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 is 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 [7, 8] 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 Rayleigh–Ritz approach, mentioning some new results. 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 in Section 7 and give 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., [12] or also [3]. As was noted in [3], 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 *with target* $\tau = 0$ from a one-dimensional subspace, as we will now briefly review.

The harmonic Rayleigh–Ritz technique [9–11] 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.

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

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

□

By this we see that 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. We will use these sets for eigenvalue inclusion regions in the next section.

It is well-known from practical situations that harmonic Ritz values θ with respect to target τ tend to avoid this τ; however, we are not aware of any concrete results showing this. The next proposition gives a result in this direction. Note that the assumption that we will make that $(A-\tau I)\,U$ is of full rank is natural: if this is not the case, τ is an eigenvalue and its corresponding eigenvector is in the space $\mathcal{U}$—a fortunate event.

Proposition 2 *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*

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

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

$$U^*(A-\tau I)^*(A-\tau I)\,Uc = (\theta-\tau)\,U^*(A-\tau I)^* Uc, \tag{4}$$

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. $\square$

Because they tend to avoid a chosen target, 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 $z = 0$ [2, 13].

Another 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 6 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

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

where the addition in the right-hand side is interpreted as elementwise: for a set S, the set $S+\tau$ is defined as

$$S + \tau := \{z + \tau : z \in S\}.$$

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

$$\frac{1}{W((A-\tau I)^{-1})} + \tau \tag{5}$$

and therefore

$$\Lambda(A) \subseteq \bigcap_{\tau\in\mathbb{C}\setminus\Lambda(A)} \frac{1}{W((A-\tau I)^{-1})} + \tau. \tag{6}$$

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

Theorem 3

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

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 3 we already see that the set (5) never includes τ itself. Indeed, we have the following proposition.

Proposition 4 *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 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 3, 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 5 *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 (5) for varying τ. The next, somewhat surprising, result shows that if $|\tau| \to \infty$, the inclusion region (5) converges to $W(A)$.

Theorem 6

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

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

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

□

We remark that in the case that A is a normal matrix this theorem has the following elegant geometric interpretation. Since a field of values is unitarily invariant, we may assume that A is a diagonal matrix and hence $(A-\tau I)^{-1} = \mathrm{diag}((a_{ii}-\tau)^{-1})$. Since the field of values of a normal matrix is the convex hull of its eigenvalues, which are its diagonal entries, we have that (5) is a circular-arc polygon with vertices at $a_{11}, \ldots, a_{nn}$. 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, $\frac{1}{L}+\tau$ 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}.

Theorems 3 and 6 imply that the inclusion set (1) is actually a rather restricted spectral inclusion set where in (6) we only take the intersection of the sets (5) for $\tau = 0$ and $\tau = \infty$. Moreover, apparently the inclusion set $W(A)$ can be "simulated" by the set (5) for $|\tau| \to \infty$.

We observe that although $W((A-\tau I)^{-1})$ is a bounded set, the set (5) 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.

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 (5) 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 6 the set (5) is a bounded inclusion region.

We are now interested in 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 we will call the boundary of (5) a *transition curve* of A, since the boundary of (5), which passes through infinity, is a transition state between (5) defining a bounded and unbounded inclusion region for the spectrum. The next result deals with the question which values of τ give rise to a transition curve.

Theorem 7 *Let $\tau \notin \Lambda(A)$. The boundary of* (5) *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 $\tau \in W(A)$. $\square$

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.

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 $W(A)$.

In addition to an inclusion region, it is suggestive here to speak of its boundary as an *inclusion* or *exclusion curve*, depending on whether it is the boundary of a bounded inclusion or exclusion region. The transition curves then form the transition stage between an inclusion curve, which is the boundary of a bounded inclusion region, and an exclusion curve, which forms the boundary of a bounded exclusion region being the complement of an unbounded inclusion region. Such an exclusion curve ξ is known to have interesting analytic properties; in particular, we know from [6, Thm. 1.1] that if ξ is an eigenvalue exclusion curve, then it bounds a “quadrature domain”.

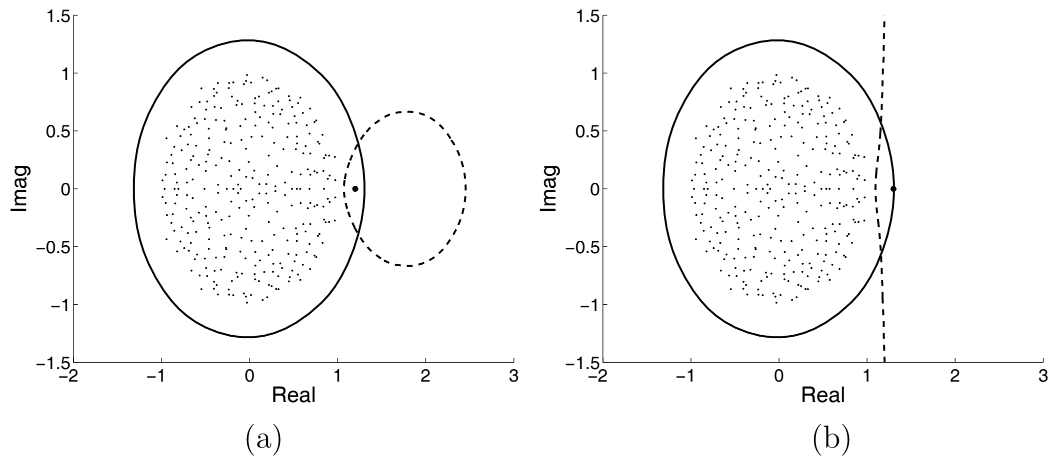


Fig. 1. (a) Spectrum, $W(A)$ (solid) and $1/W((A-\tau I)^{-1})+\tau$ (dot) of the 300×300 `randcolu` matrix for the shift $\tau = 1.2$, also indicated by an asterisk. The dotted line is the boundary of the unbounded inclusion region $1/W((A-\tau I)^{-1})+\tau$ and its bounded complement which forms an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

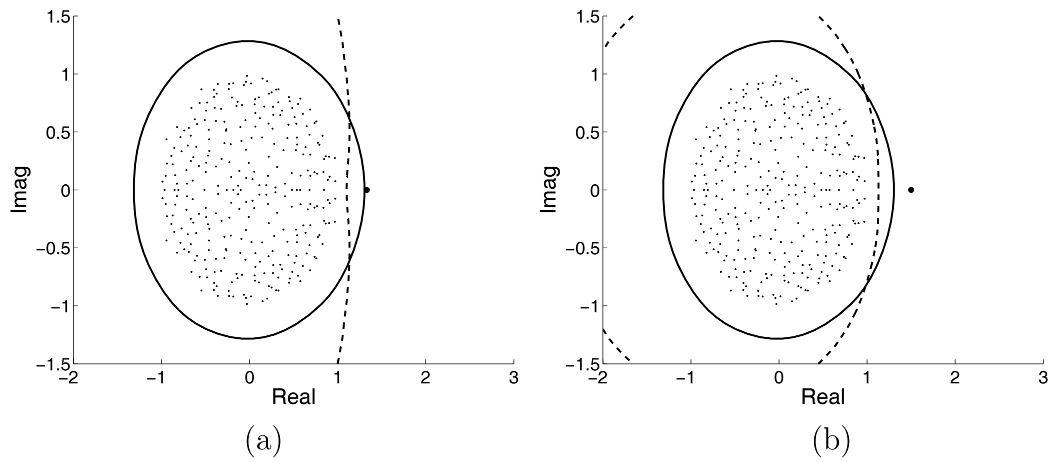


Fig. 2. (a) Same as in Figure 1 but now for shift $\tau = 1.33$, which is outside $W(A)$, which results in a bounded inclusion region. (b) Idem, but now for $\tau = 1.5$.

Suppose A is such that the boundary of $W(A)$ does not contain any line segments. Kippenhahn [5] (see also [16]) showed that the boundary of the field of values of such a matrix is an algebraic curve of class n and that its foci are the eigenvalues. Taking the reciprocal of the boundary curve of $W((A-\tau I)^{-1})$ and adding τ gives another algebraic curve such that the eigenvalues of A are foci of the curve. The conclusion is that in this case where we consider inclusion regions defined by fields of values, the entire family of curves formed by the boundaries of the sets (5) is confocal.

In the next section, we will turn our attention to Gershgorin regions.

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 [14]. 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, [14]).

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

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

and

$$\frac{1}{\Gamma((A - \tau I)^{-1})} + \tau. \tag{8}$$

Similar to Theorem 3 for the field of values, we have the following result.

Theorem 8

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

Proof: This follows immediately from Theorem 5. □

Analogous to Theorem 6, we have the following result.

Theorem 9

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

Proof: The set (7) consist 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 and discarding second and higher order terms gives

$$|1 + \eta z - (I + \eta A)_{ii}| \leq \sum_{i\neq j} |(I + \eta A)_{ij}|$$

but this just means

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

□

We note that another proof appeared in the Ph.D. thesis of the third author [15].

The concept of transition curves of the previous subsection does not carry over in a straightforward manner to the Gershgorin region since this set is a union of n disks. Since the reciprocal of a circle is a circle (or line), the boundary of each of the inclusion sets (7) is a circle (or line). When τ is large enough, we know from the last theorem that each of the sets (7) is an inclusion region, more specifically, a disk containing the eigenvalues, converging to the $\Gamma_i(A)$. If τ is close enough to an eigenvalue, then a neighborhood of the origin will often (but not always) be in one of the Gershgorin disks, in which case one of the sets (7) will be an unbounded eigenvalue inclusion region of which the complement is a disk. However, 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, [14].

By once again applying Theorem 5 we arrive at another characterization of the spectrum.

Theorem 10

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

We also get an analogue of Theorem 9, which can be proven in an almost identical way.

Theorem 11

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

Similar to Proposition 4, the sets $1/\Gamma((A-\tau I)^{-1})$ and $1/K((A-\tau I)^{-1})$ avoid τ, as the following result shows.

Proposition 12

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

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 [14]), 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

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

are often studied to better understand the behavior of nonnormal matrices; see [12] for a recent overview.

By applying Theorem 5 we have immediately the following result.

Theorem 13

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

Instead of an exact analogue of Theorem 6, we get the following result.

Theorem 14

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

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

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

Neglecting second and higher order terms yields

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

which for $|\tau| \to \infty$, $|\eta| \to 0$, gives $\Lambda_\varepsilon(A)$. $\square$

6 Subspace approximations for large matrices

The results obtained so far are mainly of theoretical interest, since the computation of the inverse of a (shifted) large matrix will often be prohibitively expensive in practice. In this section we will therefore consider approaches to numerically approximate the sets (5) and (8) using subspace methods.

A field of values may be determined numerically by the method due to Johnson [4], but this may be very expensive for large matrices. In this situation, Manteuffel and Starke [8] 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

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

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 [8], $W(A)$ is approximated by

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

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 [8] 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 [3]; [8] used a different one.)

As noted in [3, 8], 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 [3] 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 [3].

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

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

Recall from (4) 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 (5) contains the convex hull of the harmonic Ritz values with respect to shift τ.

The second approach 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

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

This approximation is not an inclusion but may be satisfactory provided that $\|(A-\tau I)^{-1}\|_2$ and $\|(H-\tau I)^{-1}\|_2$ are not too large.

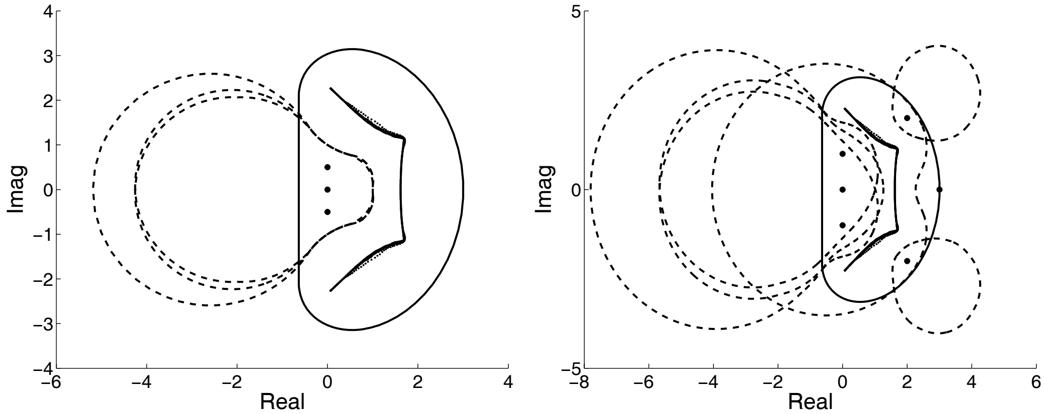
To numerically approximate $\Gamma((A-\tau I)^{-1})$ and $K((A-\tau I)^{-1})$ using the Krylov subspace $\mathcal{U}_k$, we can approximate $(A-\tau I)^{-1} \approx U_k (H_k - \tau I)^{-1} U_k^*$. The inspiration for this is the following. For arbitrary $w \in \mathbb{C}^n$, $v = (A-\tau I)^{-1} w$ can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

To further illustrate the results in this paper we 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.

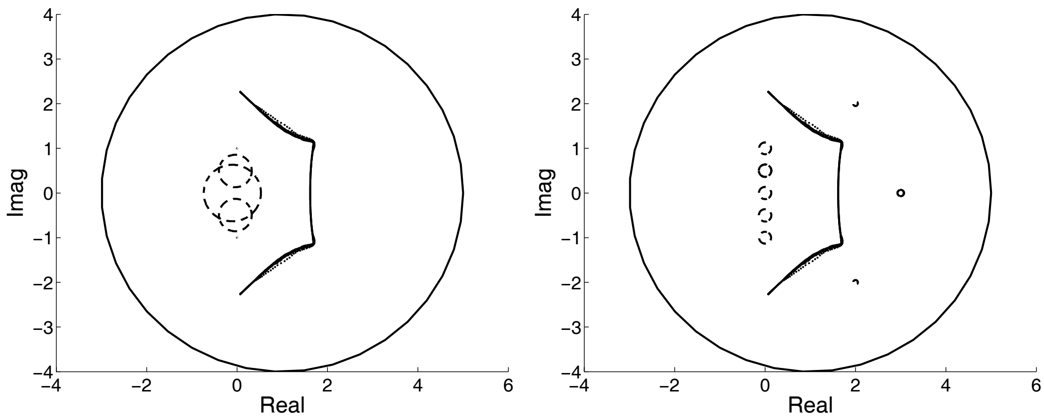


(a) (b)

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 an asterisk. (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.

For Figure 3(b) we plot *approximations* to the inclusion regions (5) 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 [3]. 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 a dotted line as boundary are inclusion regions, except for the region corresponding to $\tau = 3$, which is an exclusion region, being the bounded complement of an unbounded inclusion region. Without further details, we mention that the approximations (11) with the same targets were all "inclusion regions" of poor quality: they included not all eigenvalues, or even none.

In Figure 4(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 $\tau = 0, \;\pm i/2, \;\pm i$. We see that the plotted inclusion regions based on the Gershgorin regions seem less promising than those based on the field of values.



(a) (b)

Fig. 4. (a) Spectrum, $\Gamma(A)$ (solid) and $1/\Gamma((A-\tau I)^{-1})+\tau$ (dot) of the 1000×1000 `grcar` matrix for shifts $\tau = 0, \quad \pm i/2, \quad \pm i$. 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.

For Figure 4(b) we plot *approximations* to the inclusion regions (8) 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 exclusion regions are not as informative as those derived with the fields of values in Figure 4(b).

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.

The emphasis of this paper is on the theoretical properties of the inclusion sets and relations with the harmonic Rayleigh–Ritz technique. However, 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.

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.

References

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

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

[3] 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.

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

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

[6] J. C. Langer and D. A. Singer, *Foci and foliations of real algebraic curves*, Preprint, Department of Mathematics, Case Western Reserve University, July 2007. Submitted.

[7] 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.

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

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

[10] 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.

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

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

[13] 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.

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

[15] 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`.

[16] P. F. Zachlin and M. E. Hochstenbach, *On the field of values of a matrix*, CASA report 07-06, Department of Mathematics, TU Eindhoven, The Netherlands, January 2007. Revised June 2007. To appear in Linear and Multilinear Algebra. English translation with comments and corrections of [5].