of $A \in M_n$. One can always find a $2n$-by-$2n$ *normal* dilation of A, for example, the matrix B defined by (1.6.11), in which case $F(B) = \mathrm{Co}(\sigma(B))$. It is clear that $F(A)$ is contained in the intersection of the fields of values of all the normal dilations of A, and it is a pleasant observation that this intersection is *exactly* $F(A)$.

1.6.10 Theorem. Let $A \in M_n$ be given. Then

$$F(A) = \bigcap \left\{ F(B) \colon B = \begin{bmatrix} A & * \\ * & * \end{bmatrix} \in M_{2n} \text{ is normal} \right\}$$

Proof: Since we have already noted that $F(A)$ is a subset of the given intersection, we need only prove the reverse inclusion. Because $F(A)$ is closed and convex, it is the intersection of all the closed half-planes that contain it, a fact used in the preceding section to develop a numerical algorithm to compute $F(A)$. Thus, it is sufficient to show that for each closed half-plane that contains $F(A)$, there is some normal matrix $B = \begin{bmatrix} A & * \\ * & * \end{bmatrix} \in M_{2n}$ such that $F(B)$ is contained in the same half-plane. By the translation and scalar multiplication properties (1.2.3-4), there is no loss of generality to assume that the given half-plane is the closed right half-plane RHP_0. Thus, we shall be done if we show that whenever $A + A^*$ is positive semidefinite, that is, $F(A) \subset RHP_0$, then A has a normal dilation $B \in M_{2n}$ such that $B + B^*$ is positive semidefinite, that is, $F(B) \subset RHP_0$. Consider

$$B \equiv \begin{bmatrix} A & A^* \\ A^* & A \end{bmatrix} \tag{1.6.11}$$

Then B is a dilation of A and

$$B^*B = \begin{bmatrix} A^*A + AA^* & A^{*2} + A^2 \\ A^2 + A^{*2} & AA^* + A^*A \end{bmatrix} = BB^*$$

so B is normal, independent of any assumption about $F(A)$. Moreover, if $A + A^*$ is positive semidefinite, then

$$B + B^* = \begin{bmatrix} A + A^* & A + A^* \\ A + A^* & A + A^* \end{bmatrix} \tag{1.6.12}$$

is positive semidefinite, since if $y, z \in \mathbb{C}^n$ and $x \equiv \begin{bmatrix} y \\ z \end{bmatrix} \in \mathbb{C}^{2n}$, then $x^*(B + B^*)x = (y + z)^*(A + A^*)(y + z) \geq 0$. □

Since the field of values of a normal matrix is easily shown to be convex and the intersection of convex sets is convex, Theorem (1.6.10) suggests a clean conceptual proof of the convexity property of the field of values (1.2.2). Unfortunately, the convexity of $F(A)$ is used in a crucial way in the proof of Theorem (1.6.10); thus, it would be very pleasant to have a different proof that does not rely on the Toeplitz-Hausdorff theorem.

The matrix defined by (1.6.11) gives a $2n$-by-$2n$ normal dilation of any given matrix $A \in M_n$. It is sometimes useful to know that one can find dilations with even more structure. For example, one can choose the dilation $B \in M_{2n}$ to be of the form $B = cV$, where $V \in M_{2n}$ is unitary and $c \equiv \max\{|||A|||_2, 1\}$ (see Problem 22). Moreover, for any given integer $k \geq 1$, one can find a dilation $B \in M_{(k+1)n}$ that is a scalar multiple of a unitary matrix and has the property that $B^m = \begin{bmatrix} A^m & * \\ * & * \end{bmatrix}$ for all $m = 1, 2, \ldots, k$ (see Problem 25).

Problems

1. If $A \in M_2$, show that $F(A)$ is a possibly degenerate ellipse with center at $\frac{1}{2}\operatorname{tr} A$ and foci at the two eigenvalues of A. Characterize the case in which $F(A)$ is a circular disc. What is the radius of this disc? Characterize the case in which $F(A)$ is a line segment. What are its endpoints? In the remaining case of an ellipse, what is the eccentricity and what is the equation of the boundary?

2. For $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_2(\mathbb{R})$, show that $F(A)$ is an ellipse (possibly degenerate) with center at the point $\frac{1}{2}(a + d)$ on the real axis, whose vertices are the two points $\frac{1}{2}(a + d \pm [(a - d)^2 + (b + c)^2]^{\frac{1}{2}})$ on the real axis and the two

points $\frac{1}{2}(a+d \pm i|b-c|)$. Show that the degenerate cases correspond exactly to the cases in which A is normal. Can you find similarly explicit formulae if $A \in M_2(\mathbb{C})$?

3. If $A \in M_2$ has distinct eigenvalues, and if $A_0 \equiv A - (\frac{1}{2}\text{tr } A)I$, show that A_0 is nonsingular and the major axis of $F(A)$ lies in the direction $-(\det A_0)/|\det A_0|$ in the complex plane. What is the length of the major axis? What is the length of the minor axis?

4. If $A \in M_2$, and if one of the eigenvalues of A appears on the main diagonal of A, show that at least one of the off-diagonal entries of A must be zero.

5. Show that any given (possibly degenerate) ellipse (with interior) is the field of values of some matrix $A \in M_2$. Show how to construct such a matrix A for a given ellipse.

6. Show by example that 4 is the best possible value in (1.6.9), that is, for every $n \geq 5$ show that there is a nonnormal $A \in M_n$ with $F(A) = \text{Co}(\sigma(A))$.

7. Let $n \geq 3$ be given and let $1 \leq k \leq n$. Show that there is some $A \in M_n$ such that $F(A)$ has exactly k sharp points.

8. If $A \in M_n$ and if α is a sharp point of $F(A)$, show that a unit vector $x \in \mathbb{C}^n$ for which $\alpha = x^*Ax$ is an eigenvector for both A and $\text{H}(A)$.

9. Give an alternate, more geometric proof of (1.6.3) along the following lines: If $A \in M_n$ and $\lambda \in F(A)$, pick a unitary $U \in M_n$ so that λ is the 1,1 entry of U^*AU. (How may U be constructed?) Since

$$\lambda \in F[(U^*AU)(\{1,j\})]$$

it must be a sharp point of $F[(U^*AU)(\{1,j\})]$, $j = 2,\ldots, n$, if λ is a sharp point of $F(A)$. But since $F[(U^*AU)(\{1,j\})]$ is either a circular disc, a line segment, or an ellipse (with interior), it must actually be a line segment with λ as one endpoint in order for λ to be a sharp point. But then $F[(U^*AU)(\{1,j\})]$ must be diagonal, and

$$U^*AU = [\lambda] \oplus (U^*AU)(\{2,\ldots, n\})$$

that is, λ is an eigenvalue of A.

10. Give an alternate proof of (1.6.3) along these lines: Let $x \in \mathbb{C}^n$ be a unit

vector and let $\lambda = x^*Ax$ be a sharp point of $F(A)$. Let $y \in \mathbb{C}^n$ be independent of x. Let $H \in M_n$ denote the (Hermitian) orthogonal projection onto the two-dimensional subspace spanned by x and y, that is, let $\{v, w\}$ be an orthonormal basis of $V \equiv \text{Span}\{x, y\}$, let $P \equiv [v\ w] \in M_{n,2}$ be an isometric projection (see Problem 16 in Section (1.2)), and let $H \equiv PP^*$, so $Hz = z$ if $z \in V$ and $Hz = 0$ if $z \perp V$. Then λ is a sharp point of $F(HAH) \subset F(A)$. Since $F(HAH)$ is a (possibly degenerate) ellipse, it must be a line segment with λ as an endpoint, so λ is an eigenvalue of HAH and of A, with eigenvector x.

11. If $A \in M_n$ and $\lambda \in \sigma(A)$, and if $x \in \mathbb{C}^n$ is an associated eigenvector, show that λ is a normal eigenvalue of A if and only if x is an eigenvector of both A and A^*. What can be said if the multiplicity of λ is greater than one?

12. If $\lambda, \mu \in \sigma(A)$ and if $\mu \neq \lambda$, show that any eigenvector of A^* associated with $\bar{\mu}$ is orthogonal to any eigenvector of A associated with λ.

13. Let $A \in M_n$. Show that if x is an eigenvector of both A and $\text{H}(A)$, then x is an eigenvector of A^*. How are the eigenvalues related?

14. Give another proof of (1.6.6) using (1.5.7) and the result of Problem 13.

15. Let $A \in M_n$. If $F(A)$ is a polygon with at least $n-1$ vertices, show that A is normal.

16. Let $A \in M_n$. If at least $n-1$ of the eigenvalues of A lie on the boundary of $F(A)$, show that A is normal. Show that the value $n-1$ is the best possible.

17. Let $A \in M_n$. (a) If $F(A)$ has finitely many extreme points, show that they are all eigenvalues of A. (b) If A is normal, show that the extreme points of $F(A)$ are eigenvalues of A.

18. Show that $A \in M_n$ is normal if and only if A is unitarily similar to a direct sum of square matrices $A_1 \oplus \cdots \oplus A_k$, $A_i \in M_{n_i}$, $i = 1,\ldots, k$, and each $F(A_i)$ is a polygon (perhaps a line segment or a point) with n_i vertices, $i = 1,\ldots, k$.

19. Let $A \in M_n$ be nonsingular and normal, with $\sigma(A) = \{r_1 e^{i\theta_1},\ldots, r_n e^{i\theta_n}\}$, all $r_j > 0$, and all $\theta_j \in [0, 2\pi)$. If A has some diagonal entry equal to 0, show that there does *not* exist a value t such that $t \leq \theta_i < t + \pi$ for all $i = 1, 2,\ldots, n$, that is, the eigenvalues do not project onto the unit circle inside any semicircle.

20. Give another proof that the matrix B defined by (1.6.11) is normal, and that $B + B^*$ is positive semidefinite if $A + A^*$ is positive semidefinite, as follows: Show that the $2n$-by-$2n$ matrix $U \equiv \begin{bmatrix} I & I \\ -I & I \end{bmatrix} / \sqrt{2}$ is unitary, $U^* B U = \begin{bmatrix} A - A^* & 0 \\ 0 & A + A^* \end{bmatrix}$, and $U^*(B + B^*)U = \begin{bmatrix} 0 & 0 \\ 0 & 2(A + A^*) \end{bmatrix}$. Thus, B and $B + B^*$ are unitarily similar to something obviously normal and something positive semidefinite, respectively. How is the spectrum of B determined by A? Show that $|||B|||_2 \le 2\,|||A|||_2$.

21. Recall that a matrix A is a *contraction* if $|||A|||_2 \le 1$. See Problem 30 in Section (1.5) for an important link between contractions and the field of values.

(a) If there exists a unitary dilation $V = \begin{bmatrix} A & * \\ * & * \end{bmatrix}$ of a given square complex matrix A, show that A must be a contraction.

(b) Conversely, if $A \in M_n$ is a given contraction, show that A has a unitary dilation as follows: Let $A = PU$ be a polar decomposition of A, where $U \in M_n$ is unitary and $P \equiv (AA^*)^{\frac{1}{2}}$ is positive semidefinite. Let

$$Z \equiv \begin{bmatrix} A & (I - P^2)^{\frac{1}{2}} U \\ -(I - P^2)^{\frac{1}{2}} U & A \end{bmatrix} \in M_{2n} \tag{1.6.13}$$

and show that Z is unitary.

(c) If $A \in M_n$ is a contraction, show that

$$Z \equiv \begin{bmatrix} A & (I - AA^*)^{\frac{1}{2}} \\ -(I - A^*A)^{\frac{1}{2}} & A^* \end{bmatrix} \in M_{2n} \tag{1.6.14}$$

is also a unitary dilation of A.

(d) Use the dilation Z in (1.6.14) and the argument in the proof of Theorem (1.6.10) to show that if A is a *normal* contraction, then

$$F(A) = \bigcap \left\{ \{F(U) \colon U = \begin{bmatrix} A & * \\ * & * \end{bmatrix} \in M_{2n} \text{ is unitary} \right\}$$

(e) Combine the normal dilation (1.6.11) with the result in (d) and the norm bound in Problem 20 to show that if $A \in M_n$ is a contraction with $|||A|||_2 \le \frac{1}{2}$, then

$$F(A) = \bigcap \left\{ F(U) \colon U = \begin{bmatrix} A & * \\ * & * \end{bmatrix} \in M_{4n} \text{ is unitary} \right\}$$

(f) Now consider a third unitary dilation, one that does not require the polar decomposition, the singular value decomposition, or matrix square roots. The following construction can result in a dilation of smaller size than $2n$; an analogous construction gives a complex orthogonal dilation of an arbitrary matrix (see Problem 40). Let $A \in M_n$ be a given contraction and let $\delta \equiv \operatorname{rank}(I - A^*A)$, so $0 \leq \delta \leq n$ with $\delta = 0$ if and only if A is unitary; δ may be thought of as a measure of how far A is from being unitary. Explain why $\operatorname{rank}(I - AA^*) = \delta$ and why there are nonsingular $X, Y \in M_n$ with

$$I - AA^* = X \begin{bmatrix} I_\delta & 0 \\ 0 & 0 \end{bmatrix} X^* \text{ and } I - A^*A = Y^* \begin{bmatrix} 0 & 0 \\ 0 & I_\delta \end{bmatrix} Y \qquad (*)$$

and $I_\delta \in M_\delta$. Define

$$B \equiv X \begin{bmatrix} I_\delta \\ 0 \end{bmatrix} \in M_{n,\delta},\ \ C \equiv -[0 \ \ I_\delta]\, Y \in M_{\delta,n}, \text{ and}$$

$$D \equiv [0 \ \ I_\delta]\ YA^*(X^*)^{-1} \begin{bmatrix} I_\delta \\ 0 \end{bmatrix} \in M_\delta$$

and form

$$Z \equiv \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in M_{n+\delta} \qquad (1.6.15)$$

Use the definitions $(*)$ and the identity $A(I - A^*A) = (I - AA^*)A$ to show that

$$X^{-1}AY^* \begin{bmatrix} 0 & 0 \\ 0 & I_\delta \end{bmatrix} = \begin{bmatrix} I_\delta & 0 \\ 0 & 0 \end{bmatrix} X^*AY^{-1} \qquad (**)$$

What do (1.6.15) and $(**)$ become when $\delta = n$, that is, when A is a strict contraction ($|||A|||_2 < 1$)? In this special case, verify that (1.6.15) gives a unitary matrix. Now consider the general case when $\delta < n$; considerable algebraic manipulation seems to be needed to show that (1.6.15) is unitary.

(g) Let n and k be given positive integers. Show that a given $A \in M_k$ is a principal submatrix of a unitary $U \in M_n$ if and only if A is a contraction and $\operatorname{rank}(I - A^*A) \leq \min\{k, n-k\}$; the latter condition imposes no restriction on A if $k \leq n/2$.

(h) Show that a given $A \in M_{m,n}$ has a unitary dilation if and only if A is a contraction.

22. Let $A \in M_n$ be given. Show that there exists a unitary $V \in M_{2n}$ such that $cV = \begin{bmatrix} A & * \\ * & * \end{bmatrix}$ with $c \equiv \max\{|||A|||_2, 1\}$. Compare with Theorem (1.6.10).

23. Consider

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \in M_2 \text{ and } B_z = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ z & 0 & 0 \end{bmatrix} \in M_3$$

Show that $F(A) = \cap\{F(B_z): z \in \mathbf{C} \text{ and } |z| = 1\}$. Discuss this in light of Problem 21.

24. The dilation results in Problem 21 have a very interesting generalization. Let

$$Z \equiv \begin{bmatrix} A & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$$

denote any unitary dilation of a given contraction $A \in M_n$, for example, Z could be any of the dilations given by (1.6.13,14,15), and let $k \geq 1$ be a given integer. Consider the block matrix $V = [V_{ij}]_{i,j=1}^{k+1}$, in which each block V_{ij} is defined by: $V_{11} \equiv A$, $V_{1,2} \equiv Z_{12}$, $V_{k+1,1} \equiv Z_{21}$, $V_{k+1,2} \equiv Z_{22}$, and $V_{2,3} = V_{3,4} = \cdots = V_{k,k+1} \equiv I$, where this identity matrix is the same size as Z_{22}, and all other blocks are zero:

$$V \equiv \begin{bmatrix} A & Z_{12} & 0 & \cdots\cdots & 0 \\ 0 & 0 & I & 0 & \vdots \\ \vdots & \vdots & & I \ \ddots & 0 \\ 0 & 0 & 0 & \ddots & I \\ Z_{21} & Z_{22} & 0 & \cdots\cdots & 0 \end{bmatrix} \tag{1.6.16}$$

(a) When $k = 1$, show that $V = Z$.

(b) For $k \geq 2$, show that V is unitary and that $V^m = \begin{bmatrix} A^m & * \\ * & * \end{bmatrix}$ for $m = 1, \ldots, k$, that is, each of these powers of V is a unitary dilation of the corresponding power of A.

25. Let $A \in M_n$ and let $k \geq 1$ be a given integer. Show that there exists a unitary $V \in M_{(k+1)n}$ such that $(cV)^m = \begin{bmatrix} A^m & * \\ * & * \end{bmatrix}$ for $m = 1, 2, \ldots, k$, where $c \equiv$

max $\{|||A|||_2, 1\}$. Compare with Theorem (1.6.10).

26. Let $A \in M_n$ be a given contraction, that is, $|||A|||_2 \leq 1$. If $p(\cdot)$ is a polynomial such that $|p(z)| \leq 1$ for all $|z| = 1$, use Problem 25 to show that $p(A)$ is a contraction.

27. Use Theorem (1.6.6) to show that the algebraic and geometric multiplicities of the Perron root (and any other eigenvalues of maximum modulus) of a doubly stochastic matrix must be equal, and show by example that this is not true of a general nonnegative matrix. Show, however, that this *is* true of a nonnegative matrix that is *either* row or column stochastic, and give an example of such a matrix whose Perron root does not lie on the boundary of its field of values.

28. Use Corollary (1.6.7) to give a shorter proof of Theorem (1.6.1). Is there any circular reasoning involved in doing so?

29. A matrix $A \in M_n$ is said to be *unitarily reducible* if there is some unitary $U \in M_n$ such that $U^*AU = A_1 \oplus A_2$ with $A_1 \in M_k$, $A_2 \in M_{n-k}$, and $1 < k < n$; A is *unitarily irreducible* if it is not unitarily reducible. If $A \in M_n$ is unitarily irreducible, show that every eigenvalue of A lies in the (two-dimensional) interior of $F(A)$ and that the boundary of $F(A)$ is a smooth curve (no sharp points).

30. Let $A \in M_n$ be given. Show that $\mathrm{Co}(\sigma(A)) = \cap\{F(SAS^{-1})$: $S \in M_n$ is nonsingular$\}$. Moreover, if $J_{n_1}(\lambda_1) \oplus \cdots \oplus J_{n_k}(\lambda_k)$ is the Jordan canonical form of A (see (3.1.11) in [HJ]), show that there is some one nonsingular $S \in M_n$ such that $\mathrm{Co}(\sigma(A)) = F(SAS^{-1})$ if and only if $n_i = 1$ for every eigenvalue λ_i that lies on the boundary of $\mathrm{Co}(\sigma(A))$.

31. If $<\cdot,\cdot>$ is a given inner product on $\mathbb{C}^n$, the *field of values generated by* $<\cdot,\cdot>$ is defined by $F_{<\cdot,\cdot>}(A) \equiv \{<Ax,x>: x \in \mathbb{C}^n, <x,x> = 1\}$ for $A \in M_n$. Note that the usual field of values is the field of values generated by the usual Euclidean inner product $<x,y> \equiv y^*x$. Show that $\mathrm{Co}(\sigma(A)) = \cap\{F_{<\cdot,\cdot>}(A): <\cdot,\cdot>$ is an inner product on $\mathbb{C}^n\}$.

32. Let $A \in M_n$ be given. In general, $\rho(A) \leq r(A)$, but if $\rho(A) = r(A)$, we say that A is *spectral* (the term *spectraloid* is also used). In particular, A is spectral if it is radial, that is, if either $|||A|||_2 = \rho(A)$ or $|||A|||_2 = r(A)$; see Problems 23, 24, and 27 in Section (1.5). If A is spectral and λ is any eigenvalue of A such that $|\lambda| = \rho(A)$, show that λ is a normal eigenvalue of A. Show that every normal matrix $A \in M_n$ is spectral and that the converse is

true only for $n = 2$.

33. Use Problem 32 to show that a given $A \in M_n$ is spectral if and only if it is unitarily similar to a direct sum of the form $\lambda_1 I_{n_1} \oplus \cdots \oplus \lambda_k I_{n_k} \oplus B$, where $1 \leq k \leq n$, $n_1 + \cdots + n_k = n - m$, $0 \leq m \leq n-1$, $B \in M_m$, $|\lambda_1| = \cdots = |\lambda_k| = \rho(A)$, $\rho(B) < \rho(A)$, and $r(B) \leq r(A)$.

34. Use Problem 33 to show that a given $A \in M_n$ is spectral if and only if it is unitarily similar to $r(A)(U \oplus B)$, where $U \in M_k$ is unitary, $1 \leq k \leq n$, and $B \in M_{n-k}$ has $\rho(B) < \rho(A)$ and $r(B) \leq r(A)$.

35. Show that a given $A \in M_n$ is spectral if and only if $r(A^k) = r(A)^k$ for all $k = 1, 2, \ldots$. Conclude that if A is spectral, then every positive integer power of A is spectral. For each $n \geq 2$, exhibit a nonspectral matrix $B \in M_n$ such that B^m is spectral for all $m \geq 2$.

36. Let $A \in M_n$ be given with $r(A) \leq 1$. Show that if $\rho(A) = 1$, then $r(A^m) = 1$ for all $m = 1, 2, \ldots$. Goldberg, Tadmor, and Zwas have proved a converse: If $r(A) \leq 1$, then $\rho(A) = 1$ if and only if $r(A^m) = 1$ for some positive integer m not less than the degree of the minimal polynomial of A. Use this result to establish the following:

(a) A is spectral if and only if $r(A^m) = r(A)^m$ for some positive integer m not less than the degree of the minimal polynomial of A.

(b) Show that the 3-by-3 nilpotent Jordan block

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

has $\rho(A) = 0$ and $r(A^2) = r(A)^2 = \frac{1}{2}$, but A is not spectral. Explain why the bound on the exponent in the Goldberg-Tadmor-Zwas theorem is sharp.

(c) Show that A is spectral if and only if $r(A^n) = r(A)^n$.

(d) When $n = 2$, show that A is normal if and only if $r(A^2) = r(A)^2$.

Compare the Goldberg-Tadmor-Zwas theorem with Ptak's theorem on radial matrices and its generalization in Problems 25 and 26 in Section (1.5).

37. Let $A \in M_n$ be given. Combine the results of the preceding five problems and verify that the following conditions are equivalent:

(a) A is spectral.

(b) $\rho(A) = r(A)$.
(c) $r(A^n) = r(A)^n$.
(d) $r(A^m) = r(A)^m$ for some integer m not less than the degree of the minimal polynomial of A.
(e) $r(A^k) = r(A)^k$ for all $k = 1, 2, \ldots$.
(f) A is unitarily similar to $r(A)(U \oplus B)$, where $U \in M_k$ is unitary, $1 \leq k \leq n$, and $B \in M_{n-k}$ has $\rho(B) < 1$ and $r(B) \leq 1$.

Compare with the conditions for a matrix to be *radial* given in Problem 27 in Section (1.5).

38. Use Problem 37(f) to prove the assertion in Problem 27(g) in Section (1.5).

39. The main geometrical content of Theorem (1.6.6) is the following fact: Any eigenvector of A corresponding to an eigenvalue that is a boundary point of the field of values is orthogonal to any eigenvector of A corresponding to any other (different) eigenvalue. Provide details for the following alternative proof of this fact. We write $<x,y>$ for the usual inner product. (1) Let $\lambda \in \partial F(A)$, and let $\mu \neq \lambda$ be any other eigenvalue of A. Let x, y be unit vectors with $Ax = \lambda x$ and $Ay = \mu y$. Since $F(A)$ is compact and convex, and λ is a boundary point of $F(A)$, there is a supporting line for $F(A)$ that passes through λ. Show that this geometrical statement is equivalent to the existence of a nonzero $c \in \mathbb{C}$ and a real number a such that $\operatorname{Im} c<Az,z> \geq a$ for all unit vectors z, with equality for $z = x$, that is, $\operatorname{Im} c\lambda = a$. (2) Let $A_1 \equiv cA - iaI$, $\lambda_1 \equiv c\lambda - ia$, and $\mu_1 \equiv c\mu - ia$. Show that $\operatorname{Im} <A_1z,z> \geq 0$ for all unit vectors z, λ_1 is real, $\bar{\mu}_1 \neq \lambda_1$, and $\operatorname{Im} \mu_1 \neq 0$. (3) For arbitrary $\xi, \eta \in \mathbb{C}$, note that $\operatorname{Im} <A_1(\xi x + \eta y),(\xi x + \eta y)> \geq 0$ and deduce that $\operatorname{Im} [|\eta|^2\mu_1 + (\lambda_1 - \bar{\mu}_1)\xi\bar{\eta}<x,y>] \geq 0$. (4) Show that you may choose η so that $|\eta| = 1$ and $\operatorname{Im} [(\lambda_1 - \bar{\mu}_1)\bar{\eta}<x,y>] = |(\lambda_1 - \bar{\mu}_1)<x,y>|$; deduce that $\operatorname{Im} \mu_1 + |(\lambda_1 - \bar{\mu}_1)<x,y>|\xi \geq 0$ for all $\xi \in \mathbb{R}$. (5) Conclude that $<x,y> = 0$, as desired. *Remark:* The proof we have outlined is of interest because it does not rely on an assumption that A is a (finite-dimensional) matrix. It is valid for any bounded linear operator A on a complex Hilbert space with inner product $<\cdot,\cdot>$, and uses the fact that $F(A)$ is a compact convex set. Thus, the geometrical statement about orthogonality of eigenvectors given at the beginning of this problem is valid for any bounded linear operator A on a complex Hilbert space.

40. Although a given matrix has a unitary dilation if and only if it is a contraction (Problem 21), *every complex matrix has a complex orthogonal* *dilation.*

(a) Let $A \in M_n$ and let $\delta \equiv \text{rank}(I - A^TA)$, so $0 \leq \delta \leq n$ with $\delta = 0$ if and only if A is complex orthogonal; δ may be thought of as a measure of how far A is from being complex orthogonal. Explain why $\text{rank}(I - AA^T) = \delta$ and use Corollary (4.4.4) in [HJ] to show that there are nonsingular $X, Y \in M_n$ with

$$I - AA^T = X\begin{bmatrix} I_\delta & 0 \\ 0 & 0 \end{bmatrix} X^T \text{ and } I - A^TA = Y^T\begin{bmatrix} 0 & 0 \\ 0 & I_\delta \end{bmatrix} Y \qquad ({}^T)$$

and $I_\delta \in M_\delta$. Define

$$B \equiv X\begin{bmatrix} I_\delta \\ 0 \end{bmatrix} \in M_{n,\delta},\ C \equiv -[0\ \ I_\delta]\, Y \in M_{\delta,n}, \text{ and}$$

$$D \equiv [0\ \ I_\delta]\ YA^T(X^T)^{-1}\begin{bmatrix} I_\delta \\ 0 \end{bmatrix} \in M_\delta$$

and form

$$Q \equiv \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in M_{n+\delta} \qquad (1.6.17)$$

Use the definitions $({}^T)$ and the identity $A(I - A^TA) = (I - AA^T)A$ to show that

$$X^{-1}AY^T\begin{bmatrix} 0 & 0 \\ 0 & I_\delta \end{bmatrix} = \begin{bmatrix} I_\delta & 0 \\ 0 & 0 \end{bmatrix} X^TAY^{-1} \qquad ({}^{TT})$$

Notice that these calculations carry over exactly to (or from) (**) in Problem 21; one just interchanges * and ^{T}. What do (1.6.17) and $({}^{TT})$ become when $\delta = n$? In this special case, verify that (1.6.17) gives a complex orthogonal matrix. The algebraic manipulations necessary to establish the general case are analogous to those needed to establish the general case in Problem 21.

(b) Assuming that (1.6.17) gives a complex orthogonal dilation of A, explain why the construction of (1.6.16) in Problem 24 gives a complex orthogonal dilation $P = \begin{bmatrix} A & * \\ * & * \end{bmatrix} \in M_{n+k\delta}$ such that $P^m = \begin{bmatrix} A^m & * \\ * & * \end{bmatrix}$ for $m = 1,\ldots, k$.

(c) Let n and k be given positive integers. Show that a given $A \in M_k$ is a

principal submatrix of a complex orthogonal $Q \in M_n$ if and only if $\operatorname{rank}(I - A^T A) \le \min\{k, n-k\}$, which imposes no restriction on A if $k \le n/2$.
(d) Show that every $A \in M_{m,n}$ has a complex orthogonal dilation.

Notes and Further Readings. The proof given for Theorem (1.6.1) is due to Onn Chan. Theorem (1.6.3) may be found in Section I of R. Kippenhahn, Über den Wertevorrat einer Matrix, *Math. Nachr.* 6 (1951), 193-228, where a proof different from the three given here may be found (see Problems 9 and 10), as well as many additional geometric results about the field of values; however, see the comments at the end of Section (1.8) about some of the quaternion results in Section II of Kippenhahn's paper. Theorems (1.6.6) and (1.6.8) are from C. R. Johnson, Normality and the Numerical Range, *Linear Algebra Appl.* 15 (1976), 89-94, where additional related references may be found. A version of (1.6.1) and related ideas are discussed in O. Taussky, Matrices with Trace Zero, *Amer. Math. Monthly* 69 (1962), 40-42. Theorem (1.6.10) and its proof are in P. R. Halmos, Numerical Ranges and Normal Dilations, *Acta Sci. Math. (Szeged)* 25 (1964), 1-5. The constructions of unitary dilations of a given contraction given in (1.6.13) and (1.6.16) are in E. Egervary, On the Contractive Linear Transformations of n-Dimensional Vector Space, *Acta Sci. Math. (Szeged)* 15 (1953), 178-182. The constructions for (1.6.15-17) as well as sharp bounds on the sizes of unitary and complex orthogonal dilations with the power properties discussed in Problems 24 and 40 are given in R. C. Thompson and C.-C. T. Kuo, Doubly Stochastic, Unitary, Unimodular, and Complex Orthogonal Power Embeddings, *Acta. Sci. Math. (Szeged)* 44 (1982), 345-357. Generalizations and extensions of the result in Problem 26 are in K. Fan, Applications and Sharpened Forms of an Inequality of Von Neumann, pp. 113-121 of [UhGr]; $p(z)$ need not be a polynomial, but may be any analytic function on an open set containing the closed unit disc, and there are matrix analytic versions of classical results such as the Schwarz lemma, Pick's theorem, and the Koebe $\frac{1}{4}$-theorem for univalent functions. The theorem referred to in Problem 36 is proved in M. Goldberg, E. Tadmor, and G. Zwas, The Numerical Radius and Spectral Matrices, *Linear Multilinear Algebra* 2 (1975), 317-326. The argument in Problem 39 is due to Ky Fan, A Remark on Orthogonality of Eigenvectors, *Linear Multilinear Algebra* 23 (1988), 283-284.

1.7 Products of matrices

We survey here a broad range of facts relating products of matrices to the field of values. These fall into four categories:

(a) Examples of the failure of submultiplicativity of $F(\cdot)$;
(b) Results about the usual product when zero is not in the field of values of one of the factors;
(c) Discussion of simultaneous diagonalization by congruence;
(d) A brief survey of the field of values of a Hadamard product.

Examples of the failure of submultiplicativity for $F(\cdot)$

Unfortunately, even very weak multiplicative analogs of the subadditivity property (1.2.7) do not hold. We begin by noting several examples that illustrate what is *not* true; these place in proper perspective the few facts that *are* known about products.

The containment

$$F(AB) \subset F(A)F(B), \qquad A, B \in M_n \tag{1.7.1}$$

fails to hold both "angularly" and "magnitudinally."

1.7.2 Example. Let $A = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix}$. Then $F(A) = F(B) = F(A)F(B) =$ the unit disc, while $F(AB)$ is the line segment joining 0 and 4. Thus, $F(AB)$ contains points much further (four times as far) from the origin than $F(A)F(B)$, so this example shows that the numerical radius $r(\cdot)$ is not a matrix norm. However, $4r(\cdot)$ is a matrix norm (see Problem 22 in Section (5.7) of [HJ]).

1.7.3 Example. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then $F(A) = F(B) = F(A)F(B) =$ the line segment joining -1 and 1, while $F(AB)$ is the line segment joining $-i$ and i. Also, $F'(A) = F'(B) = F'(A)F'(B) =$ the real line, while $F'(AB) =$ the imaginary axis. Note that $r(AB) \leq r(A)r(B)$ in this case, but that (1.7.1) still fails—for angular reasons; $F'(AB)$ is not contained in $F'(A)F'(B)$ either.

1.7.4 Example. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$. Then $F(A)$ is the line segment joining

1 and i, while $F(A^2)$ is the line segment joining -1 and 1. Thus, $F(A^2)$ is not even contained in $F(A)^2$, since $0 \in F(A^2)$ and $0 \notin F(A)$. However, $\mathrm{Co}(F(A)^2) = \mathrm{Co}\{-1, i, 1\}$, so $F(A^2) \subset \mathrm{Co}(F(A)^2)$.

One reason it is not surprising that (1.7.1) does not hold in general is that $F(A)F(B)$ is not generally a convex set, whereas $F(AB)$ is always convex. But, since $F(A)F(B)$ is convex in the case of both examples (1.7.2) and (1.7.3), the inclusion

$$F(AB) \subset \mathrm{Co}(F(A)F(B)), \quad A, B \in M_n \tag{1.7.1a}$$

also fails to hold in general. Examples (1.7.2) and (1.7.3) also show that another weaker statement

$$\sigma(AB) \subset F(A)F(B) \tag{1.7.1b}$$

fails to hold, too.

Product results when zero is not in the field of values of a factor

Now that we know a few things that we can *not* prove, we turn to examining some of what *can* be said about the field of values and products of matrices. The importance of the condition that zero not be in the field of values of one of the factors should be noted at the outset. This means, for example, that the field of values may be rotated into the right half-plane, an observation that links many of the results presented.

1.7.5 Observation. If $A \in M_n$ is nonsingular, then $F'(A^{-1}) = \overline{F'(A)}$.

Proof: Employing the congruence property (1.2.12), we have $F'(A^{-1}) = F'(AA^{-1}A^*) = F'(A^*) = \overline{F'(A)}$. □

Exercise. Verify that $F(A^*) = \overline{F(A)}$ and, therefore, that $F'(A^*) = \overline{F'(A)}$.

Although (1.7.1b) fails to hold, the following four facts indicate that certain limited statements can be made about the relationship between the spectrum of a product and the product of the fields of values.

1.7.6 Theorem. Let $A, B \in M_n$ and assume that $0 \notin F(B)$. Then $\sigma(AB^{-1}) \subset F(A)/F(B)$.

Proof: Since $0 \notin F(B)$, we know from the spectral containment property (1.2.6) that B is nonsingular. The set ratio $F(A)/F(B)$ (which has the usual algebraic interpretation) makes sense and is bounded. If $\lambda \in \sigma(AB^{-1})$, then there is a unit vector $x \in \mathbf{C}^n$ such that $x^*AB^{-1} = \lambda x^*$. Then $x^*A = \lambda x^*B$, $x^*Ax = \lambda x^*Bx$, and $x^*Bx \neq 0$; hence $\lambda = x^*Ax/x^*Bx \in F(A)/F(B)$. $\square$

1.7.7 Corollary. Let $A, B \in M_n$. If B is positive semidefinite, then $\sigma(AB) \subset F(A)F(B)$.

Proof: First suppose that B is nonsingular and set $B = C^{-1}$. If $\lambda \in \sigma(AB) = \sigma(AC^{-1})$, then by (1.7.6) we must have $\lambda = a/c$ for some $a \in F(A)$ and some $c \in F(C)$. If β_{min} and β_{max} are the smallest and largest eigenvalues of B, then β_{min}^{-1} and β_{max}^{-1} are the largest and smallest eigenvalues of the positive definite matrix C, and $F(C)$ is the interval $[\beta_{max}^{-1}, \beta_{min}^{-1}]$. Therefore, $c^{-1} \in [\beta_{min}, \beta_{max}] = F(B)$ and $\lambda = ac^{-1} \in F(A)F(B)$. The case in which B is singular may be reduced to the nonsingular case by replacing B with $B + \epsilon I$, $\epsilon > 0$, and then letting $\epsilon \to 0$. $\square$

1.7.8 Theorem. Let $A, B \in M_n$ and assume that $0 \notin F(B)$. Then $\sigma(AB) \subset F'(A)F'(B)$.

Proof: Since $0 \notin F(B)$, B is nonsingular and we may write $B = C^{-1}$. By (1.7.6), for each $\lambda \in \sigma(AB) = \sigma(AC^{-1})$ we have $\lambda = a/c$ for some $a \in F(A) \subset F'(A)$. For some unit vector $x \in \mathbf{C}^n$ we have $c = x^*Cx = (Cx)^*B^*(Cx) = y^*B^*y \in F'(B^*) = \overline{F'(B)}$, where we have set $y \equiv Cx$. Then $\bar{c} \in F'(B)$, $c \neq 0$, and $c^{-1} = \bar{c}/|c|^2 \in F'(B)$. Thus, $\lambda = ac^{-1} \in F'(A)F'(B)$. $\square$

1.7.9 Theorem. Let $A \in M_n$ and $\lambda \in \mathbf{C}$ be given with $\lambda \neq 0$. The following are equivalent:

(a) $\lambda \in F'(A)$;
(b) $\lambda \in \sigma(HA)$ for some positive definite $H \in M_n$; and
(c) $\lambda \in \sigma(C^*AC)$ for some nonsingular $C \in M_n$.

Proof: Since H is positive definite if and only if $H = CC^*$ for some nonsingular $C \in M_n$, and since $\sigma(CC^*A) = \sigma(C^*AC)$, it is clear that (b) and (c) are equivalent. To show that (c) implies (a), suppose that $\lambda \in \sigma(C^*AC)$ with an associated unit eigenvector y. Then $\lambda = y^*C^*ACy = (Cy)^*A(Cy) = x^*Ax$ for $x \equiv Cy \neq 0$, and hence $\lambda \in F'(A)$. Conversely, suppose $\lambda = x^*Ax$. Then $x \neq 0$ since $\lambda \neq 0$, and we let $C_1 \in M_n$ be any nonsingular matrix whose first column is x. The 1,1 entry of $C_1^*AC_1$ is λ. Let v^T be the row vector formed by the remaining $n-1$ entries of the first row of $C_1^*AC_1$ and let $z^T = -v^T/\lambda$. Then $C_2 = \begin{bmatrix} 1 & z^T \\ 0 & I \end{bmatrix} \in M_n$ is nonsingular and $C_2^*(C_1^*AC_1)C_2 = \begin{bmatrix} \lambda & 0 \\ * & * \end{bmatrix}$. Thus, $\lambda \in \sigma(C^*AC)$, where $C = C_1C_2$ is nonsingular. □

It is now possible to characterize the so-called *H-stable matrices*, that is, those $A \in M_n$ such that all eigenvalues of HA have positive real part for all positive definite matrices H.

1.7.10 **Corollary**. Let $A \in M_n$. Then $\sigma(HA) \subset RHP$ for all positive definite $H \in M_n$ if and only if $F(A) \subset RHP \cup \{0\}$ and A is nonsingular.

Exercise. Prove (1.7.10) using (1.7.9).

We have already seen the special role played by matrices for which zero is exterior to the field of values. Another example of their special nature is the following result, which amounts to a characterization.

1.7.11 **Theorem**. Let $A \in M_n$ be nonsingular. The following are equivalent:

(a) $A^{-1}A^* = B^{-1}B^*$ for some $B \in M_n$ with $0 \notin F(B)$.

(b) A is *congruent to a normal matrix.

(b′) A is *congruent to a normal matrix via a positive definite congruence.

(c) $A^{-1}A^*$ is similar to a unitary matrix.

Proof: Since $A^{-1}A^*$ is unitary if and only if $(A^{-1}A^*)^{-1} = (A^{-1}A^*)^*$, which holds if and only if $A^*A = AA^*$, we see that $A^{-1}A^*$ is unitary if and only if A is normal. The equivalence of (b) and (c) then follows directly from the calculation $S^{-1}A^{-1}A^*S = S^{-1}A^{-1}(S^*)^{-1}S^*A^*S = (S^*AS)^{-1}(S^*AS)^*$; if $S^{-1}A^{-1}A^*S$ is unitary, then S^*AS (a *congruence of A) must be normal, and conversely.

To verify that (c) implies (a), suppose that $U \equiv S^{-1}A^{-1}A^*S$ is unitary, and let $U^{\frac{1}{2}}$ be one of the (several) unitary square roots of U such that all eigenvalues of $U^{\frac{1}{2}}$ lie on an arc of the unit circle of length less than π. By (1.2.9), $0 \notin F(U^{\frac{1}{2}})$ and by (1.2.12) and (1.7.5), $0 \notin F((S^{-1})^* U^{-\frac{1}{2}} S^{-1})$, where $U^{-\frac{1}{2}} \equiv (U^{\frac{1}{2}})^{-1}$. Now calculate $A^{-1}A^* = SUS^{-1} = S(U^{-\frac{1}{2}})^{-1}(U^{-\frac{1}{2}})^* S^{-1} = S(U^{-\frac{1}{2}})^{-1} S^* (S^{-1})^* (U^{-\frac{1}{2}})^* S^{-1} = B^{-1}B^*$, with $B \equiv (S^{-1})^* U^{-\frac{1}{2}} S^{-1}$ and $0 \notin F(B)$ as (a) asserts. Assuming (a), we may suppose $0 \notin F(A)$ and then we may suppose, without loss of generality, that $\mathrm{H}(A)$ is positive definite by (1.3.5). We now show that (a) implies (b′) by writing $A = H + S$, where $H \equiv \mathrm{H}(A) = \frac{1}{2}(A + A^*)$ is positive definite and $S \equiv \mathrm{S}(A) = \frac{1}{2}(A - A^*)$. If $H^{-\frac{1}{2}}$ is the inverse of the unique positive definite square root of H, we have $(H^{-\frac{1}{2}})^* A H^{-\frac{1}{2}} = H^{-\frac{1}{2}}(H + S)H^{-\frac{1}{2}} = I + H^{-\frac{1}{2}} S H^{-\frac{1}{2}}$, which is easily verified to be normal. Thus, A is congruent to a normal matrix via the positive definite congruence $\mathrm{H}(A)^{-\frac{1}{2}}$. Finally, (b′) trivially implies (b). □

We next relate angular information about the spectrum of A to the angular field of values of positive definite multiples of A. Compare the next result to (1.7.9).

1.7.12 Theorem. Let Γ be an open angular sector of $\mathbf{C}$ anchored at the origin, with angle not greater than π, and let $A \in M_n$. The following are equivalent:

(a) $\sigma(A) \subset \Gamma$
(b) $F'(HA) \subset \Gamma$ for some positive definite $H \in M_n$

Proof: By (3.1.13) in [HJ] there is for each $\epsilon > 0$ a nonsingular $S_\epsilon \in M_n$ such that $S_\epsilon^{-1} A S_\epsilon$ is in modified Jordan canonical form: In place of every off-diagonal 1 that occurs in the Jordan canonical form of A is ϵ. Then, for sufficiently small ϵ, $F(S_\epsilon A S_\epsilon^{-1}) \subset \Gamma$ if $\sigma(A) \subset \Gamma$. But, by (1.2.12), we have $F'(S_\epsilon^* S_\epsilon A S_\epsilon^{-1} S_\epsilon) = F'(S_\epsilon A S_\epsilon^{-1}) \subset \Gamma$. Letting $H = S_\epsilon^* S_\epsilon$ demonstrates that (a) implies (b).

The proof that (b) implies (a) is similar. If $F'(HA) \subset \Gamma$, then $F'(H^{\frac{1}{2}} A H^{-\frac{1}{2}}) = F'(H^{-\frac{1}{2}} HAH^{-\frac{1}{2}}) \subset \Gamma$, using (1.2.12) again, where $H^{\frac{1}{2}}$ is the positive definite square root of H and $H^{-\frac{1}{2}}$ is its inverse. But $\sigma(A) = \sigma(H^{\frac{1}{2}} A H^{-\frac{1}{2}}) \subset F(H^{\frac{1}{2}} A H^{-\frac{1}{2}}) \subset F'(H^{\frac{1}{2}} A H^{-\frac{1}{2}}) \subset \Gamma$, which completes the proof. □

Simultaneous diagonalization by congruence

We now return to the notion of simultaneous diagonalization of matrices by congruence (*congruence), which was discussed initially in Section (4.5) of [HJ].

1.7.13 Definition. We say that $A_1, A_2,\ldots, A_m \in M_n$ are *simultaneously diagonalizable by congruence* if there is a nonsingular matrix $C \in M_n$ such that each of C^*A_1C, $C^*A_2C,\ldots, C^*A_mC$ is diagonal.

A notion that arises in the study of simultaneous diagonalization by congruence and links it to the field of values, especially those topics studied in this section, is that of zeroes of the Hermitian form x^*Ax.

1.7.14 Definition. A nonzero vector $x \in \mathbb{C}^n$ is said to be an *isotropic vector* for a given matrix $A \in M_n$ if $x^*Ax = 0$, and x is further said to be a *common isotropic vector* for $A_1,\ldots, A_m \in M_n$ if $x^*A_ix = 0$, $i = 1,\ldots, m$.

Two simple observations will be of use.

1.7.15 Replacement Lemma. Let $\alpha_1,\ldots, \alpha_m \in \mathbb{C}$ and $A_1,\ldots, A_m \in M_n$ be given. The matrices $A_1,\ldots, A_m$ are simultaneously diagonalizable by congruence if and only if the matrices

$$\sum_{i=1}^{n} \alpha_iA_i,\ A_2,\ A_3,\ldots,\ A_m,\quad \alpha_1 \neq 0$$

are simultaneously diagonalizable by congruence.

Exercise. Perform the calculation necessary to verify (1.7.15).

If $m = 2$ and if $A_1, A_2 \in M_n$ are Hermitian, we often study $A = A_1 + iA_2$ in order to determine whether A_1 and A_2 are simultaneously diagonalizable by congruence. Note that $A_1 = \mathrm{H}(A)$ and $iA_2 = \mathrm{S}(A)$.

1.7.16 Observation. Let $A_1, A_2 \in M_n$ be Hermitian. Then A_1 and A_2 are simultaneously diagonalizable by congruence if and only if $A = A_1 + iA_2$ is congruent to a normal matrix.

Proof: If A_1 and A_2 are simultaneously diagonalizable by congruence via $C \in M_n$, then $C^*AC = C^*A_1C + iC^*A_2C = D_1 + iD_2$, where D_1 and D_2 are real diagonal matrices. The matrix C^*AC is then diagonal and, thus, is normal.

Conversely, suppose that a given nonsingular matrix $B \in M_n$ is such that B^*AB is normal. Then there is a unitary matrix $U \in M_n$ such that $U^*B^*ABU = D$ is diagonal. Now set $C = BU$ and notice that $\mathrm{H}(D) = \mathrm{H}(C^*AC) = C^*\mathrm{H}(A)C = C^*A_1C$ is diagonal and that $-i\mathrm{S}(D) = -i\mathrm{S}(C^*AC) = -iC^*\mathrm{S}(A)C = C^*A_2C$ is diagonal. $\square$

We may now apply (1.7.11) to obtain a classical sufficient condition for two Hermitian matrices to be simultaneously diagonalizable by congruence.

Exercise. If $A_1, A_2 \in M_n$ are Hermitian and if $A = A_1 + iA_2$, show that $0 \notin F(A)$ if and only if A_1 and A_2 have no common isotropic vector.

1.7.17 Theorem. If $A_1, A_2 \in M_n$ are Hermitian and have no common isotropic vector, then A_1 and A_2 are simultaneously diagonalizable by congruence.

Proof: If A_1 and A_2 have no common isotropic vector, then $0 \notin F(A)$, where $A \equiv A_1 + iA_2$. Choosing $B = A$ in (1.7.11a), we conclude from the equivalence of (1.7.11a) and (1.7.11b) that A is congruent to a normal matrix. According to (1.7.16), then, A_1 and A_2 are simultaneously diagonalizable by congruence. $\square$

Other classical sufficient conditions for pairs of Hermitian matrices to be simultaneously diagonalizable by congruence follow directly from (1.7.17).

Exercise. Let A_1, $A_2 \in M_n$ be given. Show that the set of common isotropic vectors of A_1 and A_2 is the same as the set of common isotropic vectors of $\alpha_1 A + \alpha_2 A_2$ and A_2 if $\alpha_1 \neq 0$. Since a positive definite matrix has no isotropic vectors, conclude that A_1 and A_2 have no common isotropic vectors if a linear combination of A_1 and A_2 is positive definite.

1.7.18 Corollary. If $A_1 \in M_n$ is positive definite and if $A_2 \in M_n$ is Hermitian, then A_1 and A_2 are simultaneously diagonalizable by congruence.

1.7.19 Corollary. If A_1, $A_2 \in M_n$ are Hermitian and if there is a linear combination of A_1 and A_2 that is positive definite, then A_1 and A_2 are simultaneously diagonalizable by congruence.

Exercise. Deduce (1.7.18) and (1.7.19) from (1.7.17) using (1.7.15) in the case of (1.7.19).

Exercise. Prove (1.7.18) directly, using Sylvester's law of inertia.

Exercise. If $A_1, A_2 \in M_n$ are Hermitian, show that there is a real linear combination of A_1 and A_2 that is positive definite if and only if A_1 and A_2 have no common isotropic vector. *Hint:* Choose $0 \leq \theta < 2\pi$ so that $e^{i\theta}F(A) \subset RHP$, $A \equiv A_1 + iA_2$, and calculate $\mathrm{H}(e^{i\theta}A) = (\cos\theta)A_1 - (\sin\theta)A_2$.

1.7.20 Example. The simple example $A_1 = A_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ shows that the direct converses of (1.7.17), (1.7.18), and (1.7.19) do not hold.

Using (1.7.11) again, however, we can provide a converse to (1.7.17) that emphasizes the role of the condition "$0 \notin F(A)$" in the study of simultaneous diagonalization by congruence. We conclude this discussion of simultaneous diagonalization by congruence for two Hermitian matrices by listing this converse (1.7.21e) along with other equivalent conditions.

1.7.21 Theorem. Let $A_1, A_2 \in M_n$ be Hermitian and assume that A_1 and $A \equiv A_1 + iA_2$ are nonsingular. The following are equivalent:

(a) A_1 and A_2 are simultaneously diagonalizable by congruence;
(b) A is diagonalizable by congruence;
(c) A is congruent to a normal matrix;
(d) $A^{-1}A^*$ is similar to a unitary matrix;
(e) There is some $\hat{A} \in M_n$ with $0 \notin F(\hat{A})$, such that $A^{-1}A^* = \hat{A}^{-1}\hat{A}^*$; and
(f) $A_1^{-1}A_2$ is similar to a real diagonal matrix.

Exercise. Prove (1.7.21). These equivalences are merely a compilation of previous developments. Observation (1.7.16) shows that (a) and (c) are equivalent, independent of the nonsingularity assumptions. Show that (b) and (c) are equivalent, independent of the nonsingularity assumptions, using the theory of normal matrices. Items (c), (d), and (e) are equivalent because of (1.7.11), and the equivalence of (a) and (f) was proven in Theorem (4.5.15) in [HJ]. Notice that the assumption of nonsingularity of A_1 is necessary only because of (f) and the assumption of nonsingularity of A is necessary only because of (d) and (e).

Exercise. Many conditions for two Hermitian matrices to be simultaneously diagonalizable by congruence can be thought of as generalizations of Corollary (1.7.18). Consider (1.7.21f) and show that if A_1 is positive definite, then $A_1^{-1}A_2$ has real eigenvalues and is diagonalizable by similarity.

Extensions of the Hermitian pair case to non-Hermitian matrices and to more than two matrices are developed in Problems 18 and 19.

Hadamard products

Recall that the Hadamard product $A \circ B$ is just the entrywise product of two matrices of the same size (see Definition (5.0.1)). For comparison, we state, without proof here, some results relating $F(\cdot)$ and the Hadamard product. Tools for the first of these (1.7.22) are developed at the end of Section (4.2). The remaining four results (1.7.23-26) may be deduced from (1.7.22).

1.7.22 Theorem. If A, $N \in M_n$ and if N is normal, then $F(N \circ A) \subset \mathrm{Co}(F(N)F(A))$.

Proof: See Corollary (4.2.17).

1.7.23 Corollary. If A, $H \in M_n$ and if H is positive semidefinite, then $F(H \circ A) \subset F(H)F(A)$.

Exercise. Deduce (1.7.23) from (1.7.22) using the facts that H is normal and $F(H)F(A)$ is convex.

1.7.24 Corollary. If A, $B \in M_n$ and if either A or B is normal, then $r(A \circ B) \leq r(A)r(B)$.

Exercise. Deduce (1.7.24) from (1.7.22).

1.7.25 Corollary. If A, $B \in M_n$, then $r(A \circ B) \leq 2r(A)r(B)$.

Exercise. Deduce (1.7.25) from (1.7.24) using the representation $A = \mathrm{H}(A) + \mathrm{S}(A)$. Contrast this result with the corresponding inequality for usual matrix multiplication, given in Problem 22 in Section (5.7) of [HJ].

1.7.26 Theorem. If A, $H \in M_n$ and if H is positive definite, then $F'(H \circ A) \subset F'(A)$.

Exercise. Deduce (1.7.26) from (1.7.23).

1.7.27 Example. The example $A = B = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$ shows that the constant 2 is the best possible in (1.7.25) and that (1.7.22) cannot be generalized to arbitrary A, $N \in M_n$.

1.7.28 Example. The example $A = N = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$ shows that the "Co" cannot be dropped in (1.7.22), even when both A and N are normal.

Exercise. Show that $F'(A) \subset \cup\{F(H \circ A)\colon H \text{ is positive definite}\}$.

Problems

1. Determine all convex sets K such that KS is convex for all convex sets S.

2. In spite of the failure of (1.7.1a) and example (1.7.4), show that $F(A^2) \subset \mathrm{Co}(F(A)^2)$ for all $A \in M_n$. Conclude that $r(A^2) \leq r(A)^2$ for all $A \in M_n$.

3. The power inequality for the numerical radius, $r(A^k) \leq r(A)^k$ for all positive integers k [see Problem 23 in Section (1.5)], suggests the conjecture

$$F(A^k) \subset \mathrm{Co}(F(A)^k) \text{ for all } k = 1, 2, \ldots$$

Prove this conjecture and the corresponding numerical radius inequality when A is normal. C.-K. Li and M.-D. Choi have constructed non-normal counterexamples to the conjecture for all $k \geq 2$.

4. If $A \in M_n(\mathbb{R})$, show that $F(A)$ is symmetric with respect to the real axis, and therefore $F(A^*) = F(A^T) = F(A)$.

5. If $A \in M_n(\mathbb{R})$ is nonsingular, show that $F'(A^{-1}) = F'(A)$.

6. Under the assumptions of Theorem (1.7.6), show that $\sigma(B^{-1}A) \subset F(A)/F(B)$ also.

7. If A is normal and B is positive definite, show that the eigenvalues of AB are contained in the convex cone generated by the eigenvalues of A.

8. Suppose that $A \in M_n$ is unitary with eigenvalues $e^{i\theta_1}, \ldots, e^{i\theta_n}$ and suppose, moreover, that $B \in M_n$ is positive semidefinite with eigenvalues $\beta_1 \leq \cdots \leq \beta_n$. Let $\lambda = re^{i\theta}$, $r > 0$, be a given eigenvalue of BA. Show that $\beta_1 \leq r \leq \beta_n$, and, if all $e^{i\theta_j}$ are contained in an arc of the unit circle of length less than or equal to π, show that θ is also contained in this arc.

9. Under the assumptions of (1.7.8), explain why one can prove the apparently better result $\sigma(AB) \subset F(A)F'(B)$. Why isn't this better?

10. Theorem (1.7.8) shows that some angular information about $\sigma(AB)$ may be gained in special circumstances. Note that this might be mated with the value of a matrix norm of A and B to give magnitudinal information also. Give an example.

11. If $A = [a_{ij}] \in M_n$ has positive diagonal entries, define $\hat{A} = [\hat{a}_{ij}]$ by

$$\hat{a}_{ij} \equiv \begin{cases} (|a_{ij}| + |a_{ji}|)/(2a_{ii}) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

Let $\Gamma_t = \{re^{i\theta}: r > 0, -\theta_0 \leq \theta \leq \theta_0\}$, where $\theta_0 = \arcsin(t)$, $0 \leq \theta_0 < \pi/2$ and $0 \leq t < 1$. Show that if $A \in M_n$ has positive diagonal entries and if $\rho(\hat{A}) < 1$, then $F'(A) \subset \Gamma_{\rho(\hat{A})}$.

12. Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$. Compare the angular information about $\sigma(AB)$ given by (1.7.8) with that obtained from direct application of Geršgorin's The rem (6.1.1,2) in [HJ] to the product AB. Note that $F'(A) = F'(B) = \Gamma_{\frac{1}{2}}$ and $\Gamma_{\frac{1}{2}}\Gamma_{\frac{1}{2}} = \Gamma_{\sqrt{1/2}}$. Note also that $\rho(\hat{A}) = \rho(\hat{B}) = \frac{1}{2}$, so these are cases of equality for the result of the preceding exercise. Either use that exercise or calculate $F'(A)$ and $F'(B)$ directly.

13. Generalize (1.7.10) for sectors of $\mathbb{C}$ other than RHP.

14. Show that $F(A^k) \subset RHP$ for all positive integers k if and only if A is positive definite.

15. Use the ideas of the proof of (1.7.9) to show that the result continues to hold if all matrices are restricted to be real.

16. If $A = H + S$, with H positive definite and S skew-Hermitian, verify that $H^{-\frac{1}{2}}SH^{-\frac{1}{2}}$ is skew-Hermitian and that $H^{-\frac{1}{2}}(H + S)H^{-\frac{1}{2}}$ is normal.

17. Show that for any $A \in M_n$ there is a real number t_0 such that $A + tI$ is

congruent to a normal matrix for all $t > t_0$.

18. Let $A_1,\ldots, A_m \in M_n$. Show that the following statements are equivalent:

(a) $A_1,\ldots, A_m$ are simultaneously diagonalizable by congruence.
(b) $A_1,\ldots, A_m$ are simultaneously congruent to commuting normal matrices.
(c) The $2m$ Hermitian matrices $\mathrm{H}(A_1)$, $i\mathrm{S}(A_1)$, $\mathrm{H}(A_2)$, $i\mathrm{S}(A_2),\ldots,$ $\mathrm{H}(A_m)$, $i\mathrm{S}(A_m)$ are simultaneously diagonalizable by congruence.

19. If $A_1,\ldots, A_{2q} \in M_n$ are Hermitian, show that $A_1,\ldots, A_{2q}$ are simultaneously diagonalizable by congruence if and only if $A_1 + iA_2$, $A_3 + iA_4,\ldots,$ $A_{2q-1} + iA_{2q}$ are simultaneously diagonalizable by congruence.

20. Let a nonsingular $A \in M_n$ be given. Explain why both factors in a polar decomposition $A = PU$ are uniquely determined, where $U \in M_n$ is unitary. Prove that $F'(U) \subset F'(A)$.

21. Let $A \in M_n$ be given with $0 \notin F(A)$ and let $A = PU$ be the polar factorization of A with $U \in M_n$ unitary. Prove that U is a *cramped unitary matrix*, that is, the eigenvalues of U lie on an arc of the unit circle of length less than π. Also show that $\Theta(U) \leq \Theta(A)$, where $\Theta(\cdot)$ is the field angle (1.1.3). Show that a unitary matrix V is cramped if and only if $0 \notin F(V)$.

22. Let $G \subset \mathbf{C}$ be a nonempty open convex set, and let $A \in M_n$ be given. Show that $\sigma(A) \subset G$ if and only if there is a nonsingular matrix $S \in M_n$ such that $F(S^{-1}AS) \subset G$.

23. If $A, B \in M_n$ are both normal with $\sigma(A) = \{\alpha_1,\ldots, \alpha_n\}$ and $\sigma(B) = \{\beta_1,\ldots, \beta_n\}$, show that $F(A \circ B)$ is contained in the convex polygon determined by the points $\alpha_i \beta_j$, for $i, j = 1,\ldots, n$.

24. Let $A \in M_n$ be given. Show that the following are equivalent:

(a) $A^* = S^{-1}AS$ for some $S \in M_n$ with $0 \notin F(S)$.
(b) $A^* = P^{-1}AP$ for some positive definite $P \in M_n$.
(c) A is similar to a Hermitian matrix.
(d) $A = PK$ for $P, K \in M_n$ with P positive definite and K Hermitian.

Compare with (4.1.7) in [HJ].

25. Let $A \in M_n$ be a given normal matrix. Show that there is an $S \in M_n$ with $0 \notin F(S)$ and $A^* = S^{-1}AS$ if and only if A is Hermitian.

26. Let $A \in M_n$ be given. Show that there is a unitary $U \in M_n$ with $0 \notin F(U)$ and $A^* = U^*AU$ if and only if A is Hermitian. Consider $A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$ and $U = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ to show that the identity $A^* = U^*AU$ for a general unitary U does not imply normality of A.

27. (a) If $A \in M_n$ and $0 \in F(A)$, show that $|z| \leq |||A - zI|||_2$ for all $z \in \mathbb{C}$.
(b) Use (a) to show that $|||AB - BA - I|||_2 \geq 1$ for all $A, B \in M_n$; in particular, the identity matrix is not a commutator. See Problem 6 in Section (4.5) for a generalization of this inequality to any matrix norm.
(c) If $A \in M_n$ and there is some $z \in \mathbb{C}$ such that $|||A - zI|||_2 < |z|$, show that $0 \notin F(A)$.

28. Let nonsingular $A, B \in M_n$ be given and let $C \equiv ABA^{-1}B^{-1}$. If A and C are normal, $AC = CA$, and $0 \notin F(B)$, show that $AB = BA$, that is, $C = I$. In particular, use Problem 21 to show that if A and B are unitary, $AC = CA$, and B is a cramped unitary matrix, then $AB = BA$. For an additive commutator analog of this result, see Problem 12 in Section (2.4) of [HJ].

Notes and Further Readings. The results (1.7.6-8) are based upon H. Wielandt, On the Eigenvalues of $A + B$ and AB, *J. Research N.B.S.* 77B (1973), 61-63, which was redrafted from Wielandt's National Bureau of Standards Report #1367, December 27, 1951, by C. R. Johnson. Theorem (1.7.9) is from C. R. Johnson, The Field of Values and Spectra of Positive Definite Multiples, *J. Research N.B.S.* 78B (1974) 197-198, and (1.7.10) was first proved by D. Carlson in A New Criterion for H-Stability of Complex Matrices, *Linear Algebra Appl.* 1 (1968), 59-64. Theorem (1.7.11) is from C. R. DePrima and C. R. Johnson, The Range of $A^{-1}A^*$ in $GL(n,\mathbb{C})$, *Linear Algebra Appl.* 9 (1974), 209-222, and (1.7.12) is from C. R. Johnson, A Lyapunov Theorem for Angular Cones, *J. Research National Bureau Standards* 78B (1974), 7-10. The treatment of simultaneous diagonalization by congruence is a new exposition centered around the field of values, which includes some classical results such as (1.7.17).

1.8 Generalizations of the field of values

There is a rich variety of generalizations of the field of values, some of which have been studied in detail. These generalizations emphasize various algebraic, analytic, or axiomatic aspects of the field of values, making it one of the most generalized concepts in mathematics. With no attempt to be complete, we mention, with occasional comments, several prominent or natural generalizations; there are many others. A natural question to ask about any generalized field of values is whether or not it is always convex: For some generalizations it is, and for some it is not. This gives further insight into the convexity (1.2.2) of the usual field of values, certainly one of its more subtle properties.

The first generalization involves a natural alteration of the inner product used to calculate each point of the field.

1.8.1 *Generalized inner product:* Let $H \in M_n$ be a given positive definite matrix. For any $A \in M_n$, define

$$F_H(A) \equiv \{x^* HAx: x \in \mathbb{C}^n, x^* Hx = 1\}$$

1.8.1a Observation. Since H is positive definite, it can be written as $H = S^* S$ with $S \in M_n$ nonsingular. Then $F_H(A) = F(SAS^{-1})$, so $F_H(A)$ is just the usual field of values of a similarity of A, by a fixed similarity matrix.

Proof: $F_H(A) = \{x^* S^* SAS^{-1} Sx: x \in \mathbb{C}^n, x^* S^* Sx = 1\} = \{y^* SAS^{-1} y: y \in \mathbb{C}^n, y^* y = 1\} = F(SAS^{-1})$. □

Exercise. There are many different matrices $S \in M_n$ such that $S^* S = H$. Why does it not matter which is chosen? *Hint:* If $T^* T = S^* S$, show that ST^{-1} is unitary and apply (1.2.8).

Exercise. Consider the fixed similarity generalization $F^S(A) \equiv F(SAS^{-1})$ for a fixed nonsingular $S \in M_n$. Show that $F^S(\cdot) = F_{S^* S}(\cdot) = F_H(\cdot)$ for $H \equiv S^* S$, so that the fixed similarity and generalized inner product generalizations are the same.

Exercise. Why is $F_H(A)$ always convex?

Exercise. Show that $\sigma(A) \subset F_H(A)$ for any positive definite matrix H. Show further that $\cap\{F_H(A): H \in M_n \text{ is positive definite}\} = \text{Co}(\sigma(A))$. *Hint:* Use (2.4.7) in [HJ]. For which matrices A does there exist a positive definite H with $F_H(A) = \text{Co}(\sigma(A))$?

Another generalization of the field of values is motivated by the fact that $x^* Ax \in M_1$ when $x \in \mathbb{C}^n$, so the usual field is just the set of determinants of all such matrices for normalized x.

1.8.2 *Determinants of isometric projections:* For any $m \leq n$ and any $A \in M_n$, define

$$F_m(A) \equiv \{\det(X^* AX): X \in M_{n,m}, X^* X = I \in M_m\}$$

Exercise. What are $F_1(\cdot)$ and $F_n(\cdot)$?

Exercise. Construct an example to show that $F_m(A)$ is not always convex.

The so-called k-field of values is another generalization, which is based upon replacement of a single normalized vector x with an orthonormal set. It is more related to the trace than to the determinant.

1.8.3 *The k-field of values:* Let k be a given positive integer. For any $A \in M_n$, define

$$F^k(A) \equiv \{\tfrac{1}{k}(x_1^* Ax_1 + \cdots + x_k^* Ax_k): X = [x_1 \ldots x_k] \in M_{n,k} \text{ and } X^* X = I\}$$

The set $F^k(A)$ is always convex, but the proof is involved, and this fact is a special case of a further generalization to follow.

Exercise. Show that both $F_m(\cdot)$ and $F^k(\cdot)$ are invariant under unitary similarity of their arguments.

Exercise. What are $F^1(A)$ and $F^n(A)$?

Exercise. If A is normal with eigenvalues $\lambda_1, \ldots, \lambda_n$, what is $F^k(A)$?

Exercise. Show that $F^n(A) \subset F^{n-1}(A) \subset \cdots \subset F^2(A) \subset F^1(A)$.

A further generalization of the k-field of values (1.8.3) is the c-field of values in which the equal positive coefficients $\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k}$ are replaced by arbitrary complex numbers.

1.8.4 *The c-field of values:* Let $c = [c_1, \ldots, c_n]^T \in \mathbb{C}^n$ be a given vector. For any $A \in M_n$, define

$$F^c(A) \equiv \{c_1 x_1^* Ax_1 + \cdots + c_n x_n^* Ax_n: X = [x_1 \ldots x_n] \in M_n \text{ and } X^* X = I\}$$

Exercise. Show that $F^k(A) = F^c(A)$ for $c = [1,\dots,1,0,\dots,0]^T/k \in \mathbb{C}^n$. For what c is $F^c(A) = F(A)$?

Exercise. Is $F^c(\cdot)$ invariant under unitary similarity of its argument?

Exercise. Show that $F^c(A) = \{\mathrm{tr}(CU^*AU)\colon U \in M_n \text{ is unitary}\}$, where $C = \mathrm{diag}(c_1,\dots,c_n)$.

Unfortunately, the c-field of values $F^c(A)$ is not always convex for every $c \in \mathbb{C}^n$, but it is always convex when c is a *real* vector. More generally, if $c \in \mathbb{C}^n$ is given, $F^c(A)$ is convex for all $A \in M_n$ if and only if the entries of c are collinear as points in the complex plane. If the entries of c are not collinear, there exists a normal $A \in M_n$ such that $F^c(A)$ is not convex. When $n \geq 5$, the ordinary field of values of $A \in M_n$ can be a convex polygon without A being normal, but it is known that A is normal if and only if there is some $c \in \mathbb{R}^n$ with distinct entries such that $F^c(A)$ is a convex polygon.

A set $S \subset \mathbb{C}$ is said to be *star-shaped with respect to a given point* $z_0 \in S$ if for every $z \in S$ the whole line segment $\{\alpha z + (1-\alpha)z_0\colon\ 0 \leq \alpha \leq 1\}$ lies in S; a set $S \subset \mathbb{C}$ is convex if and only if it is star-shaped with respect to every point in S. Although some c-fields of values fail to be convex, it is known that $F^c(A)$ is always star-shaped with respect to the point $\frac{1}{n}(\mathrm{tr}\, A)(c_1 + \cdots + c_n)$.

Exercise. Construct an example of a complex vector c and a matrix A for which $F^c(A)$ is not convex.

Exercise. Consider the vector set-valued function ${}^mF(\cdot)$ defined on M_n by

$${}^mF(A) = \{[x_1^*Ax_1,\dots,x_m^*Ax_m]^T\colon X = [x_1 \dots x_m] \in M_{n,m} \text{ and } X^*X = I\}$$

Show that $F^c(A)$ is the projection of ${}^nF(A)$ into $\mathbb{C}$ by the linear functional whose coefficients are the components of the vector c. Show that ${}^mF(\cdot)$ is invariant under unitary similarity of its argument and that if $y \in {}^mF(A)$, then $Py \in {}^mF(A)$ for any permutation matrix $P \in M_n$.

Naturally associated with the c-field of values is the *c-numerical radius*

$$\begin{aligned} r_c(A) &\equiv \max\{|z|\colon z \in F^c(A)\} \\ &= \max\{|\mathrm{tr}(CU^*AU)|\colon\ U \in M_n \text{ is unitary}\} \end{aligned}$$

where $c = [c_1,\ldots, c_n]^T \in \mathbb{C}^n$ and $C \equiv \operatorname{diag}(c_1,\ldots, c_n)$. When $n \geq 2$, the function $r_c(\cdot)$ is a norm on M_n if and only if $c_1 + \cdots + c_n \neq 0$ and the scalars c_i are not all equal; this condition is clearly met for $c = e_1 = [1, 0,\ldots, 0]^T$, in which case $r_c(\cdot)$ is the usual numerical radius $r(\cdot)$. It is known that if $A, B \in M_n$ are Hermitian, then $r_c(A) = r_c(B)$ for all $c \in \mathbb{R}^n$ if and only if A is unitarily similar to $\pm B$.

A generalization of (1.8.4) is the C-field of values.

1.8.5 *The C-field of values:* Let $C \in M_n$ be given. For any $A \in M_n$, define

$$F^C(A) \equiv \{\operatorname{tr}(CU^*AU): U \in M_n \text{ is unitary}\}$$

Exercise. Show that $F^{V_1^*CV_1}(V_2^*AV_2) = F^C(A)$ if $V_1, V_2 \in M_n$ are unitary.

Exercise. If $C \in M_n$ is normal, show that $F^C(A) = F^c(A)$, where the vector c is the vector of eigenvalues of the matrix C. Thus, the C-field is a generalization of the c-field.

Exercise. Show that $F^A(C) = F^C(A)$ for all $A, C \in M_n$. Deduce that $F^C(A)$ is convex if either C or A is Hermitian, or, more generally, if either A or C is normal and has eigenvalues that are collinear as points in the complex plane.

Known properties of the c-field cover the issue of convexity of C-fields of values for normal C, but otherwise it is not known which pairs of matrices $C, A \in M_n$ produce a convex $F^C(A)$ and for which C the set $F^C(A)$ is convex for all $A \in M_n$.

Associated with the C-field of values are natural generalizations of the numerical radius, spectral norm, and spectral radius:

$$|||A|||_C \equiv \max\{|\operatorname{tr}(CUAV)|: U, V \in M_n \text{ are unitary}\}$$

$$r_C(A) \equiv \max\{|\operatorname{tr}(CUAU^*)|: U \in M_n \text{ is unitary}\}$$

$$\rho_C(A) \equiv \max\left\{\left|\sum_{i=1}^n \lambda_i(A)\,\lambda_{\pi(i)}(C)\right|: \pi \text{ is a permutation of } 1,\ldots, n\right\}$$

where the eigenvalues of A and C are $\{\lambda_i(A)\}$ and $\{\lambda_i(C)\}$, respectively. For any $A, C \in M_n$, these three quantities satisfy the inequalities

$$\rho_C(A) \le r_C(A) \le |||A|||_C \tag{1.8.5a}$$

Exercise. If $C = E_{11} = \text{diag}(1,0,\ldots,0)$, show that $\rho_C(A) = \rho(A)$, $r_C(A) = r(A)$, and $|||A|||_C = |||A|||_2$.

A generalization of a different sort is a family of objects often called the *Bauer fields of values*. They generalize the fact that the usual field of values is related to the l_2 vector norm $\|\cdot\|_2$. There is one of these generalized fields of values for each vector norm on $\mathbb{C}^n$. If $\|\cdot\|$ is a given vector norm on $\mathbb{C}^n$, let $\|\cdot\|^D$ denote the vector norm that is dual to $\|\cdot\|$. See (5.4.12-17) in [HJ] for information about dual norms that is needed here.

1.8.6 *The Bauer field of values associated with the norm $\|\cdot\|$ on $\mathbb{C}^n$:* For any $A \in M_n$, define

$$F_{\|\cdot\|}(A) \equiv \{y^*Ax : x, y \in \mathbb{C}^n \text{ and } \|y\|^D = \|x\| = y^*x = 1\}$$

Notice that $F_{\|\cdot\|}(A)$ is just the range of the sesquilinear form y^*Ax over those normalized ordered pairs of vectors $(x,y) \in \mathbb{C}^n \times \mathbb{C}^n$ that are *dual pairs* with respect to the vector norm $\|\cdot\|$ (see (5.4.17) in [HJ]). The set $F_{\|\cdot\|}(A)$ is not always convex, but it always contains the spectrum $\sigma(A)$. There are further norm-related generalizations of the field of values that we do not mention here, but an example of another of these, a very natural one, is the set function $G_F(A)$ defined in (1.5.1).

Exercise. Show that $F_{\|\cdot\|_2}(A) = F(A)$ for every $A \in M_n$.

Exercise. Show that $\sigma(A) \subset F_{\|\cdot\|}(A)$ for every norm $\|\cdot\|$ on $\mathbb{C}^n$.

Exercise. Determine $F_{\|\cdot\|_1}\begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$.

Exercise. Show that $F_{\nu(\cdot)}(A^*) = \overline{F_{\|\cdot\|}(A)}$, where $\nu(\cdot) \equiv \|\cdot\|^D$.

Exercise. If $\|x\|_S \equiv \|Sx\|$ for a nonsingular $S \in M_n$, determine $F_{\|\cdot\|_S}(A)$ in terms of $F_{\|\cdot\|}(\cdot)$.

Exercise. Give an example of an $F_{\|\cdot\|}(A)$ that is not convex.

For $A \in M_n(\mathbb{R})$, it may at first seem unnatural to study $F(A)$, which is defined in terms of vectors in $\mathbb{C}^n$. The motivation for this should be clear, though, from the results presented in this chapter so far. If $A \in M_n(\mathbb{R})$, there is a strictly real object corresponding to $F(A)$ that has been studied.

1.8.7 Definition. *The real field of values:* For any $A \in M_n(\mathbb{R})$, define

$$FR(A) \equiv \{x^TAx : x \in \mathbb{R}^n \text{ and } x^Tx = 1\}$$

Notice that $FR(A) = FR([A + A^T]/2)$, so in studying the real field of values it suffices to consider only real symmetric matrices.

Exercise. Show that $FR(A) \subset \mathbb{R}$ and that all real eigenvalues of $A \in M_n(\mathbb{R})$ are contained in $FR(A)$.

Exercise. Show that $FR(A)$ is easy to compute: It is just the real line segment joining the smallest and largest eigenvalues of $\mathrm{H}(A) = (A + A^T)/2$. In particular, therefore, $FR(A) = F(\mathrm{H}(A))$ is always convex.

Exercise. If $A \in M_n(\mathbb{R})$, show that $FR(A) = F(A)$ if and only if A is symmetric.

Exercise. Which of the basic functional properties (1.2.1-12) does $FR(\cdot)$ share with $F(\cdot)$?

If $x, y \in \mathbb{C}^n$, $y^*y = 1$, and $x^*x = 1$, then $y^*x = 1$ implies that $y = x$ (this is a case of equality in the Cauchy-Schwarz inequality). This suggests another generalization of the field of values in which the Hermitian form x^*Ax is replaced by the sesquilinear form y^*Ax, subject to a relation between y and x.

1.8.8 Definition. *The q-field of values:* Let $q \in [0,1]$ be given. For any $A \in M_n$, define

$$F_q(A) \equiv \{y^*Ax : x, y \in \mathbb{C}^n,\ y^*y = x^*x = 1, \text{ and } y^*x = q\}$$

If $0 \leq q < 1$, $F_q(\cdot)$ is defined only for $n \geq 2$.

Exercise. Show that it is unnecessary to consider $F_q(\cdot)$ for a $q \in \mathbb{C}$ that is not nonnegative, for if $q \in \mathbb{C}$ with $|q| \leq 1$ and $F_q(\cdot)$ is defined analogously ($y^*x = q$), then $F_q(A) = e^{i\theta}F_{|q|}(A)$, where $q = e^{i\theta}|q|$.

Exercise. Let $q = 1$ and show that $F_1(A) = F(A)$ for all A. Thus, $F_1(A)$ is always convex.

Exercise. Let $q = 0$ and show that $F_0(A)$ is always a disc centered at the origin, and hence $F_0(A)$ is always convex. *Hint:* If $z \in F_0(A)$, show that $e^{i\theta}z \in F_0(A)$ for all $\theta \in [0,2\pi)$. If z_1, z_2 are two given points in $F_0(A)$, show that they are connected by a continuous path lying in $F_0(A)$ (any pair of orthonormal vectors can be rotated continuously into any other pair of orthonormal vectors). It follows that $F_0(A)$ is an annulus centered at the origin. Finally, show that $0 \in F_0(A)$ by considering y^*Ax for an eigenvector x of A and some y that is orthogonal to x.

Exercise. Show that $F_0(A)$ is the set of 1,2 entries of matrices that are unitarily similar to A. The ordinary field of values may be thought of as the set of 1,1 entries of matrices that are unitarily similar to A.

N. K. Tsing has shown that $F_q(A)$ is convex for all $q \in [0,1]$ and all $A \in M_n$, $n \geq 2$.

Thus far, the classical field of values and all but two of the generalizations of it we have mentioned have been objects that lie intrinsically in two real dimensions $[{}^mF(\cdot) \subset \mathbb{C}^m$ and $FR(A) \subset \mathbb{R}$ for $A \in M_n(\mathbb{R})]$. Another generalization, sometimes called the *shell*, lies in three real dimensions in an attempt to capture more information about the matrix.

1.8.9 Definition. *The Davis-Wielandt shell:* For any $A \in M_n$, define

$$DW(A) \equiv \{[\mathrm{Re}\, x^*Ax, \mathrm{Im}\, x^*Ax, x^*(A^*A)x]^T \colon x \in \mathbb{C}^n, x^*x = 1\}$$

Exercise. Let $A \in M_n$ have eigenvalues $\{\lambda_1,\dots, \lambda_n\}$. If A is normal, show that $DW(A) = \mathrm{Co}(\{[\mathrm{Re}\,\lambda_i, \mathrm{Im}\,\lambda_i, |\lambda_i|^2]^T \colon i = 1,\dots, n\})$.

One motivation for the definition of the Davis-Wielandt shell is that the converse of the assertion in the preceding exercise is also true. This is in contrast to the situation for the classical field of values, for which the simple converse to the analogous property (1.2.9) is not valid, as discussed in Section (1.6).

There are several useful multi-dimensional generalizations of the field of values that involve more than one matrix. The first is suggested naturally by thinking of the usual field of values as an object that lies in two real dimensions. For any $A \in M_n$, write $A = A_1 + iA_2$, where $A_1 = \mathrm{H}(A) = (A + A^*)/2$ and $A_2 = -i\mathrm{S}(A) = -i(A - A^*)/2$ are both Hermitian. Then

$F(A) = \{x^*Ax: x \in \mathbb{C}^n, x^*x = 1\} = \{x^*A_1x + ix^*A_2x: x \in \mathbb{C}^n, x^*x = 1\}$, which (since x^*A_1x and x^*A_2x are both real) describes the same set in the plane as $\{(x^*A_1x, x^*A_2x): x \in \mathbb{C}^n, x^*x = 1\}$. Thus, the Toeplitz-Hausdorff theorem (1.2.2) says that the latter set in $\mathbb{R}^2$ is convex for any two Hermitian matrices $A_1, A_2 \in M_n$, and we are led to the following generalizations of $F(A)$ and $FR(A)$.

1.8.10 Definition. *The k-dimensional field of k matrices:* Let $k \geq 1$ be a given integer. For any $A_1,\ldots, A_k \in M_n$ define

$$FC_k(A_1,\ldots, A_k) = \{[x^*A_1x,\ldots, x^*A_kx]^T: x \in \mathbb{C}^n, x^*x = 1\} \subset \mathbb{C}^k$$

Similarly, if $A_1,\ldots, A_k \in M_n(\mathbb{R})$, define

$$FR_k(A_1,\ldots, A_k) \equiv \{[x^TA_1x,\ldots, x^TA_kx]^T: x \in \mathbb{R}^n, x^Tx = 1\} \subset \mathbb{R}^k$$

Notice that when $k = 1$, $FC_1(A) = F(A)$ and $FR_1(A) = FR(A)$. For $k = 2$, $FC_2([A + A^*]/2, -i[A - A^*]/2)$ describes a set in a real two-dimensional subspace of $\mathbb{C}^2$ that is the same as $F(A)$. If the matrices $A_1,\ldots, A_k$ are all Hermitian, then $FC_k(A_1,\ldots, A_k) \subset \mathbb{R}^k \subset \mathbb{C}^k$. In considering FC_k, the case in which all of the k matrices are Hermitian is quite different from the general case in which the matrices are arbitrary, but in studying FR_k, it is convenient to know that we always have $FR_k(A_1,\ldots, A_k) = FR_k([A_1 + A_1^*]/2,\ldots, [A_k + A_k^*]/2)$.

Exercise. For any Hermitian matrices $A_1, A_2 \in M_n$, note that $FC_1(A_1) \subset \mathbb{R}^1$ and $FC_2(A_1,A_2) \subset \mathbb{R}^2$ are convex sets. For any real symmetric matrix $A_1 \in M_n(\mathbb{R})$, note that $FR_1(A_1) \subset \mathbb{R}^1$ is convex.

Exercise. For $n = 2$, consider the two real symmetric (Hermitian) matrices

$$A_1 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Show that $FR_2(A_1,A_2)$ is *not* convex but that $FC_2(A_1,A_2)$ *is* convex. Compare the latter set with $F(A_1 + iA_2)$. *Hint:* Consider the points in $FR_2(A_1,A_2)$ generated by $x = [1, 0]^T$ and $x = [0, 1]^T$. Is the origin in $FR_2(A_1,A_2)$?

The preceding exercise illustrates two special cases of what is known generally about convexity of the k-dimensional fields.

Convexity of FR_k:

For $n = 1$, $FR_k(A_1,\ldots, A_k)$ is a single point and is therefore convex for all $A_1,\ldots, A_k \in M_1(\mathbb{R})$ and all $k \geq 1$. For $n = 2$, $FR_1(A)$ is convex for every $A \in M_2(\mathbb{R})$, but for every $k \geq 2$ there are matrices $A_1,\ldots, A_k \in M_2(\mathbb{R})$ such that $FR_k(A_1,\ldots, A_k)$ is not convex. For $n \geq 3$, $FR_1(A_1)$ and $FR_2(A_1,A_2)$ are convex for all A_1, $A_2 \in M_n(\mathbb{R})$, but for every $k \geq 3$ there are matrices $A_1,\ldots, A_k \in M_n(\mathbb{R})$ such that $FR_k(A_1,\ldots, A_k)$ is not convex.

Convexity of FC_k for Hermitian matrices:

For $n = 1$, $FC_k(A_1,\ldots, A_k)$ is a single point and is therefore convex for all $A_1,\ldots, A_k \in M_1$ and all $k \geq 1$. For $n = 2$, $FC_1(A_1)$ and $FC_2(A_1,A_2)$ are convex for all Hermitian A_1, $A_2 \in M_2$, but for every $k \geq 3$ there are Hermitian matrices $A_1,\ldots, A_k \in M_2$ such that $FC_k(A_1,\ldots, A_k)$ is not convex. For $n \geq 3$, $FC_1(A_1)$, $FC_2(A_1,A_2)$, and $FC_3(A_1,A_2,A_3)$ are convex for all Hermitian A_1, A_2, $A_3 \in M_n$, but for every $k \geq 4$ there are Hermitian matrices $A_1,\ldots, A_k \in M_n$ such that $FR_k(A_1,\ldots, A_k)$ is not convex.

The definition (1.1.1) for the field of values carries over without change to matrices and vectors with quaternion entries, but there are some surprising new developments in this case. Just as a complex number can be written as $z = a_1 + ia_2$ with a_1, $a_2 \in \mathbb{R}$ and $i^2 = -1$, a *quaternion* can be written as $\zeta = a_1 + ia_2 + ja_3 + ka_4$ with a_1, a_2, a_3, $a_4 \in \mathbb{R}$, $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$, $jk = -kj = i$, and $ki = -ik = j$. The *conjugate* of the quaternion $\zeta = a_1 + ia_2 + ja_3 + ka_4$ is $\bar{\zeta} \equiv a_1 - ia_2 - ja_3 - ka_4$; its *absolute value* is $|\zeta| \equiv (a_1^2 + a_2^2 + a_3^2 + a_4^2)^{\frac{1}{2}}$; its *real part* is $\mathrm{Re}\,\zeta \equiv a_1$. The set of quaternions, denoted by $\mathbb{Q}$, is an algebraically closed division ring (noncommutative field) in which the inverse of a nonzero quaternion ζ is given by $\zeta^{-1} \equiv \bar{\zeta}/|\zeta|^2$. The quaternions may be thought of as lying in four real dimensions, and the real and complex fields may be thought of as subfields of $\mathbb{Q}$ in a natural way.

We denote by $M_n(\mathbb{Q})$ the set of n-by-n matrices with quaternion entries and write $\mathbb{Q}^n$ for the set of n-vectors with quaternion entries; for $x \in \mathbb{Q}^n$, x^* denotes the transpose of the entrywise conjugate of x.

1.8.11 Definition. *The quaternion field of values.* For any $A \in M_n(\mathbb{Q})$, define $FQ(A) \equiv \{x^*Ax\colon\ x \in \mathbb{Q}^n \text{ and } x^*x = 1\}$.

Although the quaternion field of values $FQ(A)$ shares many properties with the complex field of values, it need not be convex even when A is a normal complex matrix. If we set

$$A_1 \equiv \begin{bmatrix} i & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } A_2 \equiv \begin{bmatrix} i & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

then $FQ(A_1)$ is *not* convex but $FQ(A_2)$ *is* convex; in the classical case, $F(A_1)$ and $F(A_2)$ are identical. It is known that, for a given $A \in M_n(\mathbb{Q})$, $FQ(A)$ is convex if and only if $\{\mathrm{Re}\,\zeta\colon\ \zeta \in FQ(A)\} = \{\zeta\colon\ \zeta \in FQ(A) \text{ and } \zeta = \mathrm{Re}\,\zeta\}$, that is, if and only if the projection of $FQ(A)$ onto the real axis is the same as the intersection of $FQ(A)$ with the real axis.

Notes and Further Readings. The generalizations of the field of values that we have mentioned in this section are a selection of only a few of the many possibilities. Several other generalizations of the field of values are mentioned with references in [Hal 67]. Generalization (1.8.1) was studied by W. Givens in Fields of Values of a Matrix, *Proc. Amer. Math. Soc.* 3 (1952), 206-209. Some of the generalizations of the field of values discussed in this section are the objects of current research by workers in the field such as M. Marcus and his students. The convexity results in (1.8.4) are in R. Westwick, A Theorem on Numerical Range, *Linear Multilinear Algebra* 2 (1975), 311-315, and in Y. Poon, Another Proof of a Result of Westwick, *Linear Multilinear Algebra* 9 (1980), 35-37. The converse is in Y.-H. Au-Yeung and N. K. Tsing, A Conjecture of Marcus on the Generalized Numerical Range, *Linear Multilinear Algebra* 14 (1983), 235-239. The fact that the c-field of values is star-shaped is in N.-K. Tsing, On the Shape of the Generalized Numerical Ranges, *Linear Multilinear Algebra* 10 (1981), 173- 182. For a survey of results about the c-field of values and the c-numerical radius, with an extensive bibliography, see C.-K. Li and N.-K. Tsing, Linear Operators that Preserve the c-Numerical Range or Radius of Matrices, *Linear Multilinear Algebra* 23 (1988), 27-46. For a discussion of the generalized spectral norm, generalized numerical radius, and generalized spectral radius introduced in Section (1.8.5), see C.-K. Li, T.-Y. Tam, and N.-K. Tsing, The Generalized Spectral Radius, Numerical Radius and Spectral Norm, *Linear Multilinear Algebra* 16 (1984), 215-237. The convexity of $F_q(A)$ mentioned in (1.8.8) is shown in N. K. Tsing, The Constrained Bilinear Form and the C-Numerical Range, *Linear Algebra Appl.* 56 (1984), 195-206. The shell generalization (1.8.9) was considered independently and in alternate forms in C. Davis, The Shell of a Hilbert Space Operator, *Acta Sci. Math. (Szeged)* 29 (1968), 69-86 and in H. Wielandt, Inclusion Theorems for Eigenvalues, U.S. Department of Commerce, National Bureau of Standards, Applied Mathematics Series 29 (1953), 75-78. For more information about $FR_k(A_1,\dots, A_k)$ see L. Brickman, On the Field of Values of a Matrix, *Proc. Amer. Math. Soc.* 12 (1961), 61-66. For a proof that $FC_3(A_1,A_2,A_3)$ is convex when A_1, A_2, $A_3 \in M_n$ are Hermitian and $n \geq 3$, and for many related results, see P. Binding, Hermitian Forms and the Fibration of Spheres, *Proc. Amer. Math. Soc.* 94 (1985), 581-584. Also see Y.-H. Au-Yeung and N. K. Tsing, An Extension of the Hausdorff-Toeplitz Theorem on the Numerical Range, *Proc. Amer. Math. Soc.* 89 (1983), 215-218. For references to the literature and proofs of the assertions made about the quaternion field of values $FQ(A)$, see Y.-H. Au-Yeung, On the Convexity of Numerical Range in Quaternionic Hilbert Spaces, *Linear Multilinear Algebra* 16 (1984), 93-100. There is also a discussion of the quaternion field of values in Section II of the 1951 paper by R. Kippenhahn cited at the end of Section (1.6), but the reader is warned that Kippenhahn's basic Theorem 36 is *false:* the quaternion field of values is *not* always convex.

There are strong links between the field of values and Lyapunov's theorem (2.2). Some further selected readings for Chapter 1 are: C. S. Ballantine, Numerical Range of a Matrix: Some Effective Criteria, *Linear Algebra Appl.* 19 (1978), 117-188; C. R. Johnson, Computation of the Field of Values of a 2-by-2 Matrix, *J. Research National Bureau Standards* 78B (1974), 105-107; C. R. Johnson, Numerical Location of the Field of Values, *Linear Multilinear Algebra* 3 (1975), 9-14; C. R. Johnson, Numerical Ranges of Principal Submatrices, *Linear Algebra Appl.* 37 (1981), 23-34; F. Murnaghan, On the Field of Values of a Square Matrix, *Proc. National Acad. Sci. U.S.A.* 18 (1932), 246-248; B. Saunders and H. Schneider, A Symmetric Numerical Range for Matrices, *Numer. Math.* 26 (1976), 99-105; O. Taussky, A Remark Concerning the Similarity of a Finite Matrix A and A^*, *Math. Z.* 117 (1970), 189-190; C. Zenger, Minimal Subadditive Inclusion Domains for the Eigenvalues of Matrices, *Linear Algebra Appl.* 17 (1977), 233-268.