Electronic Transactions on Numerical Analysis.
Volume 43, pp. 188-212, 2015.
Copyright © 2015, Kent State University.
ISSN 1068-9613.

ON THE LOCATION OF THE RITZ VALUES IN THE ARNOLDI PROCESS[*]

GÉRARD MEURANT[†]

Abstract. In this paper we give a necessary and sufficient condition for a set of complex values $\theta_1, \dots, \theta_k$ to be the Arnoldi Ritz values at iteration k for a general diagonalizable matrix A. Then we consider normal matrices and, in particular, real normal matrices with a real starting vector. We study in detail the case $k = 2$, for which we characterize the boundary of the region in the complex plane where pairs of complex conjugate Ritz values are located. Several examples with computations of the boundary of the feasible region are given. Finally we formulate some conjectures and open problems for the location of the Arnoldi Ritz values in the case $k > 2$ for real normal matrices.

AMS subject classifications. 65F15, 65F18, 15A18

Key words. Arnoldi algorithm, eigenvalues, Ritz values, normal matrices

1. Introduction.

Approximations to (a few of) the eigenvalues (and eigenvectors) of large sparse non-Hermitian matrices are often computed with (variants of) the Arnoldi process. One of the most popular software packages is ARPACK [13]. It is used, for instance, in the Matlab function `eigs`. It uses the Implicitly Restarted Arnoldi Method. In this paper we are concerned with the standard Arnoldi process, which, for a matrix A of order n and a starting vector v (assumed to be of unit norm), computes a unitary matrix V with columns v_i, where $v_1 = Ve_1 = v$, and an upper Hessenberg matrix H with positive real subdiagonal entries $h_{j+1,j}$, $j = 1, \dots, n-1$, such that

$$AV = VH,$$

if it does not stop before iteration n, a situation that we assume throughout this paper. The approximations of the eigenvalues of A (called the Ritz values) at step k are the eigenvalues $\theta_i^{(k)}$ of H_k, the leading principal submatrix of order k of H. The approximate eigenvectors are $x_i = V_{n,k} z_i^{(k)}$, where $z_i^{(k)}$ is the eigenvector associated with $\theta_i^{(k)}$ and $V_{n,k}$ is the matrix of the first k columns of V. In the sequel we will mainly consider the step k, so we will sometimes drop the superscript (k). The relation satisfied by $V_{n,k}$ is

$$AV_{n,k} = V_{n,k} H_k + h_{k+1,k} v_{k+1} e_k^T,$$

where e_k is the last column of the identity matrix of order k. This equation indicates how to compute the next column v_{k+1} of the matrix V and the kth column of H. When A is symmetric or Hermitian, the Arnoldi process reduces to the Lanczos algorithm, in which the matrix H is a symmetric tridiagonal matrix. There are many results on the convergence of the Lanczos Ritz values in the literature; see, for instance, [17, 18, 19, 21]. Most of them are based on the Cauchy interlacing theorem, which states that the Ritz values satisfy

$$\theta_1^{(k+1)} < \theta_1^{(k)} < \theta_2^{(k+1)} < \theta_2^{(k)} < \cdots < \theta_k^{(k)} < \theta_{k+1}^{(k+1)},$$

and they are related to the eigenvalues λ_j by

$$\lambda_j < \theta_j^{(k)}, \quad \theta_{k+1-j}^{(k)} < \lambda_{n+1-j}, \quad 1 \le j \le k.$$

[*] Received September 30, 2013. Accepted November 26, 2014. Published online on January 23, 2015. Recommended by H. Sadok.

[†] 30 rue du sergent Bauchat, 75012 Paris, France (`gerard.meurant@gmail.com`).

It is generally admitted that convergence of the Lanczos process for Hermitian matrices is well understood. Unfortunately in the non-Hermitian case, concerning the convergence of the Ritz values, this is not the case in general. However, some results are known about the eigenvectors; see, for instance, [1, 2]. In fact, the Arnoldi process may even not converge at all before the very last iteration. One can construct matrices with a given spectrum and starting vectors such that the Ritz values at all iterations are prescribed at an arbitrary location in the complex plane; see [9]. It means that we can construct examples for which the Ritz values *do not* converge to the eigenvalues of A before the last step.

However, the matrices that can be built using this result have generally poor mathematical properties. In particular, they are not normal. In many practical cases, we *do* observe convergence of the Ritz values toward the eigenvalues. For understanding the convergence when it occurs, an interesting problem is to know where the location of the Ritz values is for a given matrix, in particular, for matrices with special properties like (real) normal matrices. Of course, it is well known that they are inside the field of values of A, which is defined as

$$W(A) = \{\theta \,|\, \theta = v^* A v,\ v \in \mathbb{C}^n, \|v\| = 1\}.$$

If the matrix A is normal, the field of values is the convex hull of the eigenvalues, and if the matrix is real, it is symmetric with respect to the real axis.

The inverse problem described in Carden's thesis [4] and the paper [7] is, given a matrix A and complex values $\theta_1, \dots, \theta_k$, to know if there is a subspace of dimension k such that the values θ_i are the corresponding Ritz values. If we restrict ourselves to Krylov subspaces and the Arnoldi algorithm, this amounts to know if there is a unit vector v such that the values θ_i are Ritz values for the Krylov subspace

$$\mathcal{K}_k(A, v) = \text{span}\{v, Av, \dots, A^{k-1}v\}\,.$$

A closely related problem has been considered for normal matrices by Bujanović [3]. He was interested in knowing what the location of the other Ritz values is if one fixes some of the Ritz values in the field of values of A. He gave a necessary and sufficient condition that characterize the set of k complex values occurring as Ritz values of a given normal matrix. Carden and Hansen [7] also gave a condition that is equivalent to Bujanović's. For normal matrices and $k = n - 1$, see [14], and for general matrices, see [6].

In this paper we first give a necessary and sufficient condition for a set of complex values $\theta_1, \dots, \theta_k$ to be the Arnoldi Ritz values at iteration k for a given general diagonalizable matrix A. This generalizes Bujanović's condition. Then we restrict ourselves to real normal matrices and real starting vectors. We particularly study the case $k = 2$, for which we characterize the boundary of the region in the complex plane contained in $W(A)$, where pairs of complex conjugate Ritz values are located. We give several examples with computations of the boundary for real normal matrices of order up to 8. Finally, after describing some numerical experiments with real random starting vectors, we state some conjectures and open problems for $k > 2$ for real normal matrices in Section 7. The aim of this section, which provides only numerical results, is to motivate other researchers to look at these problems.

The paper is organized as follows. In Section 2 we study the matrices H_k and characterize the coefficients of their characteristic polynomial. Section 3 gives expressions for the entries of the matrix $M = K^*K$, where K is the Krylov matrix, as a function of the eigenvalues and eigenvectors of A for diagonalizable matrices. This is used in Section 4, where we give a necessary and sufficient condition for a set of k complex numbers to be the Arnoldi Ritz values at iteration k for diagonalizable matrices. The particular case of normal matrices is studied in Section 5. The case of A being real normal and $k = 2$ is considered in Section 6, in which we characterize the boundary of the region where pairs of complex conjugate Ritz values are located. Open problems and conjectures for $k > 2$ and real normal matrices are described in Section 7. Finally we give some conclusions.

2. The matrix H_k and the Ritz values.

In this section, since the Ritz values are the eigenvalues of H_k, we are interested in characterizing the matrix H_k and the coefficients of its characteristic polynomial. It is well known (see [9, 15]) that the matrix H can be written as $H = UCU^{-1}$, where U is a nonsingular upper triangular matrix such that $K = VU$ with $K = \begin{bmatrix} v & Av & \cdots & A^{n-1}v \end{bmatrix}$ and C is the companion matrix corresponding to the eigenvalues of A. We have the following theorem which characterizes H_k as a function of the entries of U.

THEOREM 2.1 ([10]). *For $k < n$, the Hessenberg matrix H_k can be written as $H_k = U_k C^{(k)} U_k^{-1}$, with U_k, the principal submatrix of order k of U, being upper triangular and $C^{(k)} = E_k + \begin{bmatrix} 0 & U_k^{-1} U_{[1:k],k+1} \end{bmatrix}$, a companion matrix where E_k is a square down-shift matrix of order k,*

$$E_k = \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ & \ddots & \ddots & & \\ & & 1 & 0 & \\ & & & 1 & 0 \end{bmatrix}.$$

Moreover, the subdiagonal entries of H are $h_{j+1,j} = \frac{U_{j+1,j+1}}{U_{j,j}}$, $j = 1, \dots, n-1$.

Clearly the Ritz values at step k are the eigenvalues of $C^{(k)}$. We see that they only depend on the matrix U and its inverse. They are also the roots of the monic polynomial defined below. By considering the inverse of an upper triangular matrix, we note that the last column of $C^{(k)}$ can be written as

$$U_k^{-1} U_{[1:k],k+1} = -U_{k+1,k+1} (U_{k+1}^{-1})_{[1:k],k+1} = -U_{k+1,k+1} (U^{-1})_{[1:k],k+1}.$$

Hence, up to a multiplying coefficient, the last column of $C^{(k)}$ is obtained from the k first components of the $(k+1)$st column of the inverse of U. The last column of $C^{(k)}$ gives the coefficients of the characteristic polynomial of H_k. Let

$$\begin{bmatrix} \beta_0^{(k)} \\ \vdots \\ \beta_{k-1}^{(k)} \end{bmatrix} = -U_k^{-1} U_{[1:k],k+1}.$$

The Ritz values are the roots of the polynomial $q_k(\lambda) = \lambda^k + \sum_{j=0}^{k-1} \beta_j^{(k)} \lambda^j = \prod_{i=1}^{k} (\lambda - \theta_i^{(k)})$. Since the entries of U and U^{-1} are intricate functions of the eigenvalues and eigenvectors of A, the following theorem provides a simpler characterization of the coefficients of the characteristic polynomial of H_k.

THEOREM 2.2. *Let $M = K^* K$, where K is the Krylov matrix. The vector of the coefficients of the characteristic polynomial of H_k, denoted as $\begin{bmatrix} \beta_0^{(k)}, \dots, \beta_{k-1}^{(k)} \end{bmatrix}$, is the solution of the linear system,*

$$M_k \begin{bmatrix} \beta_0^{(k)} \\ \vdots \\ \beta_{k-1}^{(k)} \end{bmatrix} = -M_{[1:k],k+1}, \tag{2.1}$$

where $M_k = U_k^* U_k$.

Proof. From what we have seen above, the proof is straightforward. We have

$$U_k \begin{bmatrix} \beta_0^{(k)} \\ \vdots \\ \beta_{k-1}^{(k)} \end{bmatrix} = -U_{[1:k],k+1}.$$

Multiplying by U_k^*, we obtain

$$M_k \begin{bmatrix} \beta_0^{(k)} \\ \vdots \\ \beta_{k-1}^{(k)} \end{bmatrix} = -U_k^* U_{[1:k],k+1}.$$

Clearly $U_k^* U_{[1:k],k+1} = M_{[1:k],k+1}$. ◻

Therefore it is interesting to consider the matrix $M = K^*K = U^*U$ and its principal submatrices. This is done in the next section.

3. The matrix M.

In this section we characterize the entries of $M = U^*U = K^*K$ as functions of the eigenvalues and eigenvectors of A and of the starting vector v for diagonalizable matrices A.

THEOREM 3.1. *Let the spectral decomposition of A be $A = X\Lambda X^{-1}$ with the eigenvalues λ_i, $i = 1, \ldots, n$. The entries of $M = U^*U$ are given by*

$$M_{\ell,m} = \sum_{i=1}^{n} \sum_{j=1}^{n} (X^*X)_{i,j}\, \bar{c}_i c_j\, \bar{\lambda}_i^{\ell-1} \lambda_j^{m-1}, \quad \ell, m = 1, \ldots, n,$$

with $c = X^{-1}v$ and $\bar{\lambda}_i$ denoting the complex conjugate of λ_i. If the matrix A is normal, we have the simpler expression,

$$M_{\ell,m} = \sum_{i=1}^{n} |c_i|^2\, \bar{\lambda}_i^{\ell-1} \lambda_i^{m-1}, \quad \ell, m = 1, \ldots, n,$$

with $c = X^*v$.

Proof. Since we assumed that the matrix A is diagonalizable with eigenvalues λ_i, we have

$$K = X \begin{bmatrix} c & \Lambda c & \cdots & \Lambda^{n-1} c \end{bmatrix},$$

where $c = X^{-1}v$. Let D_c be the diagonal matrix with diagonal entries c_j, $j = 1, \ldots, n$. The matrix K is

$$K = X D_c \mathcal{V}$$

with the Vandermonde matrix

$$\mathcal{V} = \begin{bmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \cdots & \lambda_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \cdots & \lambda_n^{n-1} \end{bmatrix}.$$

We note that this factorization of the Krylov matrix has been used in [11]; see also [22]. Therefore $M = K^*K = \mathcal{V}^* D_{\bar{c}} X^* X D_c \mathcal{V}$. If A is normal, $X^*X = I$ and $M = \mathcal{V}^* D_\omega \mathcal{V}$

with $\omega_j = |c_j|^2$. The entries of M can be obtained as functions of the eigenvalues and eigenvectors of A by

$$\begin{aligned} M_{\ell,m} &= e_\ell^T M e_m = e_\ell^T \mathcal{V}^* D_{\bar{c}} X^* X D_c \mathcal{V} e_m \\ &= \begin{bmatrix} \bar{\lambda}_1^{\ell-1} & \cdots & \bar{\lambda}_n^{\ell-1} \end{bmatrix} D_{\bar{c}} X^* X D_c \begin{bmatrix} \lambda_1^{m-1} \\ \vdots \\ \lambda_n^{m-1} \end{bmatrix} = \sum_{i=1}^n \sum_{j=1}^n (X^*X)_{i,j}\, \bar{c}_i c_j\, \bar{\lambda}_i^{\ell-1} \lambda_j^{m-1}. \end{aligned}$$

If A is normal, we have $X^*X = I$ and

$$M_{\ell,m} = \sum_{i=1}^n |c_i|^2\, \bar{\lambda}_i^{\ell-1} \lambda_i^{m-1}.$$

This last result is already known from [20]. □

4. The inverse problem for diagonalizable matrices. For the first Arnoldi iteration (that is, $k = 1$) the inverse problem is always solvable. We have $h_{1,1} = v^*Av$. For $\theta^{(1)} \in W(A)$, there exists a vector v such that $\theta^{(1)} = v^*Av$. Algorithms for computing such vectors are given in [5, 8, 16]. We note that if A and v are real, the first Ritz value $\theta_1^{(1)}$ is real.

For the inverse problem at the Arnoldi iteration $k > 1$, we assume that we have a set of k given complex numbers $\theta_1, \ldots, \theta_k$ belonging to $W(A)$, and we would like to find (if possible) a vector v of unit norm such that the values θ_j are the Ritz values at iteration k when running the Arnoldi algorithm with (A, v).

From (2.1) we have an equation relating the coefficients of the characteristic polynomial of H_k and the entries of a submatrix of M. Since the Ritz values are zeros of the polynomial $\lambda^k + \sum_{j=0}^{k-1} \beta_j^{(k)} \lambda^j = \prod_{i=1}^k (\lambda - \theta_i)$, the coefficients $\beta_j^{(k)}$ are (up to the sign) elementary symmetric functions of the numbers θ_j. Therefore,

$$\beta_j^{(k)} = (-1)^{k-j} e_{(k-j)}(\theta_1, \ldots, \theta_k), \quad j = 0, \ldots, k-1, \tag{4.1}$$

with

$$e_{(i)}(\theta_1, \ldots, \theta_k) = \sum_{1 \le j_1 < j_2 < \cdots < j_i \le k} \theta_{j_1} \cdots \theta_{j_k}, \quad i = 1, \ldots, k.$$

Thus, we have the following characterization of the existence of a starting vector.

THEOREM 4.1. *There exists a starting vector $v = Xc$ of unit norm such that $\theta_1, \ldots, \theta_k$ are the Arnoldi Ritz values at iteration k if and only if the nonlinear system* (2.1) *with the unknowns c_j, $j = 1, \ldots, n$, (where the coefficients $\beta_j^{(k)}$ are defined by* (4.1)*), to which we add the equation*

$$\sum_{i,j=1}^n \bar{c}_i c_j (X^*X)_{i,j} = 1, \tag{4.2}$$

has at least one solution vector c.

Proof. Let us assume that there exists a vector v such that $\theta_1, \ldots, \theta_k$ are the Arnoldi Ritz values at iteration k. They are the roots of the characteristic polynomial whose coefficients $\beta_j^{(k)}$ are given by (4.1). Hence, by Theorem 2.2, the coefficients are solution of the linear system (2.1) and the vector c is a solution of the nonlinear system defined by (2.1) plus (4.2) because the vector v is of unit norm.

Conversely, if there is a solution c to the nonlinear system (2.1)–(4.2), then there exists a solution of the linear system (2.1) with the unknowns $\beta_j^{(k)}$, which, by Theorem 2.2, are the coefficients of the characteristic polynomial of H_k, and the complex numbers defined as the roots of the polynomial are the Ritz values at Arnoldi iteration k. $\square$

To make things clear, let us consider the case $k = 2$ with $\theta_1 = \theta_1^{(2)}$, $\theta_2 = \theta_2^{(2)}$ given. Let $p = \theta_1\theta_2$ and $s = \theta_1 + \theta_2$ be known. We note that M_2 is an Hermitian matrix. Then (2.1) is

$$M_2 \begin{bmatrix} p \\ -s \end{bmatrix} = -M_{[1:2],3}.$$

Therefore, we have the two equations,

$$p - sM_{1,2} = -M_{1,3}, \quad sM_{2,2} = M_{2,3} + p\overline{M_{1,2}}.$$

The equations to be satisfied are

$$p - s \sum_{i,j=1}^{n} \bar{c}_i c_j (X^*X)_{i,j}\lambda_j = -\sum_{i,j=1}^{n} \bar{c}_i c_j (X^*X)_{i,j}\lambda_j^2,$$
$$s \sum_{i,j=1}^{n} \bar{c}_i c_j (X^*X)_{i,j}\bar{\lambda}_i\lambda_j = \sum_{i,j=1}^{n} \bar{c}_i c_j \, (X^*X)_{i,j} \, \bar{\lambda}_i\lambda_j^2 + p \sum_{i,j=1}^{n} \bar{c}_i c_j \, (X^*X)_{i,j} \, \bar{\lambda}_i.$$

Since we need to find a vector v of unit norm, we have to add the condition $\|Xc\|^2 = c^*X^*Xc = 1$, which yields the equation

$$\sum_{i,j=1}^{n} \bar{c}_i c_j (X^*X)_{i,j} = 1.$$

Because s and p are known, these are three nonlinear complex equations in n complex unknowns c_i, $i = 1, \ldots, n$. Whether or not this system has solutions determines if θ_1 and θ_2 are feasible values since, if a solution c exists, we can then find a vector v such that the two given values θ_1 and θ_2 are Ritz values for $\mathcal{K}_2(A, v)$.

We remark that this is in general not a polynomial system because of the conjugacy in the expression $\bar{c}_i c_j$. However, we can convert this system into a polynomial system by considering the real and imaginary parts of c_i as unknowns. We have then a polynomial system of six equations in $2n$ unknowns with complex coefficients that can be converted to a polynomial system with real coefficients by taking the real and imaginary parts. The trouble then is that we have to know if there are real solutions. Unfortunately there are not many results about this problem in algebraic geometry literature. The situation is much simpler if we assume that the matrix A is normal. This case is considered in the next section.

5. The inverse problem for normal matrices.

For a normal matrix and assuming that we know the coefficients $\beta_0^{(k)}, \ldots, \beta_{k-1}^{(k)}$, we obtain a $(k+1) \times n$ linear system for the moduli squared, $\omega_i = |c_i|^2$. It yields a linear system

$$C_C\omega = f_C.$$

Putting the normalizing equation $\sum_{i=1}^{n} \omega_i = 1$ first, the entries of C_C are all 1 in the first row. The entries of the second row are

$$(C_C)_{2,m} = \sum_{i=1}^{k-1} \beta_i^{(k)} \lambda_m^i + \lambda_m^k, \qquad m = 1, \ldots, n,$$

and the other entries are

$$(C_C)_{\ell,m} = \sum_{i=0}^{k-1} \beta_i^{(k)} \bar{\lambda}_m^{\ell-2} \lambda_m^i + \bar{\lambda}_m^{\ell-2} \lambda_m^k, \quad \ell = 3, \dots, k+1,\ m = 1, \dots, n.$$

The right-hand side is all zero except for the first two components, $(f_C)_1 = 1$, $(f_C)_2 = -\beta_0^{(k)}$.

We can also turn this linear system of $k+1$ complex equations in n real unknowns into a real linear system by taking the real and imaginary parts of rows 2 to k. It gives a $(2k+1) \times n$ matrix C_R, and the right-hand side is zero except for the first three components $(f_R)_1 = 1$, $(f_R)_2 = -\text{Re}[\beta_0^{(k)}]$, $(f_R)_3 = -\text{Im}[\beta_0^{(k)}]$.

Compared to the case of a general diagonalizable matrix studied in the previous section, there are good and bad things. The good thing is that we have a linear system for the unknowns ω_i instead of a nonlinear one. The bad thing is that we need to find a solution which is real and positive. Obtaining (if possible) a real solution is easy by solving $C_R \omega = f_R$, but we still need a positive solution. The characterization of $\theta_1, \dots, \theta_k$ being feasible is given in the following theorem.

THEOREM 5.1. *Let A be a normal matrix. There exists a starting vector $v = Xc$ of unit norm such that $\theta_1, \dots, \theta_k$ are the Arnoldi Ritz values at iteration k if and only if the linear system $C_R \omega = f_R$, where the coefficients $\beta_j^{(k)}$ are defined by (4.1), has at least one solution vector ω with $\omega_i \geq 0$, $i = 1, \dots, n$. Then c is any vector such that $|c_i|^2 = \omega_i$.*

Proof. The proof is similar to that of Theorem 4.1. □

The condition given in Theorem 5.1 must be equivalent to the condition recently proposed by Bujanović ([3, Theorem 4]).

For further use let us write down the equations for $k = 2$. We have

$$\begin{aligned} p - s \sum_{i=1}^{n} |c_i|^2 \lambda_i &= - \sum_{i=1}^{n} |c_i|^2 \lambda_i^2, \\ s \sum_{i=1}^{n} |c_i|^2 \, |\lambda_i|^2 &= \sum_{i=1}^{n} |c_i|^2 \, |\lambda_i|^2 \lambda_i + p \sum_{i=1}^{n} |c_i|^2 \, \bar{\lambda}_i, \\ \sum_{i=1}^{n} |c_i|^2 &= 1. \end{aligned}$$

The problem can be further simplified if the matrix A and the starting vector are real. To the best of our knowledge, this case has not been considered by other authors. Then the eigenvalues of A are real or occur in complex conjugate pairs. If the starting vector v is real, all computed results are real in the Arnoldi algorithm (in particular the matrix H) and the Ritz values are real or appear as complex conjugate pairs which are the roots of a polynomial with real coefficients $\beta_j^{(k)}$. The two eigenvectors of A corresponding to a complex conjugate pair are conjugate, and the eigenvectors corresponding to real eigenvalues are real. Then, with v being real, if $c = X^* v$ and $\lambda_i = \bar{\lambda}_j$, we have $c_i = \bar{c}_j$. This means that when the Ritz values are known, we have only one unknown c_i for each pair of complex conjugate eigenvalues. Let us assume that the matrix A has p_C pairs of complex conjugate eigenvalues (with $2p_C \leq n$) that are listed first and $n - 2p_C$ real eigenvalues denoted by $(\lambda_{2p_C+1}, \dots, \lambda_n)$. Then, we have only $n - p_C$ unknowns that, to avoid some confusion, we denote by their initial indices ranging from 1 to n as usual for eigenvalues. That is, the unknowns are the components of the vector

$$\tilde{\omega} = \left[|c_1|^2, \quad |c_3|^2, \quad \dots, \quad |c_{2p_C-1}|^2, \quad |c_{2p_C+1}|^2, \quad |c_{2p_C+2}|^2, \quad \dots, \quad |c_n|^2 \right]^T. \tag{5.1}$$

Then, in the equations derived from the matrix M, we have to group the terms containing λ_i and $\bar{\lambda}_i$. Since only real numbers are involved, we denote the matrix as C_R even though it is different from the matrix described above. The first row of the matrix C_R is now

$$\begin{aligned} &(C_R)_{1,m} = 2, \quad m = 1, \ldots, p_C, \\ &(C_R)_{1,m} = 1, \quad m = p_C + 1, \ldots, n - p_C. \end{aligned} \tag{5.2}$$

The second row is

$$\begin{aligned} (C_R)_{2,m} &= 2 \sum_{i=1}^{k-1} \beta_i^{(k)} \mathrm{Re}(\lambda_{2m-1}^i) + 2\mathrm{Re}(\lambda_{2m-1}^k), \quad && m = 1, \ldots, p_C, \\ (C_R)_{2,m} &= \sum_{i=1}^{k-1} \beta_i^{(k)} \lambda_{p_C+m}^i + \lambda_{p_C+m}^k, && m = p_C + 1, \ldots, n - p_C, \end{aligned} \tag{5.3}$$

and the other entries are

$$\begin{aligned} (C_R)_{\ell,m} &= 2 \sum_{i=0}^{k-1} \beta_i^{(k)} \mathrm{Re}(\bar{\lambda}_{2m-1}^{\ell-2} \lambda_{2m-1}^i) + 2\mathrm{Re}(\bar{\lambda}_{2m-1}^{\ell-2} \lambda_{2m-1}^k), \\ &\qquad\qquad \ell = 3, \ldots, k+1, \ \ m = 1, \ldots, p_C, \\ (C_R)_{\ell,m} &= \sum_{i=0}^{k-1} \beta_i^{(k)} \lambda_{p_C+m}^{\ell-2+i} + \lambda_{p_C+m}^{\ell-2+k}, \\ &\qquad\qquad \ell = 3, \ldots, k+1, \ \ m = p_C + 1, \ldots, n - p_C. \end{aligned} \tag{5.4}$$

The right-hand side is all zero except for the first two components, $(f_R)_1 = 1$, $(f_R)_2 = -\beta_0^{(k)}$. Therefore, the real matrix C_R is only of size $(k+1) \times (n - p_C)$, and we have $n - p_C$ unknowns. For $k = 2$ and with $s = \theta_1 + \theta_2$, $p = \theta_1 \theta_2$, the second row is

$$\begin{aligned} (C_R)_{2,m} &= -2s\mathrm{Re}(\lambda_{2m-1}) + 2\mathrm{Re}(\lambda_{2m-1}^2), \qquad && m = 1, \ldots, p_C, \\ (C_R)_{2,m} &= -s\lambda_{p_C+m} + \lambda_{p_C+m}^2, && m = p_C + 1, \ldots, n - p_C, \end{aligned}$$

and the other entries are

$$\begin{aligned} (C_R)_{3,m} &= 2p\mathrm{Re}(\bar{\lambda}_{2m-1}) - 2s|\lambda_{2m-1}|^2 + 2\mathrm{Re}(\bar{\lambda}_{2m-1} \lambda_{2m-1}^2), \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad m = 1, \ldots, p_C, \\ (C_R)_{3,m} &= p\lambda_{p_C+m} - s\lambda_{p_C+m}^2 + \lambda_{p_C+m}^3, \qquad\qquad m = p_C + 1, \ldots, n - p_C. \end{aligned}$$

To find out if there exist a positive solution, we have to consider the cases $k+1 > n - p_C$ (overdetermined system), $k + 1 = n - p_C$ (square system), and $k + 1 < n - p_C$ (underdetermined system). When we have a positive solution, we can find a real vector c by expanding the solution and taking square roots and finally obtain a real starting vector $v = Xc$. The previous discussion is summarized in the following theorem.

THEOREM 5.2. *Let A be a real normal matrix. There exists a real starting vector $v = Xc$ of unit norm such that $\theta_1, \dots, \theta_k$, where these values are real or occur in complex conjugate pairs, are the Arnoldi Ritz values at iteration k if and only if the linear system $C_R\tilde{\omega} = f_R$, where the coefficients $\beta_j^{(k)}$ are defined by (4.1), the matrix C_R is defined by (5.2)–(5.4), and $\tilde{\omega}$ by (5.1), has at least one solution vector $\tilde{\omega}$ with $\tilde{\omega}_i \geq 0$, for $i = 1, \dots, n - p_C$. Then c is any real vector such that $|c_i|^2 = \omega_i$ where ω is given by the expansion of $\tilde{\omega}$.*

Let us now consider the problem of finding a positive solution in the case that the linear system $C_R\tilde{\omega} = f_R$ is underdetermined, that is, $k+1 < n-p_C$. Solutions of a system like this can be found by using the Singular Value Decomposition (SVD). Let us consider the generic case where C_R has full rank $k + 1$. The matrix can be factorized as

$$C_R = \hat{U} \begin{bmatrix} D & 0 \end{bmatrix} \hat{V}^*, \qquad D \text{ diagonal}, \; \hat{U}^*\hat{U} = I, \; \hat{V}^*\hat{V} = I.$$

The orthonormal matrix $\hat{U}$ is of order $k + 1$ as well as D, and $\hat{V}$ is of order $n - p_C$. The diagonal of D contains the singular values. Since all the singular values are non-zero, we can find solutions to

$$\hat{U} \begin{bmatrix} D & 0 \end{bmatrix} \hat{V}^*\tilde{\omega} = \begin{bmatrix} 1 \\ (f_R)_2 \\ 0 \end{bmatrix}.$$

Let $y = \hat{V}^*\tilde{\omega}$,

$$\hat{U} \begin{bmatrix} D & 0 \end{bmatrix} y = \begin{bmatrix} 1 \\ (f_R)_2 \\ 0 \end{bmatrix} \quad \Longrightarrow \quad \hat{y} \equiv \begin{bmatrix} y_1 \\ \vdots \\ y_{k+1} \end{bmatrix} = D^{-1}\hat{U}^* \begin{bmatrix} 1 \\ (f_R)_2 \\ 0 \end{bmatrix}.$$

The solutions are given by

$$\tilde{\omega} = \hat{V} \begin{bmatrix} y_1 \\ \vdots \\ y_{k+1} \\ \times \\ \vdots \\ \times \end{bmatrix},$$

where the symbol × denotes an arbitrary real number. Let us decompose the matrix $\hat{V}$ as $\hat{V} = [\hat{V}_1 \; \hat{V}_2]$ with $\hat{V}_1$ having $k + 1$ columns. Then, we have a positive solution if and only if there exists a vector z such that

$$-\hat{V}_2 z \leq \hat{V}_1 \hat{y} \tag{5.5}$$

and

$$\tilde{\omega} = \hat{V} \begin{bmatrix} \hat{y} \\ z \end{bmatrix}.$$

To check if there is a solution to the system of inequalities (5.5), we use the algorithm described in [12] that was intended to convert a system of linear inequalities into a representation using the vertices of the polyhedron defined by the inequalities. It relies on computing the rank of submatrices and tells us if the system is feasible or not.

6. The case A real normal and $k = 2$. In this section we further simplify the problem and concentrate on the case $k = 2$ for a real normal matrix and a real starting vector. The matrix H_2 is real and has either two real eigenvalues or a pair of complex conjugate eigenvalues. We are interested in the latter case for which we have $\theta_2 = \bar{\theta}_1$. Hence, it is enough to look for the location of the complex Ritz value θ_1 and this considerably simplifies the problem. We call the set of all the complex values θ_1 in the field of values yielding a positive solution *the feasible region*. To obtain a graphical representation of the feasible region we can proceed as in Bujanović's paper [3]. We set up a regular Cartesian mesh over the field of values (in fact over the smallest rectangle containing the upper part, $y \geq 0$, of the field of values) of A for the values of θ_1, and we check if there are positive solutions to the $3 \times (n - p_C)$ linear system $C_R\tilde{\omega} = f_R$ for each value of $\theta_1 = (x, y)$ in the mesh by considering the system of inequalities (5.5). When the system is feasible for a given value of θ_1 on the mesh, we flag this location. Hence, for each θ_1 in the marked area, we can find a real vector v such that $\theta_1, \theta_2 = \bar{\theta}_1$ are the Ritz values at iteration 2. This gives an approximation of the feasible region. For θ_1 outside of the feasible region, there does not exist a real vector v that yields $(\theta_1, \bar{\theta}_1)$ as Arnoldi Ritz values at iteration 2. Of course this way of detecting the feasible location of θ_1 by discretizing the field of values has some drawbacks since some tiny feasible parts may be missing if the discretization is not fine enough.

Figures 6.1 and 6.2 display an example (Example 1) of a matrix of order 4 with two real eigenvalues on each side of the real part of a pair of complex conjugate eigenvalues. More precisely, in Example 1 the matrix A is

$$A = \begin{bmatrix} -0.446075 & 0.358311 & -0.605655 & 1.12896 \\ -0.512738 & -0.263009 & -1.09795 & 0.285 \\ 1.15846 & 0.636041 & -0.72035 & 0.0184702 \\ -0.405993 & -1.00831 & -0.417846 & -0.456834 \end{bmatrix}.$$

The eigenvalues of A are

$$\left[\lambda_1, \bar{\lambda}_1, \lambda_3, \lambda_4\right] = \left[-0.432565 + 1.66558i, -0.432565 - 1.66558i, 0.187377, -1.20852\right].$$

In this example the matrix C_R is square of order 3 since $n - p_C = 4 - 1 = 3$ and nonsingular. The field of values is shown in red and the eigenvalues of A are the red circles. The feasible values of θ_1 (respectively θ_2) are marked with blue (respectively red) crosses. In this example the feasible region is a surface in the complex plane. In this case it is connected and convex (if we add the real values inside the region), but we will see later that this is not always the case. Of course, we can also have two real Ritz values outside this region. In this example it seems that the real Ritz values can be anywhere in the interval defined by the two real eigenvalues of A. Figure 6.2 was obtained by using 700 random real starting vectors and running the Arnoldi algorithm. We see that we obtain approximately the same shape for the feasible region.

Since we may miss some parts of the feasible region due to a too coarse discretization, it is interesting to characterize its boundary. This can be done by explicitly writing down the inverse of the matrix C_R and looking at the inequalities given by the positivity constraints for ω_j. It corresponds to the elimination of the components of ω in the equations. For simplicity let us denote

$$C_R = \begin{bmatrix} 2 & 1 & 1 \\ a & c & e \\ b & d & f \end{bmatrix}.$$

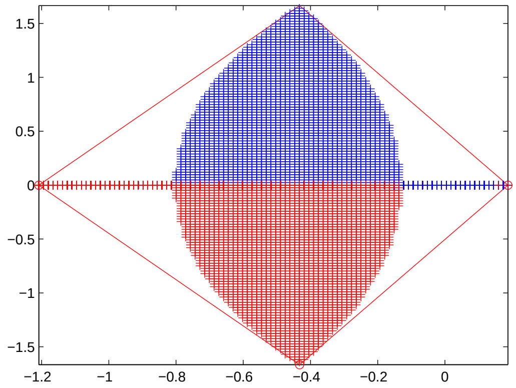

FIG. 6.1. *Location of $\theta_1 = \bar{\theta}_2$ for Example 1, $n = 4, k = 2$, A normal real.*

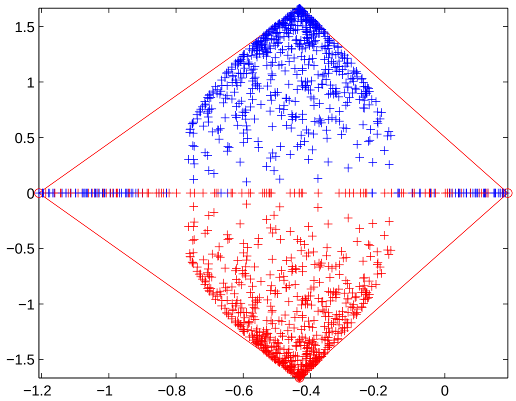

FIG. 6.2. *Location of $\theta_1 = \bar{\theta}_2$ for Example 1, $n = 4, k = 2$, A normal real, Arnoldi with random real vectors v.*

The inverse is given by

$$C_R^{-1} = \frac{1}{D}\begin{bmatrix} cf-ed & d-f & e-c \\ eb-af & 2f-b & a-2e \\ ad-cb & b-2d & 2c-a \end{bmatrix}, \quad D = a(d-f)+c(2f-b)+e(b-2d).$$

We apply the inverse to the right-hand side (which, after a change of signs, is $\begin{bmatrix}1 & p & 0\end{bmatrix}^T$, $p = |\theta_1|^2$), and we get

$$\omega = \frac{1}{D}\begin{bmatrix} cf-ed+(d-f)p \\ eb-af+(2f-b)p \\ ad-cb+(b-2d)p \end{bmatrix}.$$

We are interested in the components of ω being positive. The outside region of the feasible region is characterized by the fact that at least one component ω_j is negative. Therefore the boundary must be given by some of the components of the solution being zero. Hence, we

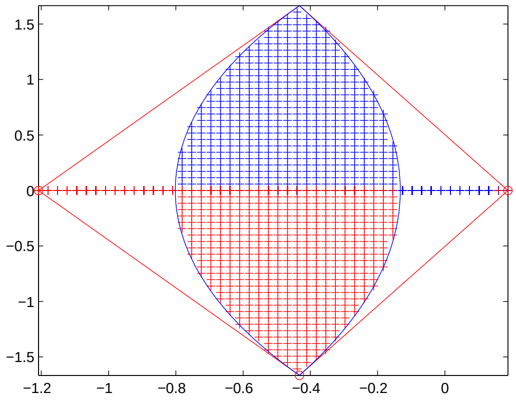

FIG. 6.3. *Location of $\theta_1 = \bar{\theta}_2$ and the boundary of the feasible region for Example 1, $n = 4, k = 2$, A normal real.*

have to look at the three equations

$$\begin{aligned} cf - ed + (d - f)p &= 0, \\ eb - af + (2f - b)p &= 0, \\ ad - cb + (b - 2d)p &= 0. \end{aligned}$$

The coefficients a, b, c, d, e, f are functions of the unknowns quantities $s = 2x = 2\mathrm{Re}(\theta_1)$ and $p = x^2 + y^2 = |\theta_1|^2$. These equations define three curves in the (x, y) complex plane. Some (parts) of these curves yield the boundary of the feasible region for θ_1. However, we note that one component can be zero on one of the curves without changing sign. Therefore, not all the curves might be relevant for the boundary. We just know that the boundary is contained in the union of the curves. Moreover, we are only interested in the parts of the curves contained in the convex hull of the eigenvalues of A. For completeness, remember that we have

$$\begin{aligned} a &= 2s\mathrm{Re}(\lambda_1) - 2\mathrm{Re}(\lambda_1^2), & b &= 2s|\lambda_1|^2 - 2|\lambda_1|^2\mathrm{Re}(\lambda_1) - 2p\mathrm{Re}(\lambda_1), \\ c &= s\lambda_3 - \lambda_3^2, & d &= s\lambda_3^2 - \lambda_3^3 - p\lambda_3, \\ e &= s\lambda_4 - \lambda_4^2, & f &= s\lambda_4^2 - \lambda_4^3 - p\lambda_4. \end{aligned}$$

The first curve involves only the real eigenvalues of A. The two other curves pass through λ_1 and $\bar{\lambda}_1$.

Figure 6.3 displays the boundary curves that we obtain for Example 1 as well as the approximation of the feasible region using a smaller number of discretization points than before. These curves were obtained using a contour plot of the level 0 for the three functions of x and y. We see that we do get the boundary of the feasible region for θ_1. We have only two curves since the one which depends only on the real eigenvalues (corresponding to the first equation) does not have points in the window we are interested in (that is, the field of values).

Let us now consider $n = 5$. The matrix A has either two pairs of complex conjugate eigenvalues $(\lambda_1, \bar{\lambda}_1)$, $(\lambda_3, \bar{\lambda}_3)$ and a real eigenvalue λ_4 (that can be on the boundary or inside the field of values) or one pair of complex conjugate eigenvalues and three real eigenvalues (there is at least one inside the field of values, eventually two). In the first case the matrix C_R

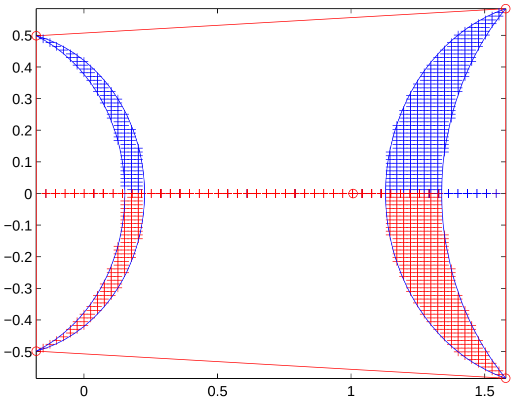

FIG. 6.4. *Location of $\theta_1 = \bar{\theta}_2$ for Example 2, $n = 5, k = 2$, A normal real.*

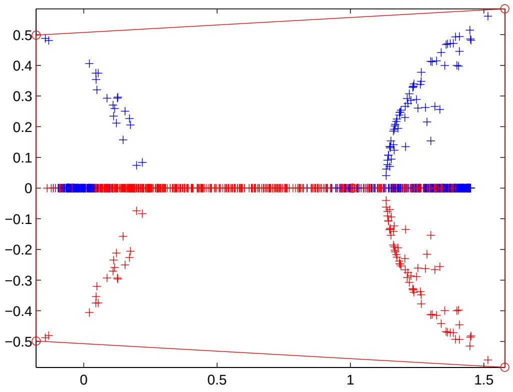

FIG. 6.5. *Location of $\theta_1 = \bar{\theta}_2$ for Example 2, $n = 5, k = 2$, A normal real, Arnoldi with random real vectors v.*

is of size 3×3 as it was for the previous example with $n = 4$. Then, we can apply the same techniques by eliminating the components of ω to obtain the boundary of the feasible region for θ_1. The only differences are that $(C_R)_{1,2} = 2$ and the values of the coefficients. We have

$$\begin{aligned} a &= 2s\mathrm{Re}(\lambda_1) - 2\mathrm{Re}(\lambda_1^2), & b &= 2s|\lambda_1|^2 - 2|\lambda_1|^2\mathrm{Re}(\lambda_1) - 2p\mathrm{Re}(\lambda_1), \\ c &= 2s\mathrm{Re}(\lambda_3) - 2\mathrm{Re}(\lambda_3^2), & d &= 2s|\lambda_3|^2 - 2|\lambda_3|^2\mathrm{Re}(\lambda_3) - 2p\mathrm{Re}(\lambda_3), \\ e &= s\lambda_4 - \lambda_4^2, & f &= s\lambda_4^2 - \lambda_4^3 - p\lambda_4. \end{aligned}$$

The equations are

$$\begin{aligned} cf - ed + (d - 2f)p &= 0, \\ eb - af + (2f - b)p &= 0, \\ ad - bc + 2(b - d)p &= 0. \end{aligned}$$

These equations only differ slightly from those above by a multiplicative factor of 2 at some terms. Figure 6.4 displays the feasible region and the boundary for a case with two pairs of complex conjugate eigenvalues and a real eigenvalue inside the field of values (Example 2).

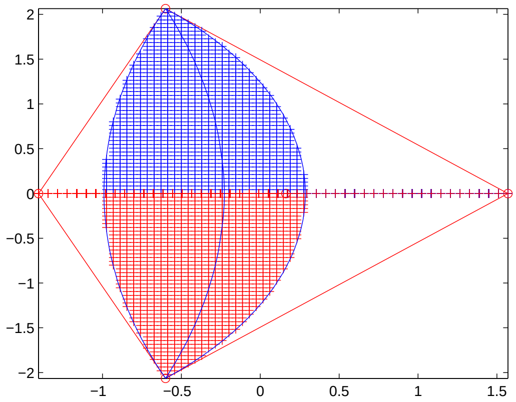

FIG. 6.6. *Location of $\theta_1 = \bar{\theta}_2$ for Example 3, $n = 5$, $k = 2$, A normal real.*

In this example the feasible region is neither connected nor convex. Figure 6.5 shows that random starting vectors do not always yield a good rendering of the feasible region. In Example 2 the matrix is

$$A = \begin{bmatrix} 0.513786 & -0.419578 & 0.156842 & 0.447046 & 0.540983 \\ -0.789795 & 0.767537 & -0.451475 & 0.12333 & 0.202036 \\ 0.0825256 & -0.091751 & 1.31755 & 0.5561 & -0.00409194 \\ 0.179105 & 0.7687 & -0.247999 & 1.31189 & 0.0474895 \\ -0.174622 & -0.329046 & -0.185905 & 0.403025 & -0.101738 \end{bmatrix}.$$

The eigenvalues of A are

$$\big[-0.178102 + 0.498599i, -0.178102 - 0.498599i, \\ 1.5788 + 0.584391i, 1.5788 - 0.584391i, 1.00762 \big].$$

Figure 6.6 displays an example with only one pair of complex conjugate eigenvalues and three real eigenvalues with two of them on the same side of the real part of the complex eigenvalues (Example 3). We see that we have one piece of the curve which is inside the feasible region. It can be considered a “spurious” curve (even though we will see later that these curves can also have some interest). The matrix of Example 3 is

$$A = \begin{bmatrix} -1.07214 & -0.549535 & 0.809383 & -0.0826907 & 0.345094 \\ -0.779134 & 1.06039 & 0.100179 & 0.621762 & -0.184854 \\ -0.33126 & -0.0693308 & -0.551724 & 1.39559 & 1.19566 \\ -0.24838 & 0.134568 & -0.902458 & -0.0781342 & -1.22051 \\ -0.551853 & -0.844358 & -1.41854 & 0.206828 & -0.233364 \end{bmatrix}.$$

The eigenvalues of A are

$$\big[-0.600433 + 2.06392i, -0.600433 - 2.06392i, -1.40594, 1.56985, 0.161981 \big].$$

For n larger than 5, the problem of computing the boundary is more complicated. We generally have more than 3 unknowns (except for $n = 6$ with three pairs of complex conjugate eigenvalues) and therefore an underdetermined linear system for the unknowns ω_j. When prescribing a value of θ_1 (with $\theta_2 = \bar{\theta}_1$), as we have seen before, we can check the feasibility by using the SVD of the rectangular matrix $C_R = \hat{U} S \hat{V}^T$.

Concerning the boundary of the feasible region, the pieces of the boundary correspond to some of the components of ω being zero. Therefore, we can apply the same elimination technique as before by considering the matrices of order 3 corresponding to all the triples of eigenvalues, a pair of complex conjugate ones counting only for one. It corresponds to considering only three components of ω putting the other components to zero. We have to consider 3×3 matrices similar as the ones we had before with the first row being $(2, 2, 2)$, $(2, 1, 1)$, or $(1, 1, 1)$. The number of curves is three times the number of triples of eigenvalues.

Doing this corresponds to the handling of linear constraints in linear programming (LP) whose solution components must be positive. Let us assume that we have linear equality constraints $Cx = b$ defined by a real $m \times n$ matrix C of full rank with $m < n$. This procedure just amounts to taking m independent columns of C, putting the other components of the solution to zero, and solving. By possibly permuting columns, we can write $C = [B\ E]$ with B nonsingular of order m. Then

$$x = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$$

is called a basic solution. It is degenerate if some components of $B^{-1}b$ are zero. A basic feasible solution (BFS) is a basic solution that satisfies the constraints of the LP. The feasible region defined by the constraints is convex, closed, and bounded from below. The feasible region is a polyhedron, and it can be shown that the BFS are extreme points (vertices or corners) of the feasible region.

This is similar to what we are doing. We have a polyhedron in the ω-space defined by the system with the matrix C_R, consider all the 3×3 matrices (provided they are nonsingular), and symbolically compute the basic solutions. The feasible ones (with $\omega_j \geq 0$) correspond to some vertices of the polyhedron. Clearly these curves are located where components of ω may change signs as a function of $x = \text{Re}(\theta_1)$ and $y = \text{Im}(\theta_1)$. They also give a parametric description of the vertices of the polyhedron.

Figure 6.7 corresponds to an example with $n = 6$ and three pairs of complex conjugate eigenvalues (Example 4). In this example the matrix C_R is square of order 3, ω has only three components, and there is no spurious curve. We see that the shape of the feasible region can be quite complicated. The matrix A of Example 4 is

$$A = \begin{bmatrix} -0.401151 & 0.0951597 & 0.336817 & -0.0155421 & 0.342989 & 0.059462 \\ 0.0435544 & -0.711607 & -0.0851345 & -0.100931 & 0.19691 & -0.0848016 \\ -0.3473 & 0.0330741 & -0.458265 & 0.338473 & 0.161655 & -0.163792 \\ 0.0903354 & 0.144387 & 0.0427703 & -0.167152 & 0.14634 & -0.661259 \\ -0.309586 & -0.118264 & 0.148041 & -0.196687 & -0.517635 & -0.205145 \\ 0.131915 & -0.142526 & 0.380522 & 0.570822 & -0.0924846 & -0.165907 \end{bmatrix}.$$

The eigenvalues of A are

$$\begin{aligned}
\big[&-0.0640196 + 0.732597i, -0.0640196 - 0.732597i, \\
&-0.390646 + 0.477565i, -0.390646 - 0.477565i, \\
&-0.756193 + 0.125533i, -0.756193 - 0.125533i\big].
\end{aligned}$$

Figure 6.8 corresponds to an example with two pairs, one real eigenvalue on the boundary of the field of values, and one real eigenvalue inside (Example 5). We have two spurious curves. The matrix A is

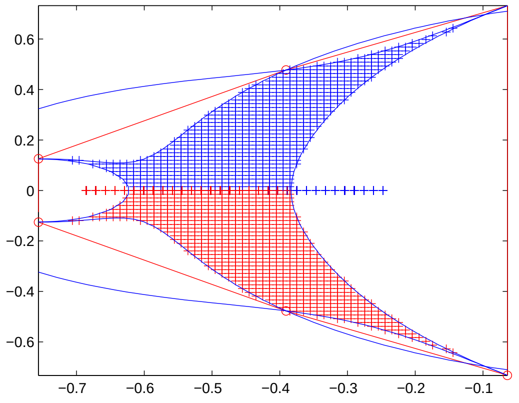

FIG. 6.7. *Location of $\theta_1 = \bar{\theta}_2$ for Example 4, $n = 6, k = 2$, A normal real.*

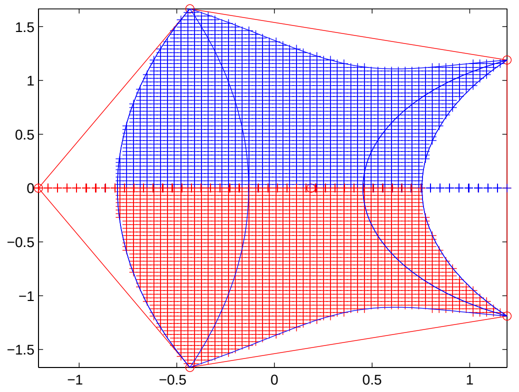

FIG. 6.8. *Location of $\theta_1 = \bar{\theta}_2$ for Example 5, $n = 6, k = 2$, A normal real.*

$$A = \begin{bmatrix} -0.500411 & 0.25411 & 0.499092 & -0.15696 & 1.26376 & -0.690147 \\ -0.850536 & 0.662412 & 0.12518 & 0.666057 & -0.873974 & -0.503358 \\ -0.095158 & 0.54861 & -0.0510311 & -0.42028 & -0.209823 & 0.122187 \\ 0.307198 & 0.827682 & -0.341422 & -0.437352 & 0.0411078 & -0.835649 \\ -1.00153 & 0.456062 & -0.0256999 & -0.551469 & 0.191305 & 1.01331 \\ 0.762756 & 0.970216 & 0.404506 & 0.804347 & 0.368779 & 0.630639 \end{bmatrix}.$$

The eigenvalues of A are

$$\begin{aligned} &[-0.432565 + 1.66558i, -0.432565 - 1.66558i, \\ &1.19092 + 1.18916i, 1.19092 - 1.18916i, -1.20852, 0.187377.] \end{aligned}$$

To visualize the feasible region for θ_1, it is useful to get rid of the "spurious" curves. This can be done approximately in the following way. We can compute points on the curves by solving an equation in x for a given value of y (or vice-versa) for each equation defining the boundary. When we get a point on a curve, we can check points surrounding it in the complex plane. If there is at least one of those points which is not feasible, then our given

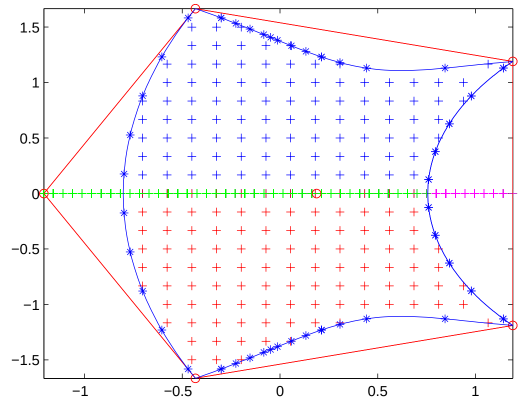

FIG. 6.9. *Boundary of the feasible region for Example 5, $n = 6, k = 2$.*

point is on the boundary and the piece of the curve on which it is located is a part of the boundary. This is not always easy because of rounding errors and because we could have some curves which are almost tangent to each other. Of course this process is not foolproof since the result depends on the choice of the surrounding points and also on some thresholds. But in many cases it works fine. Figure 6.9 shows what we get for the previous example. The blue stars are the points computed on the boundary (using the Matlab function `fzero`). Note that we get rid of the two spurious curves since we keep only the curves on which there is at least one boundary point.

There is another way to visualize the boundary of the feasible region in Matlab. The `contour` function that we use is evaluating the function on a grid and then finding the curve of level 0 by interpolation. Therefore, we can set up a routine that, given x and y, computes a solution of the underdetermined system for the point $(x + iy, x - iy)$ using the SVD. If the point is not feasible, then we return a very small negative value. However, this process is very expensive since the evaluation of the function cannot be vectorized. An example is given below in Figure 6.11 for the next Example 6. Of course we do not have spurious curves and not even the parts of the curves that are not relevant. But we have some wiggles in the curve because we set the values for non-feasible points to a small negative value introducing discontinuities in the function values.

Figure 6.10 displays an example with two pairs (one inside the field of values) and two real eigenvalues (Example 6). The feasible region has an interesting shape. Figure 6.11 shows the boundary for Example 6 computed using the SVD. The matrix is

$$A = \begin{bmatrix} 0.0433091 & 1.59759 & -0.318964 & -0.787924 & -1.5765 & 0.538701 \\ 0.222478 & -0.276959 & 0.775185 & 1.54146 & 1.8561 & 0.818277 \\ 0.348846 & -0.0614769 & 1.02246 & -0.677541 & -0.498161 & 0.193331 \\ 1.05979 & -1.7532 & -0.176368 & 0.214925 & -0.563343 & -0.580403 \\ 2.01859 & -0.900034 & 0.21777 & -1.05788 & -0.388673 & -1.0512 \\ 0.825456 & -0.837442 & 0.298154 & -0.554189 & 0.812614 & 0.77613 \end{bmatrix}.$$

The eigenvalues of A are

$$\big[-1.2413 + 3.27037i, -1.2413 - 3.27037i,$$
$$0.566382 + 0.768588i, 0.566382 - 0.768588i, 0.917215, 1.82382\big].$$

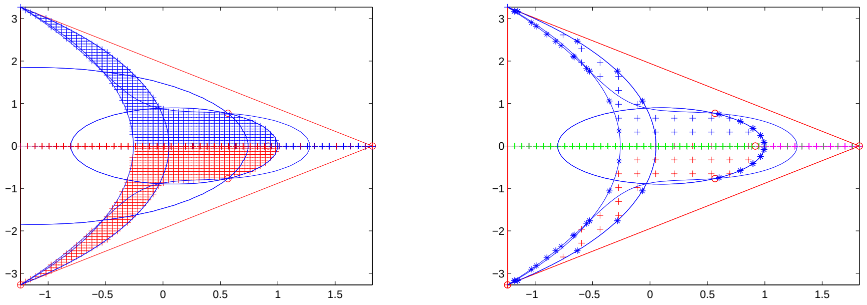

FIG. 6.10. *Location of $\theta_1 = \bar{\theta}_2$ for Example 6, $n = 6, k = 2$, A normal real.*

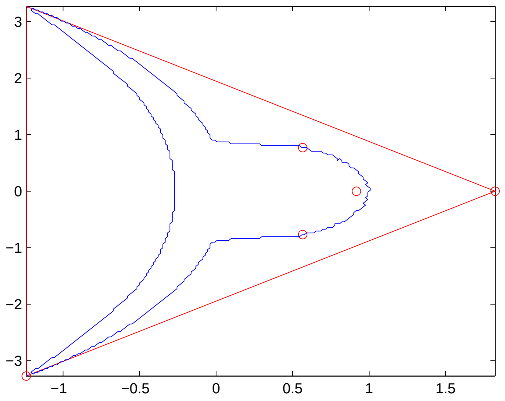

FIG. 6.11. *Boundary obtained with the SVD for Example 6, $n = 6, k = 2$, A normal real.*

Figure 6.12 displays an example with $n = 8$ (Example 7). We can see that (unfortunately) we have many spurious curves that are useless for the boundary. On the right part of Figure 6.12, we got rid of some of these curves but not all of them. The matrix of Example 7 is

$$A =$$
$$\begin{bmatrix} 0.541379 & 0.36045 & 0.724658 & -0.835226 & -0.882172 & 0.0513467 & -0.231744 & -0.316297 \\ -0.454221 & 0.575524 & -0.100099 & -0.312607 & -0.365987 & -0.122991 & 0.143776 & 0.447837 \\ 0.210676 & -0.0931479 & 0.852157 & 0.39926 & -0.119268 & -0.722606 & 0.199469 & 0.255216 \\ -0.921745 & -0.357353 & 0.0571532 & -0.569208 & -1.24529 & 1.17068 & 0.120452 & -0.304355 \\ -0.719429 & -0.137593 & 0.470774 & -1.33238 & -0.162772 & 1.02581 & -0.277858 & 0.154487 \\ 0.451727 & 0.489061 & -0.0903518 & 0.835521 & 1.06541 & 0.646274 & -0.158683 & 0.856737 \\ 0.00891101 & 0.0305841 & -0.23076 & -0.0649839 & 0.0463489 & 0.236475 & 0.810799 & -0.356549 \\ 0.682085 & -0.398763 & 0.179775 & -0.759383 & -0.268957 & 0.158633 & -0.112359 & 1.03084 \end{bmatrix}.$$

The eigenvalues of A are

$$\begin{aligned} &\big[1.68448 + 0.780709i, 1.68448 - 0.780709i, \\ &\ 0.418673 + 0.888289i, 0.418673 - 0.888289i, \\ &\ 0.882938 + 0.19178i, 0.882938 - 0.19178i, -2.9958, 0.748615\big]. \end{aligned}$$

We remark that, using the same technique as before, we can compute the boundary of the feasible region for θ_2 when θ_1 is prescribed for $k = 2$ and for complex normal matrices. Here we have to consider basic solutions for the real matrix which is of size $5 \times n$. Hence, we compute the solutions for all 5×5 matrices extracted from the system for ω. In this case

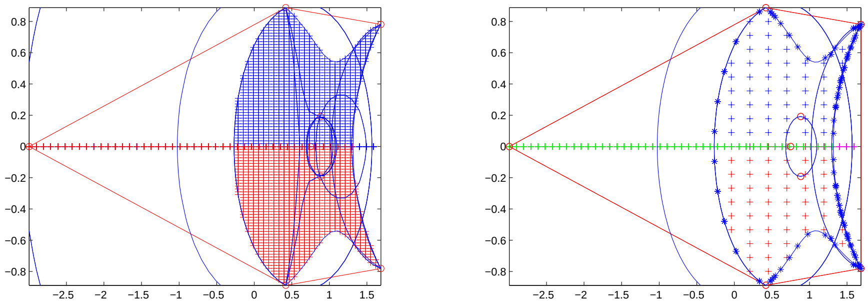

FIG. 6.12. *Location of* $\theta_1 = \bar{\theta}_2$ *for Example 7,* $n = 8, k = 2$, A *normal real.*

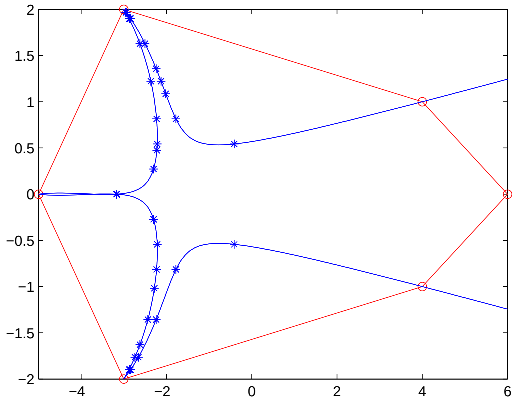

FIG. 6.13. *Boundary of the feasible region for* $\theta_2^{(2)}$ *for the Example in Figure 1(b) of [3],* $n = 6, k = 2$, $\theta_1^{(2)} = -4$.

we compute the solutions numerically and not symbolically for a given point (x, y). Then we check that the curves are indeed parts of the boundary using the same perturbation technique as before. We consider the problem of Bujanović [3, Figure 1 (b)]. The eigenvalues of A are

$$\big[-5, -3+2i, -3-2i, 4+i, 4-i, 6\big].$$

We fix $\theta_1^{(2)} = \theta_1 = -4$. The boundary of the feasible region for θ_2 for this particular value of θ_1 is displayed in Figure 6.13.

One can compare with [3] and see that we indeed find the boundary of the region for θ_2. However, such regions do not give a good idea of the location of the Ritz values because we would have to move θ_1 all over the field of values to see where the Ritz values can be located. Figure 6.14 displays the location of the Ritz values for $k = 2$ to 5 when running the Arnoldi method with a complex diagonal matrix with the given eigenvalues and random real starting vectors. We see that we have Ritz values almost everywhere. Things are strikingly different if we construct a real normal matrix with the given eigenvalues (which are real or occur in complex conjugate pairs) and run the Arnoldi method with real starting vectors. The Ritz values are shown in Figure 6.15. We see that they are constrained in two regions of the complex plane and on the real axis. Of course things would have been different if we would have used complex starting vectors. The Ritz values would have looked more like those in Figure 6.14. There is much more structure in the feasible region if everything is real-valued.

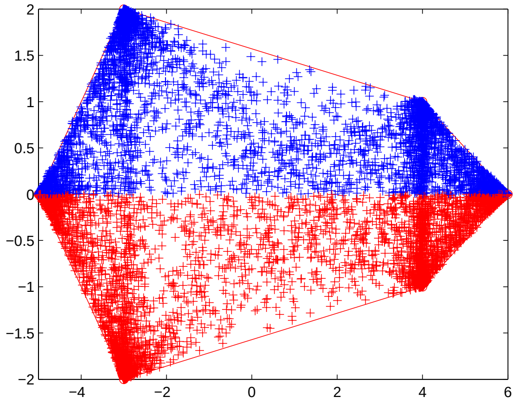

FIG. 6.14. *Location of the Ritz values,* $n = 6$*, all* $k = 2 : 5$*,* A *complex diagonal, Arnoldi with random real vectors* v*.*

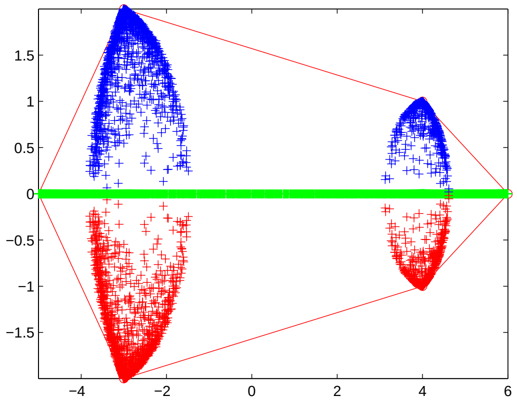

FIG. 6.15. *Location of the Ritz values,* $n = 6$*, all* $k = 2 : 5$*,* A *normal real, Arnoldi with random real vectors* v*.*

7. Open problems and conjectures for $k > 2$ and real normal matrices. In this section we describe some numerical experiments with $k > 2$ for real normal matrices. We also state some open problems and formulate some conjectures. We are interested in the iterations $k = 3$ to $k = n - 1$. We would like to know where the Ritz values are located when using real starting vectors. Clearly we cannot do the same as for $k = 2$ because, for instance, for $k = 3$, we have either three real Ritz values or a pair of complex conjugate Ritz values and a real one. Of course, we can fix the location of the real Ritz values and look for the region where the pairs of complex conjugate Ritz values may be located, but this is not that informative since it is not practical to explore all the possible positions of the real Ritz values.

Let us do some numerical experiments with random starting vectors and first consider Example 6 of the previous section with $n = 6$. For each value of $k = 2$ to $n - 1$, we generate 700 random initial vectors of unit norm, and we run the Arnoldi algorithm computing the Ritz values at iteration k. In Figure 7.1 we plot the pairs in blue and red and the real eigenvalues in green for all the values of k, and we superimpose the boundary curves computed for $k = 2$. We observe that all the Ritz values belong to the feasible region that was computed

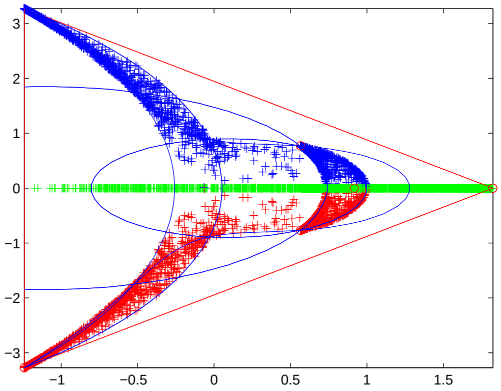

FIG. 7.1. *Location of the Ritz values for Example 6, $n = 6$, all $k = 2 : 5$, A normal real, Arnoldi with random real vectors v.*

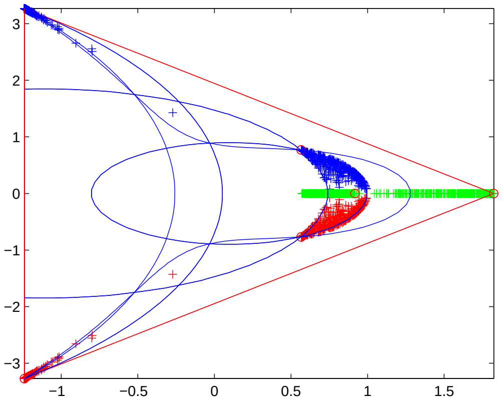

FIG. 7.2. *Location of the Ritz values for Example 6, $n = 6$, $k = 4$, A normal real, Arnoldi with random real vectors v.*

for $k = 2$. We conjecture that this is true for any real normal matrix and a real starting vector. But there is more than that.

Figure 7.2 displays the Ritz values at iteration 4. We see that some of the Ritz values are contained in a region for which one part of the boundary is one piece of a curve that was considered as “spurious” for $k = 2$. Figure 7.3 shows the Ritz values at iteration 5 (that is, the next to last one); there is an accumulation of some Ritz values on this spurious curve as well as close to the other pair of complex conjugate eigenvalues. It seems that some of the spurious curves look like “attractors” for the Ritz values, at least for random real starting vectors. It would be interesting to explain this phenomenon.

Figures 7.4–7.8 illustrate results for Example 7 with $n = 8$. Here again we observe that the Ritz values are inside the boundary for $k = 2$ and, at some iterations, Ritz values are located preferably on or close to some of the spurious curves.

Another open question is if there exist real normal matrices for which the feasible region for $k = 2$ completely fill the field of values for real starting vectors. In this paper we concentrated on pairs of complex conjugate Ritz values, but an interesting problem is to locate the real Ritz values in the intersection of the field of values with the real axis.

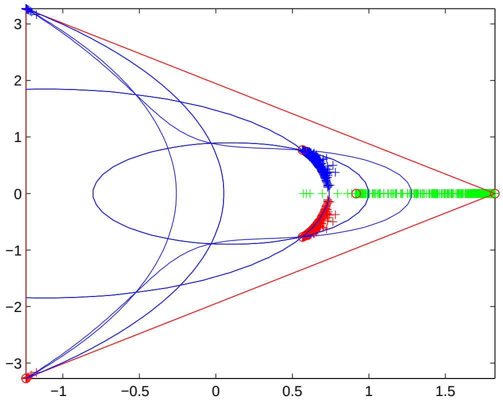

FIG. 7.3. *Location of the Ritz values for Example 6, $n = 6$, $k = 5$, A normal real, Arnoldi with random real vectors v.*

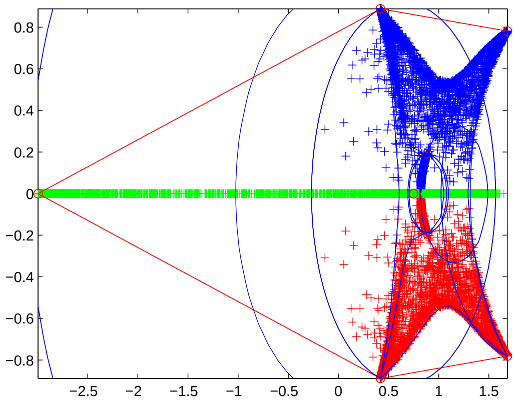

FIG. 7.4. *Location of the Ritz values for Example 7, $n = 8$, all $k = 2 : 7$, A normal real, Arnoldi with random real vectors v.*

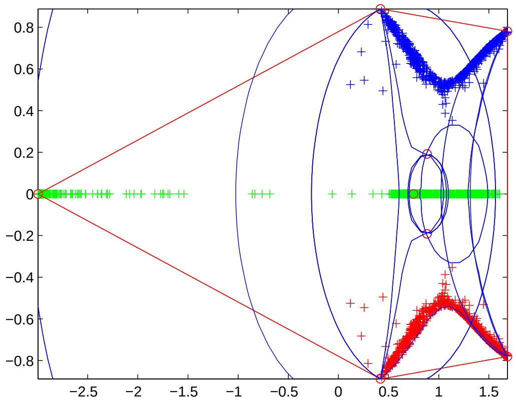

FIG. 7.5. *Location of the Ritz values for Example 7, $n = 8$, $k = 4$, A normal real, Arnoldi with random real vectors v.*

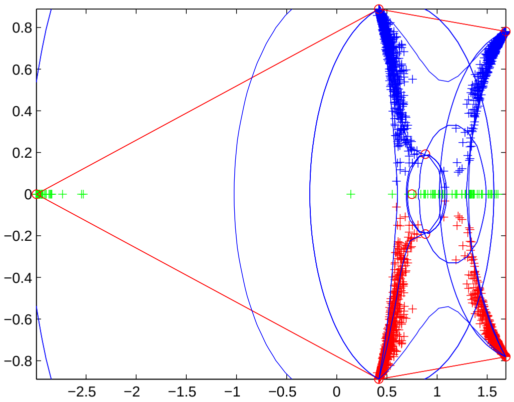

FIG. 7.6. *Location of the Ritz values for Example 7, $n = 8$, $k = 5$, A normal real, Arnoldi with random real vectors v.*

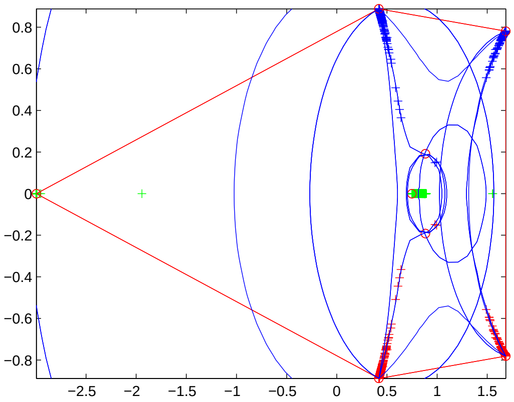

FIG. 7.7. *Location of the Ritz values for Example 7, $n = 8$, $k = 6$, A normal real, Arnoldi with random real vectors v.*

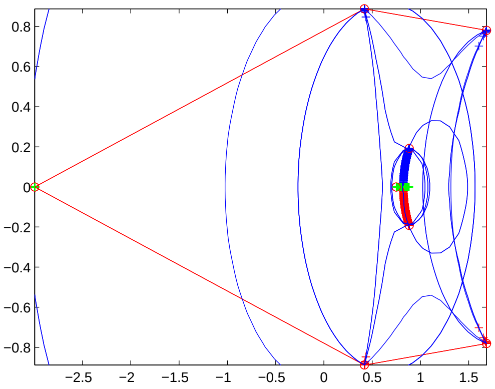

FIG. 7.8. *Location of the Ritz values for Example 7, $n = 8$, $k = 7$, A normal real, Arnoldi with random real vectors v.*

Numerical experiments not reported here seem to show that the properties described above for the Arnoldi Ritz values are not restricted to the Arnoldi algorithm. For a real normal matrix, if one constructs a real orthogonal matrix V and defines $H = V^T AV$, the Ritz values, being defined as the eigenvalues of H_k, the principal submatrix of order k of H, are also constrained in some regions inside the field of values of A. This deserves further studies.

8. Conclusion. In this paper we gave a necessary and sufficient condition for a set of complex values $\theta_1, \dots, \theta_k$ to be the Arnoldi Ritz values at iteration k for a general diagonalizable matrix A. This generalizes previously known conditions. The condition stated in this paper simplifies for normal matrices and particularly for real normal matrices and real starting vectors. We studied the case $k = 2$ in detail, for which we characterized the boundary of the region in the complex plane contained in $W(A)$, where pairs of complex conjugate Ritz values are located. Several examples with a computation of the boundary of the feasible region were given. Finally, after describing some numerical experiments with random real starting vectors, we formulated some conjectures and open problems for $k > 2$ for real normal matrices.

Acknowledgments. The author thanks J. Duintjer Tebbens for some interesting comments and the referees for remarks that helped improve the presentation.

REFERENCES

[1] M. BELLALIJ, Y. SAAD, AND H. SADOK, *On the convergence of the Arnoldi process for eigenvalue problems*, Tech. Report umsi-2007-12, Minnesota Supercomputer Institute, University of Minnesota, 2007.
[2] ———, *Further analysis of the Arnoldi process for eigenvalue problems*, SIAM J. Numer. Anal., 48 (2010), pp. 393–407.
[3] Z. BUJANOVIĆ, *On the permissible arrangements of Ritz values for normal matrices in the complex plane*, Linear Algebra Appl., 438 (2013), pp. 4606–4624.
[4] R. CARDEN, *Ritz Values and Arnoldi Convergence for Non-Hermitian Matrices*, PhD. Thesis, Computational and Applied Mathematics, Rice University, Houston, 2011.
[5] ———, *A simple algorithm for the inverse field of values problem*, Inverse Problems, 25 (2009), 115019 (9 pages).
[6] R. CARDEN AND M. EMBREE, *Ritz value localization for non-Hermitian matrices*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 1320–1338.
[7] R. CARDEN AND D. J. HANSEN, *Ritz values of normal matrices and Ceva's theorem*, Linear Algebra Appl., 438 (2013), pp. 4114–4129.
[8] C. CHORIANOPOULOS, P. PSARRAKOS, AND F. UHLIG, *A method for the inverse numerical range problem*, Electron. J. Linear Algebra, 20 (2010), pp. 198–206.
[9] J. DUINTJER TEBBENS AND G. MEURANT, *Any Ritz value behavior is possible for Arnoldi and for GMRES*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 958–978.
[10] ———, *On the convergence of QOR and QMR Krylov methods for solving linear systems*, in preparation.
[11] I. C. F. IPSEN, *Expressions and bounds for the GMRES residual*, BIT, 40 (2000), pp. 524–535.
[12] A. KOVAČEC AND B. RIBEIRO, *Convex hull calculations: a Matlab implementation and correctness proofs for the lrs-algorithm*, Tech. Report 03-26, Department of Mathematics, Coimbra University, Coimbra, Portugal, 2003.
[13] R. B. LEHOUCQ, D. C. SORENSEN, AND C.-C. YANG, *Arpack User's Guide*, SIAM, Philadelphia, 1998.
[14] S. M. MALAMUD, *Inverse spectral problem for normal matrices and the Gauss-Lucas theorem*, Trans. Amer. Math. Soc., 357 (2005), pp. 4043–4064.
[15] G. MEURANT, *GMRES and the Arioli, Pták, and Strakoš parametrization*, BIT, 52 (2012), pp. 687–702.
[16] ———, *The computation of isotropic vectors*, Numer. Algorithms, 60 (2012), pp. 193–204
[17] C. C. PAIGE, *The Computation of Eigenvalues and Eigenvectors of Very Large Sparse Matrices*, PhD. Thesis, Institute of Computer Science, University of London, London, 1971.
[18] ———, *Computational variants of the Lanczos method for the eigenproblem*, J. Inst. Math. Appl., 10 (1972), pp. 373–381.
[19] ———, *Accuracy and effectiveness of the Lanczos algorithm for the symmetric eigenproblem*, Linear Algebra Appl., 34 (1980), pp. 235–258.
[20] B. N. PARLETT, *Normal Hessenberg and moment matrices*, Linear Algebra Appl., 6 (1973), pp. 37–43.

ETNA
Kent State University
http://etna.math.kent.edu

[21] ———, *The Symmetric Eigenvalue Problem*, Prentice Hall, Englewood Cliffs, 1980.

[22] H. SADOK, *Analysis of the convergence of the minimal and the orthogonal residual methods*, Numer. Algorithms, 40 (2005), pp. 201–216.