NUMERICAL APPROXIMATION OF THE FIELD OF VALUES OF THE INVERSE OF A LARGE MATRIX

MICHIEL E. HOCHSTENBACH, DAVID A. SINGER, AND PAUL F. ZACHLIN

Dedicated to Prof. Bebiano

ABSTRACT. We consider the approximation of the field of values of the inverse of a large sparse matrix, without explicitly computing the inverse or using its action (i.e., accurately solving a linear system with this matrix). We review results by Manteuffel and Starke and give an alternative that may yield better approximations in practice. We give connections with the harmonic Rayleigh–Ritz approach. Several properties and applications of the studied concepts as well as numerical examples are provided.

1. INTRODUCTION

Let A be a large sparse nonsingular $n \times n$ matrix. The field of values

$$W(A) = \{\boldsymbol{x}^* A \boldsymbol{x} : \|\boldsymbol{x}\| = 1\},$$

where $\|\cdot\|$ denotes the 2-norm, is a convex set in the complex plane that is of interest for many problems and applications. For instance, it is useful to obtain (approximate) eigenvalue inclusion regions for large matrices (see, e.g., [6, 4]).

In this paper we are interested in the numerical approximation of $W(A^{-1})$, the field of values of the inverse of A, without computing or approximating A^{-1}. In Section 2 we first focus on the numerical approximation of $W(A)$, both for small and large matrices. Section 3 reviews some applications where knowledge of $W(A^{-1})$ may be useful. We give some relations between $W(A)$ and $W(A^{-1})$ in Section 4.

2000 *Mathematics Subject Classification.* 65F15, 65F50, 65F35, 47A12, 65F30, 65F10.

Key words and phrases. Field of values, numerical range, matrix inverse, large sparse matrix, Ritz values, harmonic Rayleigh–Ritz, harmonic Ritz values, GMRES convergence, Arnoldi, numerical radius, numerical abscissa, inner numerical radius, inclusion region.

The first author was supported by a Vidi research grant from the Netherlands Organization for Scientific Research (NWO).

In Section 5 we propose two approximation methods for $W(A^{-1})$. We review an approach by Manteuffel and Starke and give relations with the harmonic Rayleigh–Ritz method. We also present an alternative approach that comes with no inclusion guarantee, but may work better in practice. We provide some analysis and numerical examples. We end with some conclusions in Section 6.

2. Krylov approximations to the field of values

An attractive aspect of $W(A)$ is the availability of efficient numerical methods to approximate this set, both for small and large matrices. For small matrices, Johnson [7] pointed out that $W(A)$ may be efficiently approximated by computing the maximal and minimal eigenvalue of the Hermitian part of $e^{i\alpha}A$:

$$\tfrac{1}{2}\left(e^{i\alpha}A + (e^{i\alpha}A)^*\right)$$

for a number of angles $\alpha \in [0, \pi)$. Hereby the equalities

$$\begin{aligned} \max_{z\in W(A)} \mathrm{Re}(e^{-i\alpha}z) &= \tfrac{1}{2}\,\lambda_{\max}(e^{i\alpha}A + (e^{i\alpha}A)^*), \\ \min_{z\in W(A)} \mathrm{Re}(e^{-i\alpha}z) &= \tfrac{1}{2}\,\lambda_{\min}(e^{i\alpha}A + (e^{i\alpha}A)^*) \end{aligned} \tag{2.1}$$

are used for every angle α. Here, Re denotes the real part of a complex number, and $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and smallest eigenvalue of a Hermitian matrix.

For matrices of large dimension, we may use the Lanczos method (see, e.g., [15]) to approximate the largest and smallest eigenvalue of the Hermitian matrices $e^{i\alpha}A + (e^{i\alpha}A)^*$. This method generates a low-dimensional Krylov subspace to approximate eigenpairs of large (sparse) matrices. The Lanczos method generally approximates the extremal eigenvalues well, particularly the largest and smallest eigenvalue. We may run a new Lanczos process for every α in a chosen discrete set; for each angle α the largest eigenpair will be approximated using a different Krylov subspace, generated by a different matrix of the form $e^{i\alpha}A + (e^{i\alpha}A)^*$ and an initial vector, for instance a random vector, or the approximate eigenvector for a previous value of α.

We would like to point out that there is a computationally less expensive alternative using only a *single* Krylov subspace for all angles α as follows. We may perform a single run of the Arnoldi process (see, e.g., [15]) on A (or on $e^{i\alpha}A$ for any fixed angle α) and an initial vector $\boldsymbol{u}_1$ of unit length (for instance a random vector). Let

$$\mathcal{U}_k = \mathcal{K}_k(A, \boldsymbol{u}_1) = \mathrm{span}(\boldsymbol{u}_1,\, A\boldsymbol{u}_1, \ldots, A^{k-1}\boldsymbol{u}_1)$$

be the Krylov space of dimension k generated by A and $\boldsymbol{u}_1$. Performing k steps of Arnoldi yields the decomposition

$$AU_k = U_k\, H_k + h_{k+1,k}\, \boldsymbol{u}_{k+1}\, \boldsymbol{e}_k^* = U_{k+1}\, \underline{H}_k, \tag{2.2}$$

where the columns of U_k form an orthonormal basis for $\mathcal{U}_k$ with $\boldsymbol{u}_1$ as its first column, H_k is an upper Hessenberg matrix, $\boldsymbol{e}_k$ is the kth canonical basis vector, and $\underline{H}_k = \begin{bmatrix} H_k \\ h_{k+1,k}\, \boldsymbol{e}_k^* \end{bmatrix}$ is a $(k+1) \times k$ upper Hessenberg matrix with an extra row.

As originally suggested by Manteuffel and Starke [8] one can approximate $W(A)$ by $W(U_k^* A U_k) = W(H_k)$. This approximation has the following pleasant monotonic inclusion property.

Proposition 2.1.

$$W(H_k) \subseteq W(H_{k+1}) \subseteq W(A).$$

PROOF. This follows from the observation

$$\begin{aligned} W(U_k^* A U_k) &= \{\boldsymbol{c}^* U_k^* A U_k \boldsymbol{c} : \boldsymbol{c} \in \mathbb{C}^k,\ \|\boldsymbol{c}\| = 1\} \\ &= \{\boldsymbol{x}^* A \boldsymbol{x} : \boldsymbol{x} \in \mathcal{U}_k,\ \|\boldsymbol{x}\| = 1\} \end{aligned}$$

and the fact that $\mathcal{U}_k \subset \mathcal{U}_{k+1}$. □

From this proposition it follows that the convex set $W(H_k) = W(U_k^* A U_k)$ may be interpreted as the field of values of A restricted to the Krylov subspace $\mathcal{U}_k$. In particular, we know that after k steps $W(H_k)$, and therefore also $W(A)$, contains the convex hull of the eigenvalues of H_k, which are called the Ritz values of A with respect to $\mathcal{U}_k$. Since k is assumed to be much smaller than n (the main idea of a subspace method), determining $W(H_k)$ is computationally very efficient, for instance by the method proposed by Johnson.

We note that this second approach via a single Arnoldi decomposition is in practice often very sensible. While a different search space per α may often give a (slightly) better result (that is, a larger maximal eigenvalue per angle and hence a larger approximate field of values which is still included in the true $W(A)$), both approaches may give nearly the same approximation, as for instance in the following example.

Example 2.1. As a first example, consider for A the 256×256 `grcar` matrix [9], which is banded upper Hessenberg; this example is also used in [8]. The eigenvalues (dots) and $W(A)$ are depicted in Figure 1(a). We approximate $W(A)$ by 10-dimensional Krylov subspaces, with a random initial vector $\boldsymbol{b}$. From Figure 1(b), we see that taking different Krylov subspaces $\mathcal{K}_k\left(\frac{1}{2}(e^{i\alpha}A + e^{-i\alpha}A^*), \boldsymbol{b}\right)$

for each of the 32 angles (dashed) instead of one Krylov subspace $\mathcal{K}_k(A, \boldsymbol{b})$ for all angles (solid) hardly improves the numerical approximation to $W(A)$, while the first approach is much more expensive.

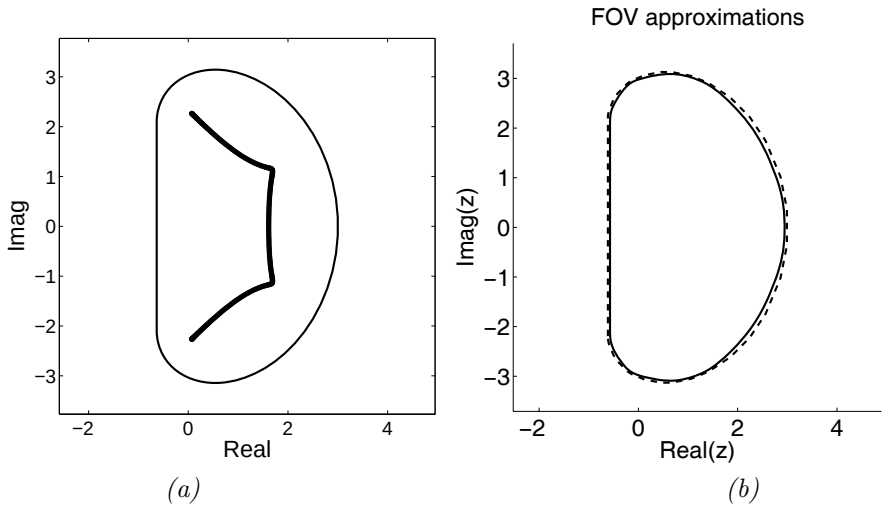

FIGURE 1. (a) Eigenvalues (dots) and $W(A)$ (solid) of the 256×256 `grcar` matrix. Approximations of $W(A)$ using different Krylov subspaces for each of the 32 angles (dashed), respectively one Krylov subspace for all angles (solid). All Krylov spaces have dimension $k = 10$.

In the following result, we see that the approximations may be even exactly the same for some combinations of matrices and initial vectors.

Proposition 2.2. *Let A be an $n \times n$ tridiagonal matrix such that the subdiagonal entries of A and $e^{i\alpha}A + (e^{i\alpha}A)^*$, for each $\alpha \in \{\alpha_1, \ldots, \alpha_m\}$, do not vanish. Suppose that $\boldsymbol{b} = \boldsymbol{e}_1$ is the initial vector of the Krylov space. Then the following two approximation methods to $W(A)$ produce exactly the same result.*

(a) *Run k steps of Arnoldi (2.2) on A and approximate $W(A) \approx W(H_k)$, where $W(H_k)$ on its turn is approximated by a set of angles $\{\alpha_1, \ldots, \alpha_m\}$.*
(b) *For each angle $\alpha \in \{\alpha_1, \ldots, \alpha_m\}$, run k steps of Lanczos on $\frac{1}{2}(e^{i\alpha}A + (e^{i\alpha}A)^*)$ resulting in $k \times k$ tridiagonal matrices $T_k(\alpha)$, determine the maximal and minimal eigenvalues $\lambda_{\max}(\alpha)$ and $\lambda_{\min}(\alpha)$ of the $T_k(\alpha)$,*

and approximate $W(A)$ by the intersection of all half planes of the form $\{z : \lambda_{\min}(\alpha) \le e^{-i\alpha}\,\mathrm{Re}(z) \le \lambda_{\max}(\alpha)\}$.

PROOF. The crucial observation is that, because of the assumptions of the nonzero subdiagonal elements, the Krylov spaces generated in both cases:

(a) $\mathcal{K}_k\left(\frac{1}{2}(e^{i\alpha}A + e^{-i\alpha}A^*), \boldsymbol{e}_1\right)$ for each α; and
(b) $\mathcal{K}_k(A, \boldsymbol{e}_1)$,

are all equal to $\mathrm{span}\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k\}$ and, consequently, the approximations to the field of values will be exactly the same in both approaches. □

In the remainder of this paper we concentrate on the field of values for the inverse of a large sparse invertible matrix A. Hereby, we do not want to compute A^{-1} explicitly, since this is often prohibitively expensive, but we also want to avoid the costs of accurately solving a linear system of the form $A\boldsymbol{x} = \boldsymbol{b}$. Instead, we will use a single Krylov space generated by A to approximate $W(A^{-1})$.

3. The field of values of the inverse of a matrix: applications

We review a few applications where estimates for the field of values of A^{-1} may be useful. First, we recall the definitions (see, e.g., [14]) of the *inner numerical radius*

$$\nu(A) = \min_{z \in W(A)} |z|,$$

the *numerical radius*

$$\mu(A) = \max_{z \in W(A)} |z|,$$

and the the *numerical abscissa*

$$\omega(A) = \max_{z \in W(A)} \mathrm{Re}(z).$$

From (2.1) we know that $\omega(A) = \frac{1}{2}\,\lambda_{\max}(A + A^*)$; this quantity is also called the *logarithmic norm* of A. We will now review some areas where it may be of interest to approximate the quantities $\nu(A^{-1})$, $\mu(A^{-1})$, or $\omega(A^{-1})$.

The first area we consider is the study of the convergence of iterative methods for linear systems. Consider the linear system $A\boldsymbol{x} = \boldsymbol{b}$ with $\boldsymbol{x}_0$ as initial guess, $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$ as corresponding initial residual, and the Krylov space $\mathcal{U}_k = \mathcal{K}(A, \boldsymbol{r}_0)$ as search space. In minimum residual methods, such as MINRES for symmetric A or GMRES for nonsymmetric A, one approximates $\boldsymbol{x}$ by

$$\boldsymbol{x} \approx \boldsymbol{x}_k = (AU_k)^+\boldsymbol{b} := (U_k^*A^*AU_k)^{-1}U_k^*A^*\boldsymbol{b},$$

where we assume that $A\mathcal{U}_k$ is of full rank k. This choice minimizes the residual over the search space: $\boldsymbol{x}_k = \operatorname{argmin}_{\boldsymbol{y}_k \in \mathcal{U}_k} \|\boldsymbol{b} - A\boldsymbol{y}_k\|$. Eiermann and Ernst [2, Section 6.1.3], [1, Cor. 6.2] show that for this minimum residual

$$\|\boldsymbol{r}_k^{\mathrm{MR}}\| := \|\boldsymbol{b} - A\boldsymbol{x}_k\|$$

we have the upper bound

$$\frac{\|\boldsymbol{r}_k^{\mathrm{MR}}\|}{\|\boldsymbol{r}_0\|} \leq \left(1 - \nu(A)\,\nu(A^{-1})\right)^{k/2}.$$

Note in particular that the right-hand side is independent of the choice of the initial guess and the resulting Krylov search space.

Moreover, we have the following property. Since $\boldsymbol{x}_k \in \mathcal{K}_k(A, \boldsymbol{b})$, we can write $x_k = q_{k-1}(A)\,\boldsymbol{b}$ for a certain polynomial q_{k-1} of degree $k-1$, and $\boldsymbol{r}_k = p_k(A)\,\boldsymbol{b}$ for a certain polynomial p_k of degree k. Then the zeros ζ of the *GMRES polynomial* p_k satisfy $\zeta^{-1} \in W(A^{-1})$, see [8] and also below.

In fact, these zeros are the *harmonic Ritz values* [10, 3]. The harmonic Rayleigh–Ritz approach, which was introduced in [10], is a popular tool for the numerical computation of interior eigenvalues close to a given target τ. Note that these eigenvalues are exterior eigenvalues of the shifted and inverted matrix $(A - \tau I)^{-1}$, where I denotes the identity matrix. For subspace methods it is often easier to compute exterior eigenvalues, as compared to interior eigenvalues. Given a search space $\mathcal{U}_k$ for the eigenvector, this suggests to determine (or extract) approximate eigenpairs $(\theta, \boldsymbol{u})$, where $\boldsymbol{u} \in \mathcal{U}_k$, by imposing the Galerkin condition

$$(A - \tau I)^{-1}\boldsymbol{u} - (\theta - \tau)^{-1}\boldsymbol{u} \perp \widetilde{\mathcal{U}}_k.$$

Here $\widetilde{\mathcal{U}}_k$ is a test space; to avoid working with the inverse of a large sparse matrix the harmonic Rayleigh–Ritz approach takes $\widetilde{\mathcal{U}}_k = (A-\tau I)^*(A-\tau I)\,\mathcal{U}_k$. In this case the subspace extraction is determined by the projected (and hence low-dimensional) generalized eigenvalue problem

$$U_k^*(A - \tau I)^*(A - \tau I)\,U_k\boldsymbol{c} = (\theta - \tau)\,U_k^*(A - \tau I)^*U_k\boldsymbol{c},$$

where we write $\boldsymbol{u} = U_k\boldsymbol{c}$, with $\boldsymbol{c} \in \mathbb{C}^k$. The pair $(\theta, \boldsymbol{u}) = (\theta, U_k\boldsymbol{c})$ is called a harmonic Ritz pair. For eigenpair approximations, one is interested in the pair with the harmonic Ritz value θ closest to τ, see, e.g., [13]. Given a harmonic Ritz vector $\boldsymbol{u}$, the corresponding harmonic Ritz value is given by

$$\frac{\boldsymbol{u}^*(A - \tau I)^*(A - \tau I)\,\boldsymbol{u}}{\boldsymbol{u}^*(A - \tau I)^*\boldsymbol{u}}.$$

The roots of the GMRES polynomial are the harmonic Ritz values with target $\tau = 0$ [10, 3]. By the equalities

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

we see that the inverses of these harmonic Ritz values are in $W(A^{-1})$. These equalities may also be interpreted as: $W(A^{-1})$ is the set of the inverses of all harmonic Ritz values that may be obtained after one step of the Arnoldi iteration (see [14, p. 166] for a similar statement about $W(A)$ and Ritz values). We will come back to the relation of harmonic Ritz values and the field of values in Section 5.

For another area of application, we consider the linear differential-algebraic equation with constant coefficients

$$A\boldsymbol{u}'(t) = \boldsymbol{u}(t), \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0.$$

If A is expensive to invert, this equation may be considered "effectively implicit" in the sense that it is computationally not easy to explicitly express $\boldsymbol{u}'(t)$ in terms of $\boldsymbol{u}(t)$. The numerical abscissa of A^{-1} determines the behavior of $\|e^{tA^{-1}}\|$ as $t \to 0$ (see, e.g., [14, (14.2)]):

$$\frac{d}{dt} \|e^{tA^{-1}}\|_{t=0} = \omega(A^{-1}).$$

In fact, we also have $\|e^{tA^{-1}}\| \leq e^{t\omega(A^{-1})}$ for all $t \geq 0$; see, e.g., [14, Thm. 17.1].

We can also use $W(A^{-1})$ for estimates for the norm of the matrix inverse and the condition number $\kappa(A) = \|A\| \, \|A^{-1}\|$. We have

$$\mu(A^{-1}) \leq \|A^{-1}\| \leq 2\,\mu(A^{-1})$$

(and a similar result for powers of A^{-1}, cf. [14, (17.6)]). Combined with an estimate for $\|A\|$, for instance using $\mu(A) \leq \|A\| \leq 2\,\mu(A)$, this gives an approximation for the condition number; see also the following section.

The field of values of A and of A^{-1} may also be used to determine inclusion regions for matrix eigenvalues. Particularly, Manteuffel and Starke [8] mention the inclusion region

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

Here $1/S$ for a subset S of the complex numbers is defined elementwise: $1/S = \{z^{-1} : z \in S\}$. This approach was extended to inverses of shifted matrices in [6] to give very useful and efficiently generated spectral inclusion regions.

Finally, we note that $W(A^{-1})$ also arises in the location of certain parameters in the recently developed IDR method [11].

4. Relations involving $\mu(A)$, $\nu(A)$ $\mu(A^{-1})$, and $\nu(A^{-1})$

Before considering the practical approximation of the field of values of A^{-1} in the next section, we look at some bounds for $\nu(A^{-1})$ and $\mu(A^{-1})$, and their relations with the quantities $\mu(A)$, $\nu(A)$, $\|A\|$, and $\|A^{-1}\|$. Eiermann and Ernst [1] note that

$$\nu(A^{-1}) = \min_{\boldsymbol{x}\neq\boldsymbol{0}} \left|\frac{\boldsymbol{x}^*A^{-1}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}}\right| = \min_{\boldsymbol{y}\neq\boldsymbol{0}} \left|\frac{\boldsymbol{y}^*A^*\boldsymbol{y}}{\boldsymbol{y}^*\boldsymbol{y}}\right| \cdot \left|\frac{\boldsymbol{y}^*\boldsymbol{y}}{\boldsymbol{y}^*A^*A\boldsymbol{y}}\right| \geq \frac{\nu(A)}{\|A\|^2}. \tag{4.1}$$

Using similar techniques we now derive new upper bounds for $\nu(A^{-1})$, and upper and lower bounds for $\mu(A^{-1})$.

Proposition 4.1. (a) $\dfrac{\nu(A)}{\|A\|^2} \leq \nu(A^{-1}) \leq \min\left\{\dfrac{\nu(A)}{\sigma_{\min}^2(A)}\,,\ \dfrac{\mu(A)}{\|A\|^2}\right\}.$

(b) $\max\left\{\dfrac{\nu(A)}{\sigma_{\min}^2(A)}\,,\ \dfrac{\mu(A)}{\|A\|^2}\right\} \leq \mu(A^{-1}) \leq \dfrac{\mu(A)}{\sigma_{\min}^2(A)}.$

Proof. With an eye on (4.1) we have

$$\nu(A^{-1}) \leq \min_{\boldsymbol{y}\neq\boldsymbol{0}} \left|\frac{\boldsymbol{y}^*A^*\boldsymbol{y}}{\boldsymbol{y}^*\boldsymbol{y}}\right| \cdot \max_{\boldsymbol{y}\neq\boldsymbol{0}} \left|\frac{\boldsymbol{y}^*\boldsymbol{y}}{\boldsymbol{y}^*A^*A\boldsymbol{y}}\right| = \frac{\nu(A)}{\sigma_{\min}^2(A)}.$$

Also,

$$\nu(A^{-1}) \leq \max_{\boldsymbol{y}\neq\boldsymbol{0}} \left|\frac{\boldsymbol{y}^*A^*\boldsymbol{y}}{\boldsymbol{y}^*\boldsymbol{y}}\right| \cdot \min_{\boldsymbol{y}\neq\boldsymbol{0}} \left|\frac{\boldsymbol{y}^*\boldsymbol{y}}{\boldsymbol{y}^*A^*A\boldsymbol{y}}\right| = \frac{\mu(A)}{\|A\|^2}.$$

Part (b) follows from

$$\mu(A^{-1}) = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \left|\frac{\boldsymbol{x}^*A^{-1}\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}}\right| \leq \max_{\boldsymbol{y}\neq\boldsymbol{0}} \left|\frac{\boldsymbol{y}^*A^*\boldsymbol{y}}{\boldsymbol{y}^*\boldsymbol{y}}\right| \cdot \max_{\boldsymbol{y}\neq\boldsymbol{0}} \left|\frac{\boldsymbol{y}^*\boldsymbol{y}}{\boldsymbol{y}^*A^*A\boldsymbol{y}}\right| = \frac{\mu(A)}{\sigma_{\min}^2(A)},$$

$$\mu(A^{-1}) \geq \min_{\boldsymbol{y}\neq\boldsymbol{0}} \left|\frac{\boldsymbol{y}^*A^*\boldsymbol{y}}{\boldsymbol{y}^*\boldsymbol{y}}\right| \cdot \max_{\boldsymbol{y}\neq\boldsymbol{0}} \left|\frac{\boldsymbol{y}^*\boldsymbol{y}}{\boldsymbol{y}^*A^*A\boldsymbol{y}}\right| = \frac{\nu(A)}{\sigma_{\min}^2(A)},$$

and

$$\mu(A^{-1}) \geq \max_{\boldsymbol{y}\neq\boldsymbol{0}} \left|\frac{\boldsymbol{y}^*A^*\boldsymbol{y}}{\boldsymbol{y}^*\boldsymbol{y}}\right| \cdot \min_{\boldsymbol{y}\neq\boldsymbol{0}} \left|\frac{\boldsymbol{y}^*\boldsymbol{y}}{\boldsymbol{y}^*A^*A\boldsymbol{y}}\right| = \frac{\mu(A)}{\|A\|^2}.$$

□

In particular, we have from this

$$\|A\| \leq (\nu(A^{-1}))^{-1}\,\frac{\mu(A)}{\|A\|} \leq (\nu(A^{-1}))^{-1}$$

and therefore $\|A\| \leq \min\{2\,\mu(A),\ \nu(A^{-1})^{-1}\}$.

From a different perspective the bounds in this subsection, together with bounds such as $\mu(A^{-1}) \leq \|A^{-1}\| \leq 2\,\mu(A^{-1})$, imply bounds for $\|A\|$, $\|A^{-1}\|$ and the condition number $\kappa(A) = \|A\|\,\|A^{-1}\|$ in terms of $W(A)$ and $W(A^{-1})$.

5. Two numerical approximation methods for $W(A^{-1})$

To approximate the field of values of A^{-1}, note that we have

$$W(A^{-1}) \supseteq W(\widehat{U}_k^* A^{-1} \widehat{U}_k)$$

for any $n \times k$ matrix $\widehat{U}_k$ with orthonormal columns that form a basis for a space $\widehat{\mathcal{U}}_k$. To avoid matrix inversion of a large matrix, one idea is to take $\widehat{\mathcal{U}}_k = A\mathcal{U}_k$. In what follows we will assume that this space is of full dimension k. We can compute the resulting approximation efficiently as follows. If $\underline{H}_k = Q_k R_k$ is the reduced QR-decomposition of $\underline{H}_k$, then $AU_k R_k^{-1} = U_{k+1}\underline{H}_k R_k^{-1} = U_{k+1}Q_k$ has orthonormal columns, and

$$W(A^{-1}) \supseteq W(R_k^{-*} U_k^* A^* A^{-1} A U_k R_k^{-1}) = W(R_k^{-*} H_k^* R_k^{-1}). \tag{5.1}$$

This approximation is a true subset of $W(A^{-1})$ and behaves monotonically as function of k; cf. also Proposition 2.1.

Proposition 5.1. *Write $\widehat{H}_k = R_k^{-*} H_k^* R_k^{-1}$, and suppose that $\widehat{\mathcal{U}}_k \subset \widehat{\mathcal{U}}_{k+1}$. Then*

$$W(\widehat{H}_k) \subseteq W(\widehat{H}_{k+1}) \subseteq W(A^{-1}).$$

Proof. This follows from the fact that $W(\widehat{H}_k)$ is the restriction of $W(A^{-1})$ to the subspace $\widehat{\mathcal{U}}_k$. □

We note that Manteuffel and Starke [8] also suggest this approximation via a slightly different derivation. An advantage of expression (5.1) compared with [8] is that from (5.1) it is clear that this approximation to $W(A^{-1})$ is itself a field of values of a low-dimensional matrix $R_k^{-*} H_k^* R_k^{-1}$. (In [8], this approximation is described as the intersection of strips in the complex plane determined by the minimal and maximal eigenvalues of *generalized* eigenvalue problems.)

In the following we will assume that H_k is invertible. The eigenvalues of

$$\widetilde{H}_k := (U_k^* A^* U_k)^{-1} U_k^* A^* A U_k = H_k^{-*} R_k^* R_k$$

are the *harmonic Ritz values* of A with respect to search space $\mathcal{U}_k$ and target $\tau = 0$ (see [10] and Section 3). It is not difficult to prove that the eigenvalues of this matrix are the inverses of the eigenvalues of $R_k^{-*} H_k^* R_k^{-1}$. Since the eigenvalues of $\widetilde{H}_k$ are the harmonic Ritz values, our conclusion is that after

k steps we know that $W(A^{-1})$ contains the convex hull of the inverses of the harmonic Ritz values.

However, in spite of Proposition 5.1, $W(R_k^{-*}H_k^*R_k^{-1})$ may not always be an approximation to $W(A^{-1})$ of good quality. We now present a new, alternative, approach to approximate $W(A^{-1})$:

$$W(A^{-1}) \supseteq W(U_k^*A^{-1}U_k) \approx W(H_k^{-1}). \tag{5.2}$$

The key idea is now to approximate $W(A^{-1})$ using a Krylov subspace generated by A. In what is to follow we assume the Arnoldi decomposition (2.2). We prove some results on the quality of the approximation $W(A^{-1}) \approx W(H_k^{-1})$ in the following Proposition. We will see that this approximation may be superior to $W(\widehat{H}_k)$ in Example 5.1. Let $\mathrm{dist}(z,S)$ denote the distance from $z \in \mathbb{C}$ to a closed and bounded set $S \subset \mathbb{C}$: $\mathrm{dist}(z,S) = \min\{\, |z-s| \,:\, s \in S \,\}$.

Proposition 5.2. (a) *$W(H_k^{-1}) = W(U_k^*(A^{-1}+E_k)\,U_k)$, with*

$$E_k = h_{k+1,k}\,(A^{-1}\boldsymbol{u}_{k+1})\,(\boldsymbol{e}_k^*H_k^{-1}U_k^*). \tag{5.3}$$

(b) *$W(H_k^{-1}) = W(U_k^*A^{-1}(I_k+F_k)\,U_k)$, where $F_k = h_{k+1,k}\,\boldsymbol{u}_{k+1}\,(\boldsymbol{e}_k^*H_k^{-1}U_k^*)$.*

(c) *For all $z \in W(H_k^{-1})$ we have*

$$\mathrm{dist}(z,\ W(U_k^*A^{-1}U_k)) \le |h_{k+1,k}|\,\|U_k^*A^{-1}\boldsymbol{u}_{k+1}\|\,\|H_k^{-*}\boldsymbol{e}_k\|.$$

Proof. To start with, to derive (5.2) we rewrite (2.2):

$$A^{-1}U_k = U_kH_k^{-1} - h_{k+1,k}A^{-1}\boldsymbol{u}_{k+1}\boldsymbol{e}_k^*H_k^{-1},$$

so that

$$U_k^*A^{-1}U_k = H_k^{-1} - h_{k+1,k}U_k^*A^{-1}\boldsymbol{u}_{k+1}\boldsymbol{e}_k^*H_k^{-1}. \tag{5.4}$$

Discarding the last term on the right-hand side gives the approximation $U_k^*A^{-1}U_k \approx H_k^{-1}$. In fact, we see that

$$H_k^{-1} = U_k^*(A^{-1}+E_k)\,U_k = U_k^*A^{-1}(I_k+F_k)\,U_k,$$

which implies (a) and (b). Finally, from (5.4) we have

$$\boldsymbol{y}^*H_k^{-1}\boldsymbol{y} = \boldsymbol{y}^*(U_k^*A^{-1}U_k)\,\boldsymbol{y} + h_{k+1,k}\,(\boldsymbol{y}^*U_k^*A^{-1}\boldsymbol{u}_{k+1})\,(\boldsymbol{e}_k^*H_k^{-1}\boldsymbol{y})$$

from which (c) follows. □

Part (a) states that the field of values of H_k^{-1} is the true field of values of a perturbed matrix $A^{-1}+E_k = A^{-1}(I+F_k)$, restricted to the Krylov subspace $\mathcal{U}_k$. Here, the *backward error* E_k is a (hopefully small) rank-one update matrix. In particular $\|E_k\|$ will be small if $\mathcal{U}_k$ is almost an invariant subspace, which means that $|h_{k+1,k}|$ is small. Item (b) is similar but uses a multiplicative

(relative) perturbation. Part (c) means that every point of $W(H_k^{-1})$ is at most $|h_{k+1,k}|\,\|U_k^*A^{-1}\boldsymbol{u}_{k+1}\|\,\|H_k^{-*}\boldsymbol{e}_k\|$ away from $W(U_k^*A^{-1}U_k)$, a projected field of values of A^{-1} which is included in the sought set $W(A^{-1})$.

In summary, when we compare (5.1) with (5.2), the advantage of the former is that this approximation is a proper inclusion. However, a strength of (5.2) over (5.1) is that this method employs a projection that works solely with $\mathcal{U}_k$: (5.2) approximates a projection of A^{-1} onto the space $\mathcal{U}_k$, and not onto the space $A\mathcal{U}_k$. The space $A\mathcal{U}_k$ may be seen as biased, as by the multiplication with A the smallest eigenmodes will have been (greatly) reduced. Therefore, we expect that (5.2) may often yield a better approximation, especially if $|h_{k+1,k}|\,\|U_k^*A^{-1}\boldsymbol{u}_{k+1}\|\,\|H_k^{-*}\boldsymbol{e}_k\|$ is reasonably small.

We now continue with an illustrative numerical example.

Example 5.1. In Figure 2 the eigenvalues (dots) and field of values of A^{-1} (solid) are plotted, where A is the 256×256 `grcar` matrix. We take for $\mathcal{U}$ a 16-dimensional Krylov space generated using a random starting vector. We plot two approximations to $W(A^{-1})$: $W(R_k^{-*}H_k^*R_k^{-1})$ ((5.1), dash), and $W(H_k^{-1})$ ((5.2), solid). Although it is not guaranteed that $W(H_k^{-1})$ is a subset of $W(A^{-1})$, it is the case here; moreover, we see that $W(H_k^{-1})$ is a much better approximation to $W(A^{-1})$ than $W(R_k^{-*}H_k^*R_k^{-1})$. Indeed, $W(R_k^{-*}H_k^*R_k^{-1})$ contains only a small portion of the eigenvalues of A^{-1}, while $W(H_k^{-1})$ contains the entire spectrum of A^{-1}.

6. Concluding remarks

We have approximated the field of values of the inverse of a large sparse matrix by subspace methods, in particular the Arnoldi process. To this aim, we have introduced a viable alternative for a method proposed by Manteuffel and Starke [8].

The Arnoldi procedure can be used to give an approximation to the field of values of A *and* of its inverse. As the Ritz values of A with respect to search space $\mathcal{U}_k$ are always in $W(A)$, the inverses of the harmonic Ritz values are guaranteed to be in $W(A^{-*})$. As we can use the *Ritz matrix* $H_k = U_k^*AU_k$ to approximate $W(A)$, we can use the *harmonic Ritz matrix*, $\widetilde{H}_k = (U_k^*A^*U_k)^{-1}U_k^*A^*AU_k$, of which the harmonic Ritz values are the eigenvalues, to approximate $W(A^{-1})$. However, H_k^{-1}, an approximation to the projected inverse matrix $U_k^*A^{-1}U_k$, may give more promising results.

We note that in principle we can use any search space $\mathcal{U}_k$ in the approximations. For instance, we may use the Jacobi–Davidson method [12, 5] instead of the Arnoldi method. Of course, the analysis that makes use of (2.2) no

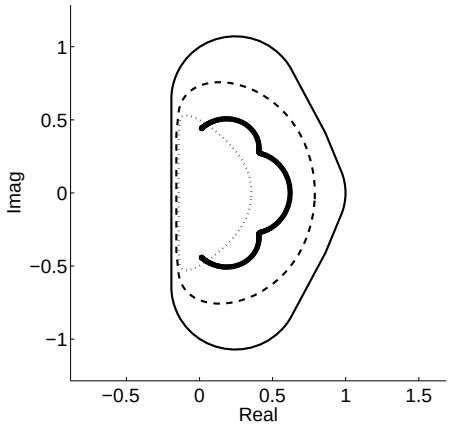

FIGURE 2. Data for A^{-1}: eigenvalues (dots), field of values $W(A^{-1})$ (solid), and the two approximations to $W(A^{-1})$: (5.1) ($W(A^{-1}|_{A\mathcal{U}})$, dotted) and (5.2) ($W(H_k^{-1})$, dash, which is an approximation to $W(A^{-1}|_{\mathcal{U}})$).

longer holds but without further details we mention that one can derive similar results using $U_k^* A U_k$ instead of H_k and the residual matrix $R_k = AU_k - U_k H_k$ instead of $h_{k+1,k}\boldsymbol{u}_{k+1}\boldsymbol{e}_k^*$. However, Jacobi–Davidson generally focuses on a selected region in the complex plane of interest, which implies that the resulting approximate field of values may globally be of lower quality. A strength of (unrestarted) Krylov methods is that they tend to well approximate the exterior eigenvalues simultaneously.

REFERENCES

[1] M. EIERMANN AND O. G. ERNST, *Geometric aspects of the theory of Krylov subspace methods*, Acta Numer., 10 (2001), pp. 251–312.

[2] O. G. ERNST, *Minimal and orthogonal residual methods and their generalizations for solving linear operator equations*. Habilitationsschrift, TU Bergakademie Freiberg, Germany, 2000.

[3] S. GOOSSENS AND D. ROOSE, *Ritz and harmonic Ritz values and the convergence of FOM and GMRES*, Numer. Linear Algebra Appl., 6 (1999), pp. 281–293.

[4] M. E. HOCHSTENBACH, *Fields of values and inclusion regions for matrix pencils*, Electron. Trans. Numer. Anal., 38 (2011), pp. 98–112.

[5] M. E. HOCHSTENBACH AND Y. NOTAY, *The Jacobi–Davidson method.*, GAMM Mitteilungen, 29 (2006), pp. 368–382.

[6] M. E. HOCHSTENBACH, D. A. SINGER, AND P. F. ZACHLIN, *Eigenvalue inclusion regions from inverses of shifted matrices*, Linear Algebra Appl., 429 (2008), pp. 2481–2496.

[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] T. A. Manteuffel and G. Starke, *On hybrid iterative methods for nonsymmetric systems of linear equations*, Numer. Math., 73 (1996), pp. 489–506.
[9] *The Matrix Market*. `http://math.nist.gov/MatrixMarket`, a repository for test matrices.
[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] V. Simoncini and D. B. Szyld, *Interpreting IDR as a Petrov-Galerkin method*, SIAM J. Sci. Comput., 32 (2010), pp. 1898–1912.
[12] G. L. G. Sleijpen and H. A. van der Vorst, *A Jacobi–Davidson iteration method for linear eigenvalue problems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 401–425.
[13] G. W. Stewart, *Matrix Algorithms. Vol. II*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.
[14] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra*, Princeton University Press, Princeton, NJ, 2005.
[15] H. A. van der Vorst, *Iterative Krylov Methods for Large Linear Systems*, vol. 13 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2003.

Department of Mathematics and Computer Science, Eindhoven University of Technology, PO Box 513, 5600 MB, The Netherlands, `www.win.tue.nl/∼hochsten/`

Department of Mathematics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-7058, USA.
E-mail address: `david.singer@case.edu`

Department of Mathematics, Lakeland Community College, 7700 Clocktower Drive, Kirtland, OH 44094-5198, USA.
E-mail address: `pzachlin@lakelandcc.edu`