Eigenvalue inclusion regions from inverses of shifted matrices

Michiel E. Hochstenbach [a], David A. Singer [b], Paul F. Zachlin [b]

[a] *Department of Mathematics and Computer Science, TU Eindhoven, PO Box 513, NL-5600 MB Eindhoven, The Netherlands. The work of this author was supported in part by NSF grant DMS-0405387.*

[b] *Department of Mathematics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-7058, USA*

Abstract

We consider eigenvalue inclusion regions based on the field of values, pseudospectra, Gershgorin region, and Brauer region of the inverse of a shifted matrix. A family of these inclusion regions is derived by varying the shift. We study several properties, one of which is that the intersection of a family is exactly the spectrum. The numerical approximation of the inclusion sets for large matrices is also examined.

Key words: Harmonic Rayleigh–Ritz, inclusion regions, exclusion regions, inclusion curves, exclusion curves, field of values, numerical range, large sparse matrix, Gershgorin regions, ovals of Cassini, Brauer regions, pseudospectra, subspace methods, Arnoldi

1 Introduction

Let A be a nonsingular $n \times n$ matrix with spectrum $\Lambda(A)$ and field of values (or numerical range)

$$W(A) = \left\{ \frac{x^*Ax}{x^*x} : x \in \mathbb{C}^n \setminus \{0\} \right\}.$$

Email addresses: `m.e.hochstenbach tue.nl` (Michiel E. Hochstenbach), `david.singer case.edu` (David A. Singer), `paul.zachlin case.edu` (Paul F. Zachlin).

While it is well known that $\Lambda(A) \subseteq W(A)$, it was noted in [10, 11] that we also have

$$\Lambda(A) \subseteq W(A) \,\cap\, \frac{1}{W(A^{-1})}. \tag{1}$$

Here $1/W(A^{-1})$ is interpreted as the set

$$\frac{1}{W(A^{-1})} := \left\{ \frac{1}{z} : z \in W(A^{-1}) \right\}.$$

In this paper we will, inspired by the harmonic Rayleigh–Ritz technique, consider generalizations of (1) and study their properties. Section 2 briefly reviews harmonic the Rayleigh–Ritz approach, mentions some new results, and gives an idea why this technique may be exploited for spectral inclusion regions. In Section 3 eigenvalue inclusion regions based on the field of values of the inverse of a shifted version of the matrix are introduced. We characterize the spectrum of a matrix in a new way as the intersection of a family of these inclusion regions. Sections 4 and 5 focus on inclusion regions derived from Gershgorin and Brauer regions, and pseudospectra of shift-and-invert matrices. The practical subspace approximation of some of the introduced sets for large matrices is considered in Section 6. We give a few examples of the techniques and a practical algorithm in Section 7 and end with some conclusions in Section 8. For other results on inclusion regions see, e.g., [1].

2 Harmonic Rayleigh–Ritz and fields of values

The sets $W(A)$ and $1/W(A^{-1})$ have close connections with eigenvalue approximations from subspaces. Indeed, $W(A)$ can be seen as the set of all possible Ritz values from a one-dimensional subspace; see, e.g., [15] or also [5]. As was noted in [5], in view of

$$\left\{ \frac{x^*x}{x^*A^{-1}x} : x \neq 0 \right\} = \left\{ \frac{y^*A^*Ay}{y^*A^*y} : y \neq 0 \right\}, \tag{2}$$

the set $1/W(A^{-1})$ is exactly the set of all possible harmonic Ritz values determined by the harmonic Rayleigh–Ritz method with respect to *target* $\tau = 0$ from a one-dimensional subspace. After a brief review of the harmonic Rayleigh–Ritz approach, we will point out a generalization of this statement in Proposition 1 below.

The harmonic Rayleigh–Ritz technique [12–14] was introduced to better approximate interior eigenvalues using subspace methods near a given target $\tau \in \mathbb{C}$. Consider the standard eigenvalue problem $Ax = \lambda x$, and let $\mathcal{U}$ be a low dimensional search space for an eigenvector x with associated search matrix U of which the columns form an orthonormal basis for $\mathcal{U}$. We are interested in an approximation $(\lambda, x) \approx (\theta, u)$ with $u \in \mathcal{U}$. Instead of the Galerkin condition $Au - \theta u \perp \mathcal{U}$ of the standard Rayleigh–Ritz extraction, the harmonic Rayleigh–Ritz extraction with target τ imposes the Galerkin condition

$$Au - \theta u \perp (A - \tau I)\mathcal{U}. \tag{3}$$

This implies that a harmonic Ritz value θ satisfies

$$\theta = \frac{u^*(A - \tau I)^* A u}{u^*(A - \tau I)^* u},$$

where u is a harmonic Ritz vector.

As a generalization of (2) we have the following result. Note that in this proposition and throughout this paper addition and division are interpreted elementwise: for a set S, we define

$$\frac{1}{S} + \tau := \left\{ \frac{1}{z} + \tau : z \in S \right\}.$$

Proposition 1

$$\frac{1}{W((A - \tau I)^{-1})} + \tau = \left\{ \frac{y^*(A - \tau I)^* A y}{y^*(A - \tau I)^* y} : y \neq 0 \right\}.$$

Proof: This follows easily from the equality

$$\left\{ \frac{x^* x}{x^*(A - \tau I)^{-1} x} + \tau : x \neq 0 \right\} = \left\{ \frac{y^*(A - \tau I)^*(A - \tau I) y}{y^*(A - \tau I)^* y} + \tau : y \neq 0 \right\}.$$

□

Therefore, the set of all possible harmonic Ritz values with respect to target τ is the reciprocal of $W((A - \tau I)^{-1})$ shifted by τ.

An important property of harmonic Ritz values θ with respect to target τ is that they tend to stay away from this τ, which we shall study more closely in Section 6 (Proposition 14). Because of this property harmonic Ritz values are exploited in several situations. For instance, the GMRES method implicitly uses harmonic Ritz values for interpolation of the function $f(z) = z^{-1}$ resulting from linear systems $Ax = b$. Also, these values have found their way into the approximation of problems involving more general matrix functions $f(A)\, b$ where one would like to avoid a specific target. One example is the sign function which has a discontinuity in $z = 0$ [4, 16]. In this paper we will use sets as the one in Proposition 1 for eigenvalue inclusion regions.

A new interesting observation is the following. If we write $\eta = \tau^{-1}$ then (3) is equivalent to $Au - \theta u \perp (\eta A - I)\mathcal{U}$. If we take the limit $|\tau| \to \infty$ or, equivalently, $\eta \to 0$, we see that the standard Rayleigh–Ritz method can be viewed as the harmonic Rayleigh–Ritz method with target at infinity; see also the related Theorem 5 in the next section.

3 Eigenvalue inclusion regions from the field of values of inverses of shifted matrices

Let τ be a complex number not equal to an eigenvalue of A. A crucial observation that we will use is that

$$\Lambda(A) = \frac{1}{\Lambda((A-\tau I)^{-1})} + \tau.$$

Since $\Lambda((A-\tau I)^{-1}) \subseteq W((A-\tau I)^{-1})$, we know that for every $\tau \notin \Lambda(A)$ the spectrum $\Lambda(A)$ is included in the set

$$\frac{1}{W((A-\tau I)^{-1})} + \tau \tag{4}$$

and therefore

$$\Lambda(A) \subseteq \bigcap_{\tau \in \mathbb{C} \setminus \Lambda(A)} \frac{1}{W((A-\tau I)^{-1})} + \tau. \tag{5}$$

The following theorem shows that this inclusion is in fact an equality.

Theorem 2

$$\Lambda(A) = \bigcap_{\tau \in \mathbb{C} \setminus \Lambda(A)} \frac{1}{W((A-\tau I)^{-1})} + \tau.$$

Proof: Suppose we have $z \in \mathbb{C} \setminus \Lambda(A)$. We still need to show that

$$z \notin \frac{1}{W((A-\tau I)^{-1})} + \tau$$

for some choice of $\tau \in \mathbb{C} \setminus \Lambda(A)$. We see this is true by letting $\tau = z$, because otherwise we would have

$$0 \in \frac{1}{W((A-zI)^{-1})},$$

which contradicts the fact that $W((A-zI)^{-1})$ is a bounded set. □

From the proof of Theorem 2 we already see that the set (4) never includes τ itself. Indeed, we have the following proposition.

Proposition 3 *If* $\|\cdot\|$ *is any subordinate norm, then*

$$\operatorname{dist}\left(\tau,\ \frac{1}{W((A-\tau I)^{-1})}+\tau\right) \geq \|(A-\tau I)^{-1}\|^{-1}.$$

Proof: This follows from the fact that the set $W((A-\tau I)^{-1})$ is inside the disk around zero with radius $\|(A-\tau I)^{-1}\|$. Note that for the two-norm we have $\|(A-\tau I)^{-1}\|_2^{-1} = \sigma_{\min}(A-\tau I)$, where $\sigma_{\min}$ indicates the minimal singular value. □

This implies that, once we already have an eigenvalue inclusion region, we can exclude a neighborhood of any $\tau \notin \Lambda(A)$, and thereby improve the inclusion region, by taking the intersection of the region with $1/W((A-\tau I)^{-1})+\tau$.

Moreover, inspecting the proof of Theorem 2, we observe that the only property of the field of values that we use is the fact that it is a bounded set that contains the eigenvalues of the matrix. Realizing this, we immediately arrive at the following theorem.

Theorem 4 *Let* G *be a set-valued function from the set of complex* $n \times n$ *matrices to subsets of* $\mathbb{C}$, *such that for any* A *the set* $G(A)$ *is bounded and contains* $\Lambda(A)$. *Then*

$$\Lambda(A) = \bigcap_{\tau \in \mathbb{C} \setminus \Lambda(A)} \frac{1}{G((A-\tau I)^{-1})} + \tau.$$

This result will be used later on in the paper.

We now study properties of the set (4) for varying τ. The next, somewhat surprising, result shows that if $|\tau| \to \infty$, the inclusion region (4) converges to $W(A)$.

Theorem 5

$$\lim_{|\tau|\to\infty} \left(\frac{1}{W((A-\tau I)^{-1})}+\tau\right) = W(A).$$

Proof: With Proposition 1 and $\eta = \tau^{-1}$, the result follows from

$$\left\{\lim_{|\tau|\to\infty} \frac{y^*(A-\tau I)^*Ay}{y^*(A-\tau I)^*y} : y \neq 0\right\} = \left\{\lim_{\eta\to 0} \frac{y^*(\eta A - I)^*Ay}{y^*(\eta A - I)^*y} : y \neq 0\right\}$$
$$= \left\{\frac{y^*Ay}{y^*y} : y \neq 0\right\} = W(A).$$

□

We remark that in the case that A is a normal matrix this theorem has the following geometric interpretation. Since the field of values is unitarily invariant, we may assume that A is a diagonal matrix and hence $(A - \tau I)^{-1} = \mathrm{diag}((a_{ii} - \tau)^{-1})$. The field of values of a normal matrix is the convex hull of its eigenvalues [6, p. 11], which in this case are its diagonal entries, and this implies that (4) is a *circular-arc polygon* with vertices at $a_{11}, \ldots, a_{nn}$. (By a circular-arc polygon we mean a closed set in the complex plane with boundary consisting of up to n circular arcs; here, the intersection of any two of these arcs is one of the eigenvalues $a_{11}, \ldots, a_{nn}$.) By choosing τ large enough the circular arcs connecting the eigenvalues get arbitrarily close to the straight line segments connecting the eigenvalues as we can see as follows. Suppose that L is a line through $(a_{ii} - \tau)^{-1}$ and $(a_{jj} - \tau)^{-1}$. Then, $\frac{1}{L} + \tau$ (interpreted elementwise as before) is the circle passing through a_{ii}, a_{jj}, and τ. Notice that as $|\tau| \to \infty$, the radius of this circle also approaches infinity, and so the circular arc approaches the straight line through a_{ii} and a_{jj}.

Kippenhahn [8] (see also [19]) showed that the field of values of an $n \times n$ matrix is the convex hull of an algebraic curve of class n called the *boundary generating curve* and that the foci of this curve are the eigenvalues of the matrix. (The *class* of a curve is the degree of the polynomial $P(u, v)$ whose zeroes give the lines $ux + vy = 1$ which are tangent to the curve. We can also characterize the class of a curve as the number of tangents—real or imaginary—that can be drawn from a point to the curve. The notion of algebraic foci generalizes the familiar definition for conics to algebraic curves of arbitrary degree. See, for example, [2, Section 3] for a recent discussion.) Taking the reciprocal of the boundary generating curve of $W((A - \tau I)^{-1})$ and adding τ gives another algebraic curve such that the eigenvalues of A are still foci of the curve. Therefore in the special case where the boundary of the field of values of the matrix A is the same as its boundary generating curve, the entire family of curves formed by the boundaries of the sets (4) is confocal, i.e., every curve in this family has the same foci.

Theorems 2 and 5 imply that the inclusion set (1) is actually a rather restricted spectral inclusion set where in (5) we only take the intersection of the sets (4) for $\tau = 0$ and $\tau = \infty$. Moreover, apparently the inclusion set $W(A)$ can be "simulated" by the set (4) for $|\tau| \to \infty$.

We observe that although $W((A-\tau I)^{-1})$ is a bounded set, the set (4) may or may not be bounded, depending on whether or not the origin is contained in $W((A-\tau I)^{-1})$. If a neighborhood of 0 is inside $W(A-\tau I)$ then $1/W((A-\tau I)^{-1})+\tau$ is an unbounded eigenvalue inclusion region with a bounded complement. For this reason, it may be more convenient to think of the bounded complement of $1/W((A-\tau I)^{-1})+\tau$ as an *eigenvalue exclusion region*. We will illustrate this with an example in Figure 1. It is suggestive here to speak of the boundary of (4) as an *inclusion* or *exclusion curve*, depending on whether it is the boundary of a bounded inclusion or exclusion region.

We will illustrate the results with the 300×300 `randcolu` matrix of MATLAB's `gallery`. In Figure 1, two targets are taken that are inside the field of values, so that the set $1/W((A-\tau I)^{-1})+\tau$ is an unbounded inclusion region; its complement can be seen as a bounded exclusion region.

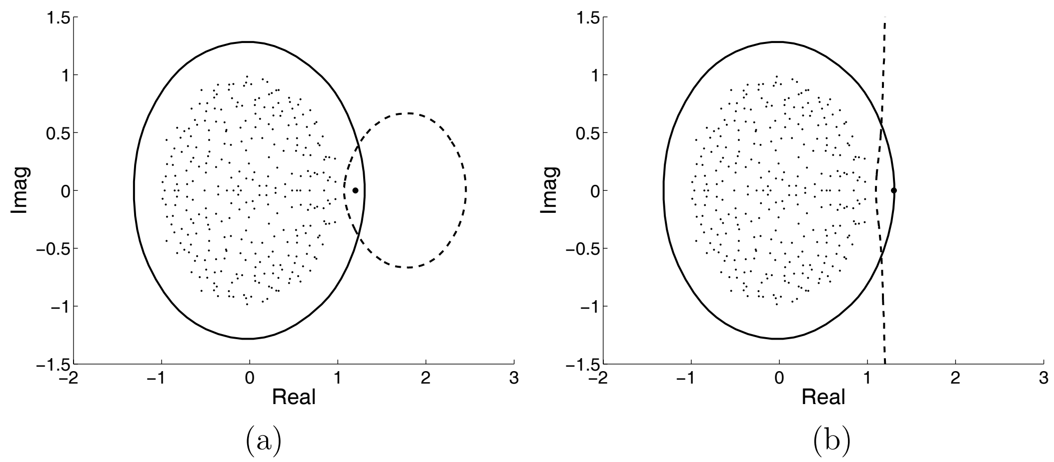
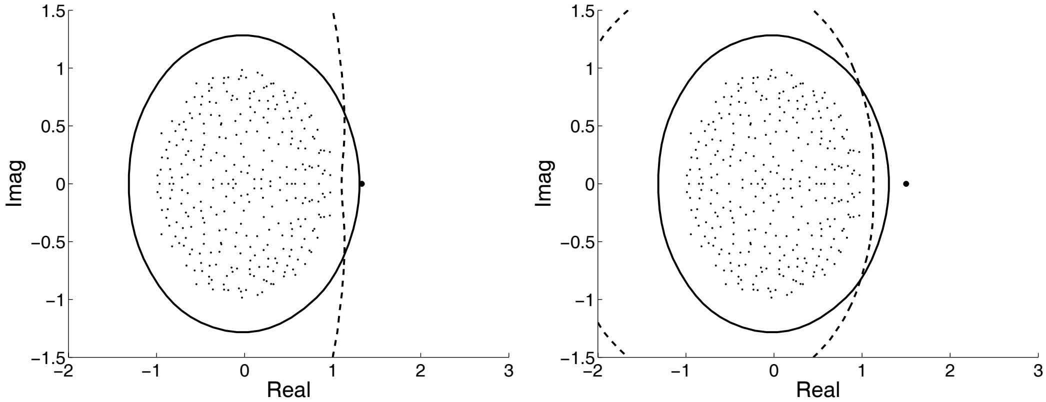


Fig. 1. (a) Spectrum, $W(A)$ (solid) and $1/W((A-\tau I)^{-1})+\tau$ (dot) of the 300×300 `randcolu` matrix for the shift $\tau=1.2$, also indicated by an asterisk. The dotted line is the boundary of the unbounded inclusion region $1/W((A-\tau I)^{-1})+\tau$ and its bounded complement which forms an exclusion region. (b) Idem but now for $\tau=1.3$, which is still inside $W(A)$. As a consequence, the inclusion region is still unbounded.

For Figure 2, two targets are taken that are outside the field of values, so that the set $1/W((A-\tau I)^{-1})+\tau$ is a bounded inclusion region. In Figure 2(b) we start to see convergence to $W(A)$.

Since the field of values contains the spectrum, a neighborhood of the origin will often be in $W(A-\tau I)$ if we choose τ sufficiently close to an eigenvalue, in which case (4) is unbounded. (Note that this is not always the case, for instance if τ is located close to an eigenvalue but just outside the field of values of a normal matrix.) On the other hand, if $|\tau|$ is large enough, then according to Theorem 5 the set (4) is a bounded inclusion region.



(a) (b)

Fig. 2. (a) Same as in Figure 1 but now for shift $\tau = 1.33$, which is outside $W(A)$, which results in a bounded inclusion region. (b) Idem, but now for $\tau = 1.5$.

Let us now consider the transition case, in which τ is such that the origin is on the boundary of $W((A - \tau I)^{-1})$. This means that $1/W((A - \tau I)^{-1}) + \tau$ is an unbounded inclusion region while the complement is also unbounded. In this case the boundary of (4) passes through infinity, and is neither an inclusion curve nor an exclusion curve, so in this case we will call the boundary of (4) a *transition curve* of A. The next result deals with the question of which values of τ give rise to a transition curve.

Theorem 6 *Let $\tau \notin \Lambda(A)$. The boundary of* (4) *is a transition curve of A if and only if τ is on the boundary of $W(A)$.*

Proof: Since τ is not an eigenvalue and

$$W((A - \tau I)^{-1}) = \left\{ \frac{x^*(A - \tau I)^* x}{x^*(A - \tau I)^*(A - \tau I)\, x} : x \neq 0 \right\}$$

we have that $0 \in W((A - \tau I)^{-1})$ if and only if there exists a nonzero x such that $x^*(A - \tau I)^* x = 0$, which means that $\tau \in W(A)$. □

In the next section, we will generalize the above results by substituting for the field of values another well-known eigenvalue inclusion region, the Gershgorin region. Before proceeding, however, we make one final geometric observation for the case of the field of values.

Suppose ξ is an algebraic curve which is the boundary of the set (4), and that the target τ is in the interior of the field of values so that ξ is an exclusion curve. Such a curve is known to have interesting analytic properties; in particular, we know from [9, Thm. 1.1] that if ξ is an eigenvalue exclusion curve, then it is the boundary of a *quadrature domain* $\mathcal{D}$. That is, there exist finitely many points $z_k = x_k + iy_k \in \mathcal{D}$ and constants $\gamma_1, \ldots, \gamma_r$, such that the following quadrature identity holds for functions $h(x,y)$ which are harmonic on the closure of $\mathcal{D}$: $\iint_{\mathcal{D}} h(x,y)dxdy = \sum_{k=1}^{r} \gamma_k h(x_k, y_k)$. In the simplest nontrivial case, where A is a 2×2 matrix, the boundaries of the sets (4) are confocal ellipses when they are inclusion curves, and inverted ellipses when they are exclusion curves. The inverted ellipse, also known as Neumann's quadrature domain, is the earliest historical example of a curve satisfying a (two-point) quadrature identity. Note, incidentally, that the two points z_1 and z_2 in this quadrature identity are *not* the eigenvalues of A; see [3] for more information.

4 Eigenvalue inclusion regions from the Gershgorin and Brauer regions of inverses of shifted matrices

The union of Gershgorin disks is another type of famous eigenvalue inclusion region. Richard Varga has proved many beautiful Gershgorin results, culminating in the book [17]. In this section, we show that similar results to some of those derived for the field of values in the previous section can be obtained for Gershgorin regions.

With $1 \leq i \leq n$ and

$$r_i(A) := \sum_{j \neq i} |a_{ij}|,$$

recall that

$$\Gamma_i(A) := \{z \in \mathbb{C} : |z - a_{ii}| \leq r_i(A)\}$$

is the ith *Gershgorin disk* of A. Now define the *Gershgorin region* to be the union of the Gershgorin disks, i.e.,

$$\Gamma(A) := \bigcup_{1 \leq i \leq n} \Gamma_i(A).$$

It is well known that the Gershgorin region is an eigenvalue inclusion region: since each eigenvalue λ is element of at least one disk, we have $\lambda \in \Gamma(A)$ (see, for example, [17]).

Now for a shifted matrix A we consider the Gershgorin region of the inverse: we will study the sets

$$\frac{1}{\Gamma_i((A - \tau I)^{-1})} + \tau \tag{6}$$

and

$$\frac{1}{\Gamma((A - \tau I)^{-1})} + \tau. \tag{7}$$

Similar to Theorem 2 for the field of values, we have the following result.

Theorem 7

$$\Lambda(A) = \bigcap_{\tau \in \mathbb{C} \setminus \Lambda(A)} \frac{1}{\Gamma((A-\tau I)^{-1})} + \tau.$$

Proof: This follows immediately from Theorem 4. □

Analogous to Theorem 5, we have the following result.

Theorem 8

$$\lim_{|\tau|\to\infty} \left(\frac{1}{\Gamma_i((A-\tau I)^{-1})} + \tau \right) = \Gamma_i(A)$$

and, consequently,

$$\lim_{|\tau|\to\infty} \left(\frac{1}{\Gamma((A-\tau I)^{-1})} + \tau \right) = \Gamma(A).$$

Proof: The set (6) consists of exactly the z for which

$$\left| \frac{1}{z-\tau} - ((A-\tau I)^{-1})_{ii} \right| \leq \sum_{i \neq j} |((A-\tau I)^{-1})_{ij}|.$$

Write $\eta = \tau^{-1}$, then we consider the situation $\eta \to 0$. Since $(A-\tau I)^{-1} = \eta\,(\eta A - I)^{-1}$ the inequality becomes

$$\left| \frac{\tau}{z-\tau} - ((\eta A - I)^{-1})_{ii} \right| \leq \sum_{i \neq j} |((\eta A - I)^{-1})_{ij}|$$

or

$$\left| \frac{1}{\eta z - 1} - ((\eta A - I)^{-1})_{ii} \right| \leq \sum_{i \neq j} |((\eta A - I)^{-1})_{ij}|.$$

Expanding in Taylor series we get

$$\left| (I + \eta A + \mathcal{O}(\eta^2))_{ii} - (1 + \eta z + \mathcal{O}(\eta^2)) \right| \leq \sum_{i \neq j} |(I + \eta A + \mathcal{O}(\eta^2))_{ij}|.$$

Since $I_{ii} = 1$ and $I_{ij} = 0$ for $j \neq i$, we may write

$$\left| (\eta A + \mathcal{O}(\eta^2))_{ii} - (\eta z + \mathcal{O}(\eta^2)) \right| \leq \sum_{i \neq j} |(\eta A + \mathcal{O}(\eta^2))_{ij}|.$$

This yields

$$|(a_{ii} - z + \mathcal{O}(\eta))| \leq \sum_{i \neq j} |(a_{ij} + \mathcal{O}(\eta))|,$$

so that in the limit $\eta \to 0$ we get

$$|z - a_{ii}| \leq \sum_{i \neq j} |a_{ij}|.$$

The second statement of the theorem follows straightforwardly. □

We note that another proof appeared in the Ph.D. thesis of the third author [18]. Since the reciprocal of a circle is a circle (or line), the boundary of each of the sets (6) is a circle (or line). When $|\tau|$ is large enough, we know from the last theorem that (6) is a disk converging to $\Gamma_i(A)$, so that for large values of τ the union of the sets (6) is an eigenvalue inclusion region. The concept of transition curves of the previous subsection does not carry over in a direct manner to the Gershgorin region since this set is a union of n disks, each containing at least one eigenvalue; in general there will not be a τ so that all n sets are unbounded.

There are several generalizations of the Gershgorin region that yield other eigenvalue inclusion regions. We next consider the arguably best-known generalization, the Brauer region. The region

$$K_{i,j}(A) := \{z \in \mathbb{C} : |z - a_{ii}| \cdot |z - a_{jj}| \leq r_i(A) \cdot r_j(A)\}$$

is called the (i, j)-th *Brauer Cassini oval* of A. The *Brauer region* is defined to be the union of the Brauer Cassini ovals, i.e.,

$$K(A) := \bigcup_{i \neq j} K_{i,j}(A).$$

The Brauer region is at least as good an eigenvalue inclusion region as the Gershgorin region: $\Lambda(A) \subseteq K(A) \subseteq \Gamma(A)$; see, for example, [17].

By once again applying Theorem 4 we arrive at another characterization of the spectrum.

Theorem 9

$$\Lambda(A) = \bigcap_{\tau \notin \Lambda(A)} \frac{1}{K((A - \tau I)^{-1})} + \tau.$$

We also get an analogue of Theorem 8, which can be proven in an almost identical way.

Theorem 10

$$\lim_{|\tau|\to\infty}\left(\frac{1}{K_{i,j}((A-\tau I)^{-1})}+\tau\right)=K_{i,j}(A)$$

and, consequently,

$$\lim_{|\tau|\to\infty}\left(\frac{1}{K((A-\tau I)^{-1})}+\tau\right)=K(A).$$

Similar to Proposition 3, the sets $1/\Gamma((A-\tau I)^{-1})$ and $1/K((A-\tau I)^{-1})$ avoid τ, as the following result shows.

Proposition 11

$$\text{dist}\left(\tau,\ \frac{1}{K((A-\tau I)^{-1})}+\tau\right)\geq \text{dist}\left(\tau,\ \frac{1}{\Gamma((A-\tau I)^{-1})}+\tau\right)\geq \|(A-\tau I)^{-1}\|_\infty^{-1}.$$

Proof: This follows from $K((A-\tau I)^{-1})\subseteq\Gamma((A-\tau I)^{-1})$ and the fact that for all $z\in\Gamma((A-\tau I)^{-1})$ we have $|z|\leq\|(A-\tau I)^{-1}\|_\infty$. □

We believe that the techniques employed so far may also be useful in studying other generalizations of the Gershgorin region (see [17]), but rather than proceed in this direction, we turn now to pseudospectra.

5 Eigenvalue inclusion regions from pseudospectra of inverses of shifted matrices

The ε-pseudospectra of A, defined by

$$\Lambda_\varepsilon(A)=\{z:\sigma_{\min}(A-zI)\leq\varepsilon\}$$

are often studied to better understand the behavior of nonnormal matrices; see [15] for a recent overview.

By applying Theorem 4 we have immediately the following result.

Theorem 12

$$\Lambda(A)=\bigcap_{\tau\notin\Lambda(A)}\frac{1}{\Lambda_\varepsilon((A-\tau I)^{-1})}+\tau.$$

Instead of an exact analogue of Theorem 5, we get the following result.

Theorem 13

$$\lim_{\tau\to\infty}\left(\frac{1}{\Lambda_{\frac{\varepsilon}{|\tau|^2}}((A-\tau I)^{-1})}+\tau\right)=\Lambda_\varepsilon(A).$$

Proof: The set

$$\frac{1}{\Lambda_{\frac{\varepsilon}{|\tau|^2}}((A-\tau I)^{-1})}+\tau$$

consists of the z for which

$$\sigma_{\min}\left(\frac{1}{z-\tau}\,I-(A-\tau I)^{-1}\right)\le\frac{\varepsilon}{|\tau|^2}.$$

Again with $\eta=\tau^{-1}$ we get

$$\sigma_{\min}\left(\frac{\eta}{\eta z-1}\,I-\eta\,(\eta A-I)^{-1}\right)\le\varepsilon\,|\eta|^2$$

Expansion in Taylor series yields

$$\sigma_{\min}\left(\eta(I+\eta A+\mathcal{O}(\eta^2))-\eta(1+\eta z+\mathcal{O}(\eta^2))\,I\right)\le\varepsilon\,|\eta|^2$$

which means

$$|\eta|^2\,\sigma_{\min}(A-zI+\mathcal{O}(\eta))\le\varepsilon\,|\eta|^2.$$

For $\eta\to 0$ this yields $\Lambda_\varepsilon(A)$. □

6 Subspace approximations for large matrices

In this and the following section we will show that the obtained results yield practical methods to approximate inclusion regions for large matrices. Since the computation of the inverse of a (shifted) large matrix will often be prohibitively expensive, we will consider approaches to numerically approximate the sets (4) and (7) using subspace methods. This section studies the case for one shift and partly contains an extension of the methods of [5] for shifted matrices. We will use the approximations in Section 7 for a method giving an approximate eigenvalue inclusion region using several suitably chosen shifts.

As already stated in Section 2, an important property of harmonic Ritz values θ with respect to target τ is that they tend to stay away from this τ. Although this behavior is known from practical situations, we are not aware of any concrete results showing this. The next proposition gives a result in this direction. As in Section 2, let $\mathcal{U}$ be a search space with associated matrix U of which the columns form an orthonormal basis for $\mathcal{U}$. We make the natural assumption that $(A - \tau I)\,U$ is of full rank; if this is not the case, τ is an eigenvalue and its corresponding eigenvector is in the space $\mathcal{U}$—a fortunate event.

Proposition 14 *Let $(A-\tau I)\,U$ be of full rank, let $U^*(A-\tau I)^*(A-\tau I)\,U = LL^*$ be the Choleski decomposition (or matrix square root decomposition), and let $\|\cdot\|$ be any subordinate norm. Then*

$$|\theta - \tau| \geq \|L^{-1}U^*(A-\tau I)^*UL^{-*}\|^{-1}.$$

Proof: Since $u \in \mathcal{U}$, we can write $u = Uc$ for $c \in \mathbb{C}^k$. Therefore, we have

$$(A-\tau I)\,u - (\theta-\tau)\,u \perp (A-\tau I)\,\mathcal{U}$$

and, via

$$U^*(A-\tau I)^*(A-\tau I)\,Uc = (\theta-\tau)\,U^*(A-\tau I)^*Uc, \tag{8}$$

this is equivalent to

$$L^{-1}U^*(A-\tau I)^*\,UL^{-*}d = (\theta-\tau)^{-1}d,$$

where $d = L^*c$. The result now follows from the fact that the spectral radius is bounded above by a subordinate matrix norm. □

A field of values may be determined numerically by the method due to Johnson [7], but this may be expensive for large matrices. In this situation, Manteuffel and Starke [11] proposed to use the Arnoldi process to approximate both $W(A)$ and $1/W(A^{-1})$ as follows. Starting from an initial vector u_1 with unit norm, let

$$AU_k = U_kH_k + h_{k+1,k}u_{k+1}e_k^* = U_{k+1}\underline{H}_k,$$

be the Arnoldi decomposition after k steps, where the columns of U_k form an orthonormal basis for the Krylov space $\mathcal{U}_k$ with u_1 as its first column, H_k is an upper Hessenberg matrix, e_k is the kth canonical basis vector, and $\underline{H}_k = \begin{bmatrix} H_k \\ h_{k+1,k}e_k^* \end{bmatrix}$ is a $(k+1)\times k$ Hessenberg matrix with an extra row.

In [11], $W(A)$ is approximated by

$$W(A) \supseteq W(U_k^*AU_k) = W(H_k). \tag{9}$$

For the approximation of $1/W(A^{-1})$, we first introduce the reduced QR-decomposition $\underline{H}_k = Q_k R_k$, so that $AU_k R_k^{-1} = U_{k+1}\underline{H}_k R_k^{-1} = U_{k+1}Q_k$ has orthonormal columns. Then $W(A^{-1})$ is approximated in [11] as follows:

$$W(A^{-1}) \supseteq W(R_k^{-*}U_k^* A^* A^{-1} A U_k R_k^{-1}) = W(R_k^{-*} H_k^* R_k^{-1}). \tag{10}$$

(In fact, this is the derivation in [5]; [11] used a different one.)

As noted in [5,11], while approximation (9) for $W(A)$ is often very reasonable, approximation (10) for $W(A^{-1})$ is frequently disappointing. In practical situations, approximation (10) may not contain the eigenvalues of A^{-1}, so that $1/W(A^{-1})$ does not contain $\Lambda(A)$. This implies that this numerical approximation of (1) does not contain the spectrum.

For this reason, in [5] the approximation

$$W(A^{-1}) \supseteq W(U_k^* A^{-1} U_k) \approx W(H_k^{-1})$$

was introduced which no longer guarantees a strict inclusion but in practice may be a much better approximation to $W(A^{-1})$.

We now discuss the numerical approximation of $W((A-\tau I)^{-1})$ for large matrices using subspace techniques. We present two alternative approximations which are (relatively straightforward) extensions of the approaches in [5].

For the first approximation, let $\underline{I}_k$ be the $k \times k$-identity with an extra $(k+1)$st zero row. With the reduced QR-decomposition $\underline{H}_k - \tau \underline{I}_k = Q_k R_k$, the matrix $(A-\tau I)\,U_k R_k^{-1}$ has orthonormal columns, and we therefore have

$$\begin{aligned} W((A-\tau I)^{-1}) &\supseteq W(R_k^{-*}U_k^*(A-\tau I)^*(A-\tau I)^{-1}(A-\tau I)\,U_k R_k^{-1}) \\ &= W(R_k^{-*}(H_k - \tau I_k)^* R_k^{-1}). \end{aligned} \tag{11}$$

Recall from (8) that the eigenvalues of

$$(U_k^*(A-\tau I)^* U_k)^{-1} U_k^*(A-\tau I)^*(A-\tau I)\,U_k = (H_k - \tau I_k)^{-*} R_k^* R_k$$

are the harmonic Ritz values shifted by $-\tau$ of A with respect to search space $\mathcal{U}_k$ and shift τ. Since the eigenvalues of $R_k^{-*}(H_k - \tau I_k)^* R_k^{-1}$ are identical to those of $R_k^{-1}R_k^{-*}(H_k - \tau I_k)^*$, we conclude that after k steps we know that (4) contains the convex hull of the harmonic Ritz values with respect to shift τ.

A second approach to approximate $W((A-\tau I)^{-1})$ is to discard the last term in the expression

$$U_k^*(A-\tau I)^{-1}U_k = (H_k - \tau I)^{-1} - h_{k+1,k} U_k^*(A-\tau I)^{-1} u_{k+1} e_k^* (H_k - \tau I)^{-1}.$$

and approximate

$$W((A-\tau I)^{-1}) \supseteq W(U_k^*(A-\tau I)^{-1}U_k) \approx W((H_k-\tau I)^{-1}). \qquad (12)$$

This approximation is not an inclusion in general but may be satisfactory provided that $\|(A-\tau I)^{-1}\|_2$ and $\|(H-\tau I)^{-1}\|_2$ are not too large. Indeed, experiments in [5] indicate that (12) is often much more favorable than (11); we will use (12) in the next subsection.

Finally, to numerically approximate $\Gamma((A-\tau I)^{-1})$ and $K((A-\tau I)^{-1})$ using the Krylov subspace $\mathcal{U}_k$, we can approximate $(A-\tau I)^{-1} \approx U_k(H_k-\tau I)^{-1}U_k^*$. The inspiration for this is the following. For arbitrary $w \in \mathbb{C}^n$, $v = (A-\tau I)^{-1}w$ can be approximated from the subspace $\mathcal{U}_k$ by

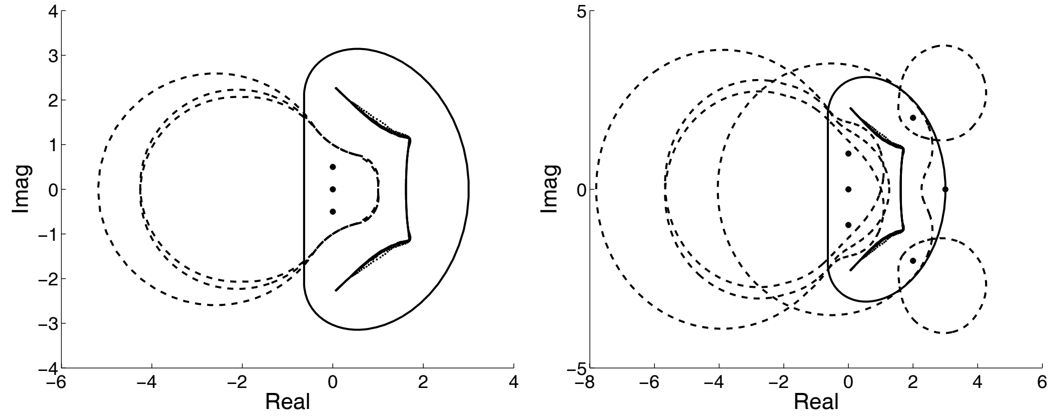
$$v \approx v_k = U_k c, \qquad w - (A-\tau I)\, v_k \perp \mathcal{U}_k,$$

so that $v_k = U_k c = U_k(H-\tau I)^{-1}U_k^* w$. The elements $((A-\tau I)^{-1})_{ij}$ that occur in $\Gamma((A-\tau I)^{-1})$ and $K((A-\tau I)^{-1})$ can then be efficiently approximated by $(e_i^*U_k)((H_k-\tau I)^{-1}(U_k^*e_j))$. We will give examples of the mentioned approaches in the next section.

7 Numerical examples and an algorithm

Before we will propose an algorithm to get an inclusion region based on the field of values, we first perform a number of experiments with the 1000×1000 `grcar` matrix. In Figure 3(a) the spectrum (dots), the inclusion region $W(A)$ (solid line), and the unbounded inclusion regions $1/W((A-\tau I)^{-1})+\tau$ (dotted line) are indicated for the targets $\tau = 0$, and $\tau = \pm i/2$. The “repelling force” of the targets resulting in exclusion regions containing the targets—the bounded complements of the inclusion regions $1/W((A-\tau I)^{-1})+\tau$—is clear. Note that the intersections of the four inclusion sets gives a much better inclusion region than $\Lambda \subseteq W(A)$ alone, but also that the regions derived with the targets $\tau = \pm i/2$ add little to the final result.

For Figure 3(b) we plot *approximations* to the inclusion regions (4) for several values of τ: 0, $\pm i$, $2 \pm 2i$, and 3. We generate a 10-dimensional Krylov space $\mathcal{K}(A, b)$ for a random vector b and take the approximation (12), which for $\tau = 0$ was found to be often better than (11) in [5]. We see for this example that as long as τ is not too close to an eigenvalue, the approximation (12) is indeed an inclusion region. All bounded regions with dotted lines as boundaries are exclusion regions, being bounded complements of unbounded inclusion regions, except for the region corresponding to $\tau = 3$, which is a bounded inclusion region. Without further details, we mention that the approximations (11) with



(a) (b)

Fig. 3. (a) Spectrum, $W(A)$ (solid) and $1/W((A-\tau I)^{-1})+\tau$ (dotted) of the 1000×1000 `grcar` matrix and targets $\tau = 0, \pm i/2$. The targets are indicated by dots. (b) The same but here the sets $1/W((A-\tau I)^{-1})+\tau$ (dot) are numerically *approximated* by (12) using a 10-dimensional Krylov space. We use the targets $\tau = 0, \pm i, 2 \pm 2i$ and 3.

the same targets were all "inclusion regions" of poor quality: they included not all eigenvalues, or even none.

Now we come to the important question how to choose sensible targets τ, given a matrix A. Since $W(A)$ is an inclusion region, it is natural to take τ on or near to the boundary of the field of values to get more restrictive inclusion regions. We already have seen an example of this in Figure 3.

Based on this idea, we now develop an automatic procedure to get an (approximate) eigenvalue inclusion region from a low-dimensional Krylov space. First we generate a low, k-dimensional Krylov space $\mathcal{K}_k(A, u_1)$ for a starting vector u_1, yielding a Krylov relation $AU_k = U_k H_k + h_{k+1,k} u_{k+1} e_k^*$. We approximate $W(A)$ by $W(H_k)$, where, in turn, we approximate $W(H_k)$ by m "angles" (that is, by multiplying H_k by factors of the form $e^{-i\alpha_j}$, $j = 1, \ldots, m$; see [7]). A low number m between, say, 8 and 32 often already gives excellent results. This process gives $2m$ points τ_j (every angle determines two points: a minimal and maximal eigenvalue of a "rotated" matrix [7]) that are on the boundary of $W(H_k)$, inside $W(A)$, and often also close to the boundary of $W(A)$.

Given these points τ_j, $j = 1, \ldots, 2m$, the sets $1/W((A-\tau_j I)^{-1})+\tau_j$ are unbounded inclusion regions of which the bounded complements are exclusion regions since the τ_j are inside $W(A)$ (see Section 3). Since $W((A-\tau_j I)^{-1})$ is computationally expensive to determine, we approximate $1/W((A-\tau_j I)^{-1})+\tau_j$ by the sets $1/W((H_k-\tau_j I_k)^{-1})+\tau_j$. We stress the fact that this is an efficient procedure since we use the *same low-dimensional Krylov space* to approximate $W(A)$ by $W(H_k)$ and $1/W((A-\tau_j I)^{-1})+\tau_j$ by $1/W((H_k-\tau_j I_k)^{-1})+\tau_j$. As discussed in the previous section, these approximations are not guaranteed to be true inclusion regions, but may be promising inclusion regions in practical situations. As a result, this procedure gives an approximate inclusion region as the intersection of $W(H_k)$ and $\bigcap_j 1/W((H_k - \tau_j I_k)^{-1}) + \tau_j$. In Algorithm 1 we summarize the algorithm.

Algorithm 1 A method to determine an (approximate) inclusion region based on fields of values and a Krylov space.

Input: A matrix A, starting vector u_1, dimension of Krylov space k, and number of angles m

Output: An (approximate) spectral inclusion region

1: Compute the Krylov space $\mathcal{K}_k(A, u_1)$
  with associated relation $AU_k = U_k H_k + h_{k+1,k} u_{k+1} e_k^*$

2: Approximate $W(A)$ by $W(H_k)$ using m angles, giving $2m$ points $\tau_j \in \mathbb{C}$
  that are inside and usually near the boundary of $W(A)$

for $j = 1 : 2m$

3: Approximate $1/W((A - \tau_j I_n)^{-1}) + \tau_j$ by $1/W((H_k - \tau_j I_k)^{-1}) + \tau_j$
  (using a low number of angles for $W((H_k - \tau_j I_k)^{-1})$)

end

4: The (approximate) spectral inclusion region is given by the intersection
  of $W(H_k)$ and $\bigcap_{j=1}^{2m} 1/W((H_k - \tau_j I_k)^{-1}) + \tau_j$

The results of Algorithm 1 for $k = 10$ (dimension of the Krylov subspace) and $m = 8$ (number of “angles” to approximate $W(H_k)$) are shown in Figure 4. Although there are many curves, the relevant region is the interior of the field of values, where the intersection with the inclusion sets $1/W((H_k - \tau_j I)^{-1}) + \tau_j$ is seen to give a tighter eigenvalue inclusion region than $W(H_k)$ alone.

Finally, we perform some experiments with inclusion regions based on Gershgorin disks. In Figure 5(a) we take the same matrix; this time we plot the spectrum (dots), and the inclusion regions defined by the Gershgorin region $\Gamma(A)$ (solid line) and $1/\Gamma((A - \tau I)^{-1}) + \tau$ (dotted lines) for $\tau = 0, \quad \pm i/2$, $\pm i$. We see that the plotted inclusion regions based on the Gershgorin regions seem less promising than those based on the field of values.

For Figure 5(b) we plot *approximations* to the inclusion regions (7) for $\tau = 0$, $\pm i/2$, $\pm i$, $2 \pm 2i$, and 3 using the same 10-dimensional Krylov subspace as before and the approximation proposed in the previous section. Clearly, the inclusion regions based on Gershgorin disks are not as informative as those based on the fields of values in Figures 3 and 4.

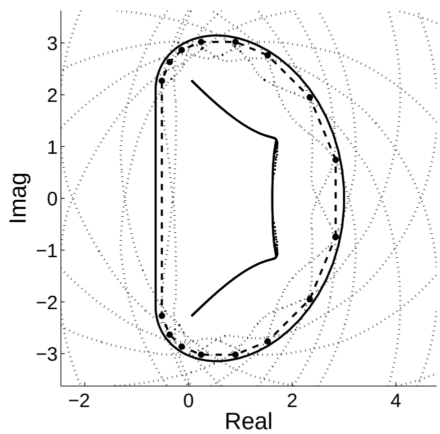


Fig. 4. Spectrum inside $W(A)$ (solid), $W(H_k)$ (dash), and inclusion sets $1/W((H_k - \tau_j I)^{-1}) + \tau_j$, $j = 1, \ldots, 16$ (dot) for the 1000×1000 `grcar` matrix and a $k = 10$ dimensional Krylov space. The targets τ_j are the points that are computed to approximate $W(H_k)$ and are indicated by dots.

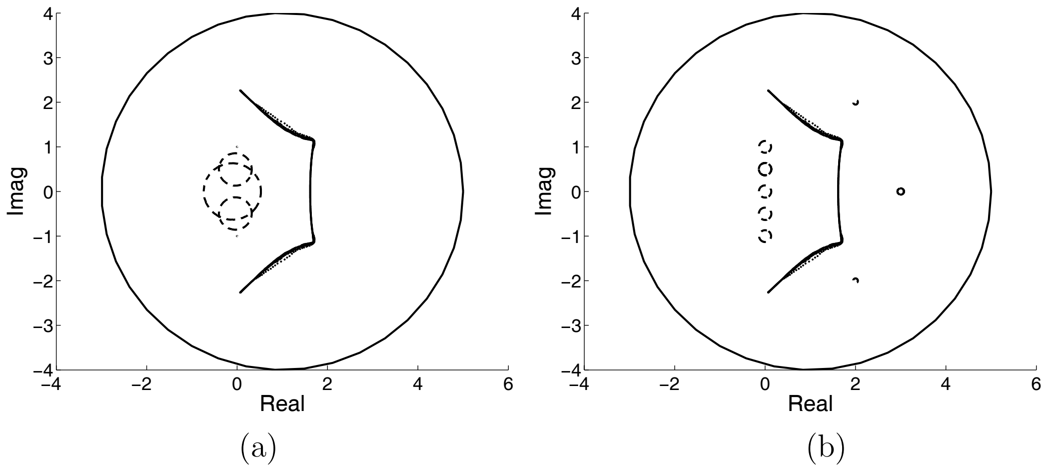


Fig. 5. (a) Spectrum inside $\Gamma(A)$ (solid) and $1/\Gamma((A - \tau I)^{-1}) + \tau$ (dot) of the 1000×1000 `grcar` matrix for shifts $\tau = 0, \ \pm i/2, \ \pm i$. Note that the exclusion regions for $\tau = \pm i$ are so tiny that they are hardly visible. (b) Spectrum, $\Gamma(A)$ (solid), and Krylov *approximations* to $1/\Gamma((A - \tau I)^{-1}) + \tau$ (dotted) for several values $\tau = 0$, $\pm i/2$, $\pm i$, $2 \pm 2i$, and 3.

8 Discussion and conclusions

We have studied eigenvalue inclusion regions derived from the field of values, pseudospectra, and Gershgorin and Brauer regions of the inverse of shifted versions of the matrix. By varying the shift we get a family of inclusion regions with surprising properties: the intersection of the family is exactly the spectrum, and an appropriate limit of the sets converges to the "mother set". In the case of the field of values, the boundaries of the inclusion regions which we called inclusion or exclusion curves form a confocal family. As a by-product, we also realized that the standard Rayleigh–Ritz method may be interpreted as a special case of harmonic Rayleigh–Ritz with target at infinity.

The emphasis of this paper is on the theoretical properties of the inclusion sets and relations with the harmonic Rayleigh–Ritz technique. In addition, in Sections 6 and 7 we have seen that the approaches may also be of practical value in determining (approximate) eigenvalue inclusion regions of large matrices via subspace approximation techniques.

In particular, we have proposed an automated procedure to get a spectral inclusion region based on a low-dimensional Krylov space. Although the resulting region is not guaranteed to be a true inclusion region, it may in practice give a much better inclusion region than the field of values alone.

Acknowledgements

This paper is dedicated with pleasure to Professor Richard Varga, with whom we have spent many very enjoyable hours. In particular, the second author is grateful for thirty years of friendship with Professor Varga. We thank the referees for very useful suggestions.

References

[1] C. Beattie and I. C. F. Ipsen, *Inclusion regions for matrix eigenvalues*, Linear Algebra Appl., 358 (2003), pp. 281–291.

[2] H.-L. Gau and P. Y. Wu, *Numerical range and Poncelet property*, Taiwanese J. Math., 7 (2003), pp. 173–193.

[3] B. Gustafsson and H. S. Shapiro, *What is a quadrature domain?*, in Quadrature domains and their applications, vol. 156 of Oper. Theory Adv. Appl., Birkhäuser, Basel, 2005, pp. 1–25.

[4] M. Hochbruck and M. E. Hochstenbach, *Subspace extraction for matrix functions*, Preprint, Dept. Math., Case Western Reserve University, September 2005. Submitted.

[5] M. E. Hochstenbach, D. A. Singer, and P. F. Zachlin, *On the field of values and pseudospectra of the inverse of a large matrix*, CASA report 07-06, Department of Mathematics, TU Eindhoven, The Netherlands, February 2007. Submitted.

[6] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[7] C. R. Johnson, *Numerical determination of the field of values of a general complex matrix*, SIAM J. Numer. Anal., 15 (1978), pp. 595–602.

[8] R. Kippenhahn, *Über den Wertevorrat einer Matrix*, Mathematische Nachrichten, 6 (1951), pp. 193–228.

[9] J. C. Langer and D. A. Singer, *Foci and foliations of real algebraic curves*, Milan J. Math., 75 (2007), pp. 225–271.

[10] T. A. Manteuffel and J. S. Otto, *On the roots of the orthogonal polynomials and residual polynomials associated with a conjugate gradient method*, Numer. Linear Algebra Appl., 1 (1994), pp. 449–475.

[11] T. A. Manteuffel and G. Starke, *On hybrid iterative methods for nonsymmetric systems of linear equations*, Numer. Math., 73 (1996), pp. 489–506.

[12] R. B. Morgan, *Computing interior eigenvalues of large matrices*, Linear Algebra Appl., 154/156 (1991), pp. 289–309.

[13] C. C. Paige, B. N. Parlett, and H. A. van der Vorst, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Num. Lin. Alg. Appl., 2 (1995), pp. 115–133.

[14] G. W. Stewart, *Matrix algorithms. Vol. II*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.

[15] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra*, Princeton University Press, Princeton, NJ, 2005.

[16] J. van den Eshof, A. Frommer, T. Lippert, K. Schilling, and H. A. van der Vorst, *Numerical methods for the QCD overlap operator: I. Sign-function and error bounds*, Comput. Phys. Comm., 146 (2002), pp. 203–224.

[17] R. S. Varga, *Geršgorin and his Circles*, vol. 36 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 2004.

[18] P. F. Zachlin, *On the Field of Values of the Inverse of a Matrix*, PhD thesis, Case Western Reserve University, 2007. Accessible via `www.ohiolink.edu/etd/view.cgi?case1181231690`.

[19] P. F. Zachlin and M. E. Hochstenbach, *On the field of values of a matrix*, Linear Multilinear Algebra, 56 (2008), pp. 185–225. English translation with comments and corrections of [7].