The distance of an eigenvector to a Krylov subspace and the convergence of the Arnoldi method for eigenvalue problems ☆

M. Bellalij

Université de Valenciennes, Le Mont Houy, 59313 Valenciennes Cedex, France

G. Meurant

30 rue du sergent Bauchat, 75012 Paris

H. Sadok

L.M.P.A, Université du Littoral, 50 rue F. Buisson, BP699, F-62228 Calais Cedex, France

Abstract

We study the distance of an eigenvector of a diagonalizable matrix A to the Krylov subspace generated from A and a given starting vector v. This distance is involved in studies of the convergence of the Arnoldi method for computing eigenvalues. Contrary to the previous studies on this problem, we provide closed-form expressions for this distance in terms of the eigenvalues and eigenvectors of A as well as the components of v in the eigenvector basis. The formulas simplify when the matrix A is normal. For A non-normal we derive upper and lower bounds that are simpler than the exact expressions. We also show how to generate starting vectors such that the distance to the Krylov subspace is equal to the worst possible case.

Keywords: Krylov subspace, Arnoldi method, Ritz values, eigenvalues, eigenvectors, diagonalizable matrices, normal matrices

1. Introduction

Some of the most important tools to compute approximate eigenvalues and eigenvectors of large sparse matrices A use projection techniques on Krylov subspaces. A well-known and widely used technique of this type is the Arnoldi method [1] developed by Saad [13]. It is an orthogonal projection method. The convergence of this kind of methods has been analyzed in detail in the Hermitian

☆Version December 14, 2015

Email addresses: Mohammed.Bellalij@univ-valenciennes.fr (M. Bellalij), gerard.meurant@gmail.com (G. Meurant), sadok@lmpa.univ-littoral.fr (H. Sadok)

case, but the analysis is much more difficult in the general non symmetric case and so there are not many results available in the literature; see [14] [15], [8], [3]. Most analyses involve the distance of an eigenvector of the matrix A to the Krylov subspace generated from A and a given starting vector v. Usually some upper bounds are derived for this distance to obtain some insights about the convergence behavior of the numerical method. In this paper we consider this problem from another view point. We derive closed-form expressions for this distance when the matrix A is diagonalizable. The distance is given in terms of the eigenvalues and eigenvectors of A as well as the components of v in the eigenvector basis. In particular, this yields bounds for the norm of the residual given by an exact eigenpair of A for the approximate problem generated by the Arnoldi method; see [3]. We hope that these new exact results can help obtaining more insights into the convergence of the Arnoldi method.

The contents of the paper are as follows. In section 2 we recall results about the Arnoldi method and its convergence. Section 3 studies the minimum distance of a given vector to a subspace of $\mathbb{C}^N$. In section 4 we consider the distance of a given eigenvector of A to a Krylov subspace of dimension k generated from a diagonalizable matrix A and a vector v. We obtain a closed-form expression in terms of the eigenvalues, the eigenvectors and the components of the starting vector in the eigenvector basis. The formulas simplify when the matrix A is normal. Bounds that are simpler than the exact formula are also given in the non-normal case. In section 5 we study whether we can find starting vectors v such that the distance of the eigenvector to the Krylov subspace is equal to 1 at a given step of the Arnoldi method. Finally, we give some conclusions.

Exact arithmetic is assumed throughout the paper. The conjugate transposes of a vector x or a matrix A are denoted by x^* and A^*. The Euclidean norm of a vector v is denoted by $\|v\|$, so $\|v\| = \sqrt{(v^*v)}$. For a matrix A, $\|A\|$ denotes the 2-norm. We denote by I the identity matrix of appropriate order and by e_k its kth column.

In this paper we will make use of the following hypothesis on the matrix A of order N and the Arnoldi starting vector v,

Hypothesis H: Let $k < N$. The matrix A is diagonalizable as $A = X\Lambda X^{-1}$ where X is the matrix of the normalized eigenvectors and Λ is the diagonal matrix of the eigenvalues denoted as λ_i, $i = 1, \ldots, N$. We assume that there are at least $k+1$ distinct eigenvalues (which we number from 1 to $k+1$) and that the first $k+1$ components of $\alpha = X^{-1}v$ are different from zero.

2. The eigenvalue problem

Let A be a square real or complex matrix of order N. Consider the eigenvalue problem: find λ belonging to $\mathbb{C}$ and x belonging to $\mathbb{C}^N$ such that

$$A\,x = \lambda x. \tag{1}$$

For a given vector $v \in \mathbb{C}^N$ of unit norm, the Krylov subspace $\mathcal{K}_k(A, v)$ is defined by

$$\mathcal{K}_k(A, v) = \text{span}\{v, A\,v, \ldots, A^{k-1}v\}. \tag{2}$$

We denote by K_k the Krylov matrix whose columns are the natural basis vectors of $\mathcal{K}_k(A, v)$ that is, $A^j v$, $j = 0, \ldots, k-1$. We assume that $k < N$ and that, according to hypothesis H, the matrix K_k is of full rank k. Throughout the paper we will denote by M_k the Gram matrix $K_k^* K_k$.

2.1. The Arnoldi method

The Arnoldi method computes approximate eigenpairs $\lambda^{(k)}, x^{(k)}$ with $x^{(k)} \in \mathcal{K}_k(A, v)$ by enforcing the Galerkin condition

$$(A\, x^{(k)} - \lambda^{(k)}\, x^{(k)}\, , A^i v) = 0 \qquad \text{for} \quad i = 0, \ldots, k-1. \tag{3}$$

Relation (3) means that the residual of the approximate eigenpair is orthogonal to the Krylov subspace of dimension k. The standard way of extracting the approximate eigenpairs from the above conditions is to use the Arnoldi process [1] which generates an orthonormal basis $v_1, \ldots, v_k$ of $\mathcal{K}_k(A, v)$ with which the relations (3) are expressed. The outputs of the algorithm at iteration k are a unitary matrix $V_k = [v_1, v_2, \ldots, v_k]$ and an unreduced upper Hessenberg matrix H_k whose entries are the scalars $h_{i,j}$ generated by the procedure. In addition, the following relations are satisfied:

- $AV_k = V_k H_k + h_{k+1,k} v_{k+1} {e_k}^T$,
- $H_k = V_k^* A V_k$.

The approximate eigenvalue problem can now be written as

$$V_k^*\, A V_k\, y^{(k)} = H_k y^{(k)} = \lambda^{(k)}\, y^{(k)}, \tag{4}$$

which is equivalent to

$$V_k^* (A - \lambda^{(k)} I) V_k y^{(k)} = 0.$$

Approximate eigenvalues (the so-called Ritz values) are the eigenvalues of H_k and the associated approximate eigenvectors are $x^{(k)} = V_k y^{(k)}$.

2.2. Convergence

The simplest analysis of the Arnoldi method uses the distance of a given eigenvector of A to the Krylov subspace. It is based on the viewpoint described below; see [15]. Let $\mathcal{P}_k$ be the orthogonal projector onto $\mathcal{K}_k(A, v)$. It can be written as $K_k (K_k^* K_k)^{-1} K_k^*$ or $V_k V_k^*$ depending on the basis we consider. Then, the approximate eigenvalue problem amounts to solving

$$\mathcal{P}_k (Ax - \lambda x) = 0, \quad x \ \in \mathcal{K}_k(A, v),$$

or in operator form

$$\mathcal{P}_k A \mathcal{P}_k x = \lambda x.$$

Let us define, $A_k \equiv \mathcal{P}_k A \mathcal{P}_k$. Note that $A_k = V_k H_k V_k^*$. The following theorem is proved in [15].

Theorem 1. *Let x be an eigenvector of A associated with the eigenvalue λ and $\gamma_k = \|\mathcal{P}_k A(I-\mathcal{P}_k)\|$. Then the residual norms of the pairs $(\lambda, \mathcal{P}_k x)$ and (λ, x) for the linear operator A_k satisfy,*

$$\|(A_k - \lambda I)\mathcal{P}_k x\| \leq \gamma_k \|(I-\mathcal{P}_k)x\|, \tag{5}$$

$$\|(A_k - \lambda I)x\| \leq \sqrt{|\lambda|^2 + \gamma_k^2}\ \|(I-\mathcal{P}_k)x\|. \tag{6}$$

Note that $\gamma_k \leq \|A\|$. Therefore, the coefficients on the right-hand sides of these inequalities are at most of the order of $\|(I-\mathcal{P}_k)x\|$. Theorem 1 states how accurate the *exact eigenpair* is with respect to the *approximate problem*. This is stated in terms of the distance of the exact eigenvector x to the Krylov subspace. The remaining issue is to compute or to estimate $\|(I-\mathcal{P}_k)x\|$. Note that $I-\mathcal{P}_k$ is the projector onto the orthogonal complement of the Krylov subspace. Upper bounds for the norm have been obtained in [3] using several tools: the eigenvectors, the Schur vectors and an approximation theory viewpoint. We remark that $\|(I-\mathcal{P}_k)x\|$ is also involved in the analysis of convergence of harmonic Ritz values and harmonic and refined Ritz vectors; see [9], [4], [10], [11]. In the following we will obtain closed-form expressions for $\|(I-\mathcal{P}_k)x\|$ when A satisfies hypothesis H.

3. The minimum distance to a subspace

In this section we analyze the minimum distance $\min_{x\,\in\,\mathcal{X}} \|w-x\|$ of a vector w to $\mathcal{X}$ which is an arbitrary subspace of dimension k. We begin by showing some simple results. First, observe that given *any* matrix W whose columns give a basis of the subspace $\mathcal{X}$, $x \in \mathcal{X}$ can be written as Wy, where $y \in \mathbb{C}^k$. Hence,

$$\|w-x\|^2 = \|w-Wy\|^2 = w^*w - w^*Wy - y^*W^*w + y^*W^*Wy.$$

The above expression is of the form

$$\|w-x\|^2 = \begin{pmatrix} 1 \\ -y \end{pmatrix}^* \underbrace{\begin{pmatrix} w^*w & w^*W \\ W^*w & W^*W \end{pmatrix}}_{\equiv C} \begin{pmatrix} 1 \\ -y \end{pmatrix}. \tag{7}$$

Note that minimizing $\|w-x\|$ over $\mathcal{X}$ is equivalent to minimizing $\|w+x\|$ over the same subspace, so the signs of the y's in the expression above can be changed when seeking the minimum distance. In the end,

$$\min_{x\,\in\,\mathcal{X}} \|w-x\|^2 = \min_{y\,\in\,\mathbb{C}^k} \|w-Wy\|^2 = \min_{y\,\in\,\mathbb{C}^k} \begin{pmatrix} 1 \\ y \end{pmatrix}^* C \begin{pmatrix} 1 \\ y \end{pmatrix}, \tag{8}$$

where C is defined in relation (7). Note that if we assume that $w \notin \mathcal{X}$ the matrix C is non singular since $C = [w,W]^*[w,W]$. Then we can characterize the distance of a vector to its orthogonal projection.

Lemma 2. *Let $\mathcal{X}$ be an arbitrary subspace of dimension k in $\mathbb{C}^N$ with a basis $W = [w_1, \cdots, w_k]$ and let $w \notin \mathcal{X}$. Let $\mathcal{P}$ be the orthogonal projector onto $\mathcal{X}$. Then, we have*

$$\|(I - \mathcal{P})w\|^2 = \frac{1}{e_1^T C^{-1} e_1}$$

where

$$C = \begin{pmatrix} w^*w & w^*W \\ W^*w & W^*W \end{pmatrix}.$$

Proof. Given an arbitrary vector $w \in \mathbb{C}^N$, observe that

$$\|(I - \mathcal{P})w\|^2 = w^*(I - \mathcal{P})(I - \mathcal{P})w = w^*(I - \mathcal{P})w = w^*w - w^*\mathcal{P}w,$$

with $\mathcal{P} = W(W^*W)^{-1}W^*$. It follows that

$$\|(I - \mathcal{P})w\|^2 = w^*w - w^*\mathcal{P}w = w^*w - w^*W(W^*W)^{-1}W^*w. \tag{9}$$

The right-hand side of relation (9) is the Schur complement of the (1,1) entry of C which, as is well-known, is the inverse of the (1,1) entry of C^{-1}.

4. The minimum distance of an eigenvector to a Krylov subspace

Let A be a diagonalizable matrix satisfying hypothesis H. We study the convergence to a given simple eigenvalue which is indexed by 1, that is, we consider x_1, the first column of X corresponding to λ_1. We have assumed that $[X^{-1}v]_1 \neq 0$ and $\|x_j\| = 1$ for all j. Note that the distance $\|(I - \mathcal{P}_k)x_1\|$ is nothing but the sine of the angle $\angle(x_1, \mathcal{K}_k)$. Bounds for the distance were given in [15] and in [8] in the case where A is not diagonalizable but non-derogatory.

In the first subsection below we will obtain explicit expressions for the minimum distance of the eigenvector x_1 to a Krylov subspace when the matrix is diagonalizable and satisfies hypothesis H. These exact expressions are quite intricate, particularly when A is non-normal. Therefore, in the second subsection we will derive lower and upper bounds for the minimum distance in the non-normal case. These bounds are simpler than the exact results and almost similar to what we have in the normal case.

4.1. An exact expression for the minimum distance

An explicit formula for $\|(I - \mathcal{P}_k)x_1\|$ is obtained immediately as a corollary of Lemma 2.

Corollary 3. *Let A be a general square matrix, $x_1 \notin \mathcal{K}_k(A, v)$ be an eigenvector of A and L_{k+1} be the rectangular matrix of $\mathbb{C}^{N\times(k+1)}$,*

$$L_{k+1} = \begin{pmatrix} \alpha_1\, x_1, & v, & A\,v, & \ldots, & A^{k-1}\,v \end{pmatrix}.$$

We assume that $\alpha \in \mathbb{C}$, $\alpha_1 \neq 0$ and L_{k+1} is of rank $k+1$. Then

$$\|(I - \mathcal{P}_k)\,\alpha_1\,x_1\|^2 = \frac{1}{e_1^T\,(L_{k+1}^* L_{k+1})^{-1}\,e_1}. \tag{10}$$

PROOF. The result follows by applying Lemma 2 with $w = \alpha_1 x_1$, and $W = [v, Av, \cdots, A^{k-1}v]$.

The reason for introducing the factor α_1 in the first column will be given in the proof of the next theorem. Note that the result in Corollary 3 is valid whatever the matrix A and the complex number $\alpha_1 \neq 0$ as long as the matrix L_{k+1} is of rank $k+1$. In the following theorem we consider a matrix satisfying hypothesis H.

Theorem 4. *Let A be a diagonalizable matrix satisfying hypothesis H, α a vector of $\mathbb{C}^k$ with components α_j such that the starting vector is $v = X\alpha$ and the two matrices*

$$D_\alpha = \begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \alpha_N \end{pmatrix} \quad \textit{and } W_{k+1} = \begin{pmatrix} 1 & 1 & \lambda_1 & \dots & \lambda_1^{k-1} \\ 0 & 1 & \lambda_2 & \dots & \lambda_2^{k-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 1 & \lambda_i & \dots & \lambda_i^{k-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 1 & \lambda_N & \dots & \lambda_N^{k-1} \end{pmatrix}. \quad (11)$$

Let $\mathcal{P}_k$ be the orthogonal projector onto $\mathcal{K}_k(A, v)$ and $\tilde{M}_{k+1} = L_{k+1}^* L_{k+1}$, where L_{k+1} is defined in Corollary 3. Then,

$$\|(I - \mathcal{P}_k)\,\alpha_1\, x_1\|^2 = \frac{1}{(\tilde{M}_{k+1}^{-1})_{1,1}} = \frac{1}{e_1^T \left(W_{k+1}^* D_\alpha^* (X^*X) D_\alpha W_{k+1}\right)^{-1} e_1}. \quad (12)$$

When A is normal the preceding formula simplifies to

$$\|(I - \mathcal{P}_k)\,\alpha_1\, x_1\|^2 = \frac{1}{e_1^T \left(W_{k+1}^* D_\alpha^* D_\alpha W_{k+1}\right)^{-1} e_1}. \quad (13)$$

PROOF. We consider a factorization of the matrix L_{k+1} when A is satisfies hypothesis H. Using the definition of α, we obtain $A^j v = X\,\Lambda^j \alpha$ for $j = 0, \dots, k-1$. Note that we have assumed that $\alpha_1 \neq 0$. Then, we have

$$\begin{aligned} L_{k+1} &= [\alpha_1 x_1, v, A\,v, \dots, A^{k-1}\,v] \\ &= X\,[\alpha_1\, e_1, \alpha, \Lambda\,\alpha, \dots, \Lambda^{k-1}\,\alpha] \\ &= X\,D_\alpha\,W_{k+1}, \end{aligned}$$

With hypothesis H the matrix L_{k+1} is of rank $k+1$. Note that α_1 was introduced as a factor of the first column of L_{k+1} to obtain the matrix D_α in the factorization. As a result it yields,

$$\tilde{M}_{k+1} = W_{k+1}^*\, D_\alpha^*\, (X^*X)\, D_\alpha\, W_{k+1}, \quad (14)$$

which is a non singular matrix. We remark that if the matrix A is normal the term within parenthesis is equal to the identity matrix. With Corollary 3, we obtain the formulas (12) and (13).

Our next goal is to express the $(1,1)$ entry of the inverse of $\tilde{M}_{k+1}$ which appears in relation (12) in terms of the eigenvalues, the eigenvectors of A and the components of α. To achieve this we use the same techniques as in [5] and [12] that is, Cramer's rule and the Cauchy-Binet formula for determinants; see [7]. Then

$$(\tilde{M}_{k+1}^{-1})_{1,1} = \frac{\det(\hat{M}_{k+1})}{\det(\tilde{M}_{k+1})},$$

where $\hat{M}_{k+1}$ is equal to $\tilde{M}_{k+1}$ except for the first column which is replaced by the first column of the identity matrix. Of course, the determinant of $\hat{M}_{k+1}$ is equal to the determinant of its submatrix with rows $[2,\ldots,k+1]$ and columns $[2,\ldots,k+1]$ which can be written as $(\tilde{M}_{k+1})_{[2:k+1],[2:k+1]}$. It yields

$$\begin{aligned}
(\tilde{M}_{k+1})_{[2:k+1],[2:k+1]} &= \begin{pmatrix} 0 & \\ \vdots & I_k \\ 0 & \end{pmatrix} \tilde{M}_{k+1} \begin{pmatrix} 0 & \cdots & 0 \\ & I_k & \end{pmatrix}, \\
&= \begin{pmatrix} 0 & \\ \vdots & I_k \\ 0 & \end{pmatrix} W_{k+1}^* \, D_\alpha^*(X^*X) D_\alpha W_{k+1} \begin{pmatrix} 0 & \cdots & 0 \\ & I_k & \end{pmatrix}, \\
&= \mathcal{V}_k^* \, D_\alpha^*(X^*X) D_\alpha \mathcal{V}_k,
\end{aligned}$$

where

$$\mathcal{V}_k = \begin{pmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{k-1} \\ 1 & \lambda_2 & \cdots & \lambda_2^{k-1} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_N & \cdots & \lambda_N^{k-1} \end{pmatrix},$$

is a rectangular Vandermonde matrix. The matrix

$$(\tilde{M}_{k+1})_{[2:k+1],[2:k+1]} = \mathcal{V}_k^* \, D_\alpha^*(X^*X) D_\alpha \mathcal{V}_k$$

is equal to the Gram matrix $M_k = K_k^* K_k$. A closed-form expression for its determinant was already given in [6].

Lemma 5 ([6]). *Let A be a diagonalizable matrix with a spectral decomposition $A = X\Lambda X^{-1}$ and let K_k be the Krylov matrix based on A and v. Then, the determinant of $M_k = K_k^* K_k$ is*

$$\det(M_k) = \sum_{I_k} \left| \sum_{J_k} \det(X_{I_k,J_k}) \alpha_{j_1} \cdots \alpha_{j_k} \prod_{\substack{j_1 \le j_p < j_q \le j_k \\ j_p, j_q \in J_k}} (\lambda_{j_q} - \lambda_{j_p}) \right|^2,$$

where the summations are over all sets of indices I_k and J_k defined as I_ℓ to be a set of ℓ indices $\big(i_1, \quad i_2, \quad \ldots, \quad i_\ell\big)$ such that $1 \le i_1 < \cdots < i_\ell \le N$ and

J_ℓ is similar with i replaced by j, X_{I_ℓ,J_ℓ} is the submatrix of X whose row and column indices of entries are defined respectively by I_ℓ and J_ℓ and $\alpha = X^{-1}v$. The product is on all pairs of indices that belong to J_k (the product being equal to 1 if $k = 1$).

If the matrix A is normal then

$$\det(M_k) = \sum_{I_k} \left[\prod_{j=1}^{k} |\alpha_{i_j}|^2 \right] \prod_{\substack{i_1 \le i_p < i_q \le i_k \\ i_p, iq \in I_k}} |\lambda_{i_q} - \lambda_{i_p}|^2,$$

with $\alpha = X^* v$.

Now we consider the determinant of $\tilde{M}_{k+1} = W_{k+1}^* \, D_\alpha^* (X^* X) D_\alpha W_{k+1}$.

Lemma 6. *Let A be a diagonalizable matrix with a spectral decomposition $A = X \Lambda X^{-1}$. Then, the determinant of $\tilde{M}_k = L_{k+1}^* \, L_{k+1}$ is*

$$\det(\tilde{M}_k) = |\alpha_1|^2 \sum_{I_{k+1}} \left| \sum_{\hat{J}_{k+1}} \det(X_{I_{k+1},\hat{J}_{k+1}}) \, \alpha_{j_2} \cdots \alpha_{j_{k+1}} \prod_{\substack{1 < j_2 \le j_p < j_q \le j_{k+1} \\ j_p, j_q \in \hat{J}_{k+1}}} (\lambda_{j_q} - \lambda_{j_p}) \right|^2,$$

where I_{k+1} is defined as in Lemma 5 and the summation with $\hat{J}_{k+1}$ is over all sets of indices $\{1, j_2, \ldots, j_{k+1}\}$ such that $1 < j_2 < \cdots < j_{k+1} \le N$.

If the matrix A is normal then

$$\det(\tilde{M}_k) = |\alpha_1|^2 \sum_{\hat{I}_{k+1}} \left[\prod_{j=2}^{k+1} |\alpha_{i_j}|^2 \right] \prod_{\substack{1 < i_2 \le i_p < i_q \le i_{k+1} \\ i_p, i_q \in \hat{I}_{k+1}}} |\lambda_{i_q} - \lambda_{i_p}|^2,$$

and the summation with $\hat{I}_{k+1}$ is over all sets of indices $\{i_2, \ldots, i_{k+1}\}$ such that $1 < i_2 < \cdots < i_{k+1} \le N$.

Proof. Let $G = X D_\alpha W_{k+1}$, then $\tilde{M}_{k+1} = G^* G$ is the product of two rectangular matrices. To compute the determinant we apply the Cauchy-Binet formula; see [7]. It yields

$$\det(\tilde{M}_k) = \sum_{I_{k+1}} |\det(G_{I_{k+1},:})|^2,$$

where I_{k+1} is defined as in Lemma 5 and $G_{I_{k+1},:}$ is the submatrix of G whose row indices are in I_{k+1}. The matrix G itself is the product of two matrices, $X D_\alpha$ and W_{k+1}. For the determinant of $G_{I_{k+1},:}$ we apply again the Cauchy-Binet formula and we obtain

$$\det(G_{I_{k+1},:}) = \sum_{J_{k+1}} \det(X_{I_{k+1},J_{k+1}}) \, \alpha_{j_1} \cdots \alpha_{j_{k+1}} \, \det(\mathcal{W}(j_1, \ldots, j_{k+1})),$$

where $\mathcal{W}(j_1,\ldots,j_{k+1})$ is obtained from the rows $j_1,\ldots,j_{k+1}$ of W_{k+1}. From the structure of the first column of W_{k+1} it is clear that $\det(\mathcal{W}(j_1,\ldots,j_{k+1}))=0$ if $1\notin\{j_1,j_2,\ldots,j_{k+1}\}$. But the indices in J_{k+1} are strictly ordered, so the determinant is different from zero only if $j_1=1$. The sum over the sets J_{k+1} reduces to a sum over sets of indices $\hat{J}_{k+1}$ which are $\{1,j_2,\ldots,j_{k+1}\}$ with $1<j_2<\cdots<j_{k+1}\leq N$. Clearly for these sets of indices we have

$$\det(\mathcal{W}(1,j_2,\ldots,j_{k+1}))=\det\begin{pmatrix}1 & \lambda_{j_2} & \cdots & \lambda_{j_2}^{k-1}\\ 1 & \lambda_{j_3} & \cdots & \lambda_{j_3}^{k-1}\\ \vdots & \vdots & & \vdots\\ 1 & \lambda_{j_{k+1}} & \cdots & \lambda_{j_{k+1}}^{k-1}\end{pmatrix}=\prod_{1<j_2\leq j_p<j_q\leq j_{k+1}}(\lambda_{j_q}-\lambda_{j_p}).$$

The indices j_p,j_q have to belong to the set $\{j_2,j_3,\ldots,j_{k+1}\}$. It yields the formula for $\det(\tilde{M}_k)$

When A is normal, $G=D_\alpha W_{k+1}$ and we do not have to apply the Cauchy-Binet formula a second time to obtain $\det(\tilde{M}_k)$.

From Lemmas 5 and 6 we obtain the following closed-form expression for the distance of an eigenvector to its orthogonal projection on the Krylov subspace.

Theorem 7. *Let A be a diagonalizable matrix satisfying hypothesis H. The distance of the eigenvector x_1 to the Krylov subspace $\mathcal{K}_k(A,v)$ is given by*

$$\|(I-\mathcal{P}_k)\,x_1\|^2=\frac{N}{D},\tag{15}$$

with

$$N=\sum_{I_{k+1}}\left|\sum_{\hat{j}_{k+1}}\det(X_{I_{k+1},\hat{j}_{k+1}})\,\alpha_{j_2}\cdots\alpha_{j_{k+1}}\prod_{\substack{1<j_2\leq j_p<j_q\leq j_{k+1}\\ j_p,j_q\in\hat{J}_{k+1}}}(\lambda_{j_q}-\lambda_{j_p})\right|^2,$$

$$D=\sum_{I_k}\left|\sum_{J_k}\det(X_{I_k,J_k})\alpha_{j_1}\cdots\alpha_{j_k}\prod_{\substack{j_1\leq j_p<j_q\leq j_k\\ j_p,j_q\in J_k}}(\lambda_{j_q}-\lambda_{j_p})\right|^2,$$

where the summation sets of indices are defined in Lemmas 5 and 6 and $\alpha=X^{-1}v$.

If, in addition, A is normal we have

$$N=\sum_{\hat{I}_{k+1}}\left[\prod_{j=2}^{k+1}|\alpha_{i_j}|^2\right]\prod_{\substack{1<i_2\leq i_p<i_q\leq i_{k+1}\\ i_p,i_q\in\hat{I}_{k+1}}}|\lambda_{i_q}-\lambda_{i_p}|^2,$$

$$D = \sum_{I_k} \left[\prod_{j=1}^{k} |\alpha_{i_j}|^2 \right] \prod_{\substack{i_1 \le i_p < i_q \le i_k \\ i_p, i_q \in I_k}} |\lambda_{i_q} - \lambda_{i_p}|^2,$$

with $\alpha = X^* v$.

Formula (15) is difficult to interpret in the non-normal case since there is a strong dependence on the eigenvectors through determinants of submatrices of X. The normal case is easier since the distance depends on the eigenvectors only through the components of $X^* v$.

The results of Theorem 7 for normal matrices can be written in a different way. The summation in the denominator D is over all sets of k indices $I_k = \{i_1, i_2, \ldots, i_k\}$ with $1 \le i_1 < i_2 < \cdots < i_k \le N$. Such sets can be split into two disjoint sets

$$I_k = \mathcal{I}_1 \cup \mathcal{I}_k, \quad \mathcal{I}_1 = \{1, i_2, \ldots, i_k\}, \quad \mathcal{I}_k = \{i_1, i_2, \ldots, i_k, \,|\, i_1 > 1\}.$$

Let us denote the sums over these two sets of indices by $S_{\mathcal{I}_1}$ and $S_{\mathcal{I}_k}$. We have

$$S_{\mathcal{I}_1} = |\alpha_1|^2 \sum_{\substack{\{i_2,\ldots,i_k\} \\ i_2 > 1}} \left[\prod_{j=2}^{k} |\alpha_{i_j}|^2 \right] \prod_{\substack{1 \le i_p < i_q \le i_k \\ i_p, i_q \in \{1, i_2, \ldots, i_k\}}} |\lambda_{i_q} - \lambda_{i_p}|^2, \tag{16}$$

$$S_{\mathcal{I}_k} = \sum_{\mathcal{I}_k} \left[\prod_{j=1}^{k} |\alpha_{i_j}|^2 \right] \prod_{\substack{1 < i_1 \le i_p < i_q \le i_k \\ i_p, i_q \in \mathcal{I}_k}} |\lambda_{i_q} - \lambda_{i_p}|^2.$$

Even though the notation is different, $S_{\mathcal{I}_k}$ is equal to the numerator N. Therefore we can write

$$\frac{N}{D} = \frac{S_{\mathcal{I}_k}}{S_{\mathcal{I}_1} + S_{\mathcal{I}_k}} = \frac{1}{1 + \frac{S_{\mathcal{I}_1}}{S_{\mathcal{I}_k}}}, \tag{17}$$

as long as $S_{\mathcal{I}_k} \neq 0$. The eigenvalue of interest λ_1 does not appear in $S_{\mathcal{I}_k}$. We see that the distance of x_1 to the Krylov subspace is equal to 1 if and only if $S_{\mathcal{I}_1} = 0$. It is small if $S_{\mathcal{I}_1}/S_{\mathcal{I}_k}$ is large. Of course, this happens if $S_{\mathcal{I}_1}$ is large or if $S_{\mathcal{I}_k}$ is small or both. Unfortunately, we cannot do the same manipulation for the non-normal case because of the term $\det(X_{I_k,J_k})$.

Let us give a small example to clarify all this and to show what kind of insights one can obtain from the previous formulas. We consider a normal matrix A of order 4 with eigenvalues $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ and $k = 2$. The sets of indices in I_3 are $(1,2,3)$, $(1,2,4)$, $(1,3,4)$. Therefore, the sets of indices in $\hat{I}_3$ are $(2,3)$, $(2,4)$, $(3,4)$. The sets of indices in I_2 are

$$(1,2),\ (1,3),\ (1,4),\ (2,3),\ (2,4),\ (3,4).$$

The three first pairs are the pairs in $\mathcal{I}_1$ and the four last ones are in $\mathcal{I}_2$. We see that $\mathcal{I}_2$ is identical to $\hat{I}_3$. Let $\beta_i = |\alpha_i|^2$. We have

$$\begin{aligned} S_{\mathcal{I}_1} &= \beta_1 \left[\beta_2|\lambda_2 - \lambda_1|^2 + \beta_3|\lambda_3 - \lambda_1|^2 + \beta_4|\lambda_4 - \lambda_1|^2\right], \\ S_{\mathcal{I}_2} &= \beta_2\beta_3|\lambda_3 - \lambda_2|^2 + \beta_2\beta_4|\lambda_4 - \lambda_2|^2 + \beta_3\beta_4|\lambda_4 - \lambda_3|^2. \end{aligned}$$

The term $S_{\mathcal{I}_1}$ will be large if at least one of the other eigenvalue is "far" from λ_1 and the projection of v on the corresponding eigenvector is not too small. The other term $S_{\mathcal{I}_k}$ is small if the products of the pairwise distances between the other eigenvalues with the moduli of the projections of v are small. If only one of the terms in the sum is not small, $S_{\mathcal{I}_k}$ cannot be small.

Let us first assume that we have a cluster of three distinct eigenvalues $\lambda_2, \lambda_3, \lambda_4$ whose small pairwise distances are of order ε and another complex eigenvalue λ_1 whose pairwise distances to the three other ones are of order 1. Assume also that all the β_i's are non zero and that no one is very small. Then $S_{\mathcal{I}_2} = c\varepsilon^2$ where $c \gg \varepsilon$ and the ratio N/D is equal to

$$\frac{1}{1 + \frac{S_{\mathcal{I}_1}}{c\varepsilon^2}} = \frac{c\varepsilon^2}{c\varepsilon^2 + S_{\mathcal{I}_1}} = \mathcal{O}(\varepsilon^2),$$

since $S_{\mathcal{I}_1}$ is of order 1. We see that with this distribution of eigenvalues $\|(I - \mathcal{P}_2)\, x_1\|$ is small of order ε.

Let us now assume that A is real. Then the eigenvalues arise as real numbers or complex conjugate pairs. Let λ_1 be a complex eigenvalue with $\lambda_2 = \overline{\lambda_1}$. Assume that λ_3 and λ_4 are complex conjugate or real with $|\lambda_4 - \lambda_3| = \varepsilon$ and the distances to λ_1 and λ_2 are of order 1. Then, $S_{\mathcal{I}_2}$ is not small unless β_2 is small. Generally, the ratio N/D is not small. We see that even though we have a tight cluster in both situations, the outcome may not be the same when A is real or complex. But, of course, the number of eigenvalues in the cluster is not the same.

To illustrate this we consider two small numerical examples. The first normal matrix is

$$A = \begin{pmatrix} 1.968 + 0.9696i & -0.1049 - 0.1038i & 0.1235 + 0.1279i & -0.05698 - 0.04881i \\ -0.1049 - 0.1038i & 1.646 + 0.6459i & 0.4355 + 0.4361i & -0.1671 - 0.1665i \\ 0.1235 + 0.1279i & 0.4355 + 0.4361i & 1.461 + 0.4628i & 0.204 + 0.205i \\ -0.05698 - 0.04881i & -0.1671 - 0.1665i & 0.204 + 0.205i & 1.926 + 0.9217i \end{pmatrix},$$

and the starting vector is

$$v = \begin{pmatrix} -0.04974 \\ 0.5969 \\ 0.5646 \\ 0.5679 \end{pmatrix}.$$

The eigenvalues of A are

$$\begin{pmatrix} 1, & 1.99 - i, & 2.01 - i, & 2 - i \end{pmatrix}.$$

It means that we are in the first case since we have a complex matrix with a cluster of three eigenvalues and the first eigenvalue is far from the cluster. The first eigenvector of A and its projection on the Krylov subspace of dimension 2 (computed as $\mathcal{P}_2 = V_2 V_2^*$) are

$$x_1 = \begin{pmatrix} -0.1745 \\ -0.5951 \\ 0.7329 \\ -0.2798 \end{pmatrix}, \quad \mathcal{P}_2 x_1 = \begin{pmatrix} -0.2157 + 0.04192i \\ -0.5967 + 0.003796i \\ 0.7205 + 0.009379i \\ -0.2692 - 0.009643i \end{pmatrix}.$$

It yields $\|x_1 - \mathcal{P}_2 x_1\|^2 = 3.9202\ 10^{-3}$. The two sums are

$$S_{\mathcal{I}_1} = 1.6624\ 10^{-2}, \quad S_{\mathcal{I}_2} = 6.5426\ 10^{-5}$$

and we can check that

$$\frac{S_{\mathcal{I}_2}}{S_{\mathcal{I}_1} + S_{\mathcal{I}_2}} = 3.9202\ 10^{-3}.$$

The two Ritz values at Arnoldi iteration 2 are

$$\left(1.99984 + i, \quad 1.00392 + 0.00392028i\right).$$

One is close to the eigenvalue 1 and the other is close to the cluster. We can also check the bounds of Theorem 1. Considering the bound (5) for the eigenvalue $\lambda_1 = 1$, we have

$$\|(A_2 - I)P_2 x_1\| = 5.5566\ 10^{-3}, \gamma_2 = 8.8749\ 10^{-2}, \|x_1 - \mathcal{P}_2 x_1\| = 6.2611\ 10^{-2},$$

$$\gamma_2 \|x_1 - \mathcal{P}_2 x_1\| = 5.5567\ 10^{-3}.$$

The bound is quite sharp. For the other bound we have,

$$\|(A_2 - I)x_1\| = 6.2857207\ 10^{-2}, \quad \sqrt{1 + \gamma_2^2} = 1.0039,$$

$$\sqrt{1 + \gamma_2^2}\, \|x_1 - \mathcal{P}_2 x_1\| = 6.2857211\ 10^{-2}.$$

This is also a very good bound.

We now consider a real normal matrix,

$$A = \begin{pmatrix} 1.894 & 0.09975 & -0.2124 & -0.3811 \\ 0.4172 & 1.137 & 0.3024 & 0.8461 \\ 0.1531 & -0.7204 & 1.649 & -0.1911 \\ 0.05344 & -0.6726 & -0.6651 & 1.32 \end{pmatrix},$$

and the starting vector,

$$v = \begin{pmatrix} -0.3019 \\ -0.2286 \\ -0.8316 \\ 0.4063 \end{pmatrix}.$$

The eigenvalues of A are

$$\left(1+i, \quad 1-i, \quad 2-0.01i, \quad 2+0.01i\right).$$

We have a cluster of two eigenvalues near 2 and a pair of complex conjugate eigenvalues far from this cluster. The first eigenvector of A corresponding to $\lambda_1 = 1+i$ and its projection on the Krylov subspace of dimension 2 are

$$x_1 = \begin{pmatrix} -0.1967 + 0.119i \\ 0.657 + 0i \\ 0.1591 + 0.3876i \\ -0.06602 + 0.5793i \end{pmatrix}, \quad \mathcal{P}_2 x_1 = \begin{pmatrix} 0.07054 + 0.01742i \\ 0.09665 + 0.1865i \\ 0.2707 + 0.3541i \\ 0.04567 + 0.5402i \end{pmatrix}.$$

The distance is $\|x_1 - \mathcal{P}_2 x_1\|^2 = 0.67682$. The two sums are

$$S_{\mathcal{I}_1} = 0.15514, \quad S_{\mathcal{I}_2} = 0.13115.$$

$S_{\mathcal{I}_1}$ is not large and $S_{\mathcal{I}_2}$ is not small. Consequently, the distance between x_1 and its projection cannot be small. The two Ritz values at iteration 2 are real, being $(1.8123, 1.1039)$. They are not close to any eigenvalue of A. Concerning the bounds, we have

$$\|(A_2 - (1+i)I)P_2 x_1\| = 0.67421, \gamma_2 = 0.99644, \|x_1 - \mathcal{P}_2 x_1\| = 0.67682,$$

$$\gamma_2 \|x_1 - \mathcal{P}_2 x_1\| = 0.67441,$$

and

$$\|(A_2 - (1+i)I)x_1\| = 1.1708, \quad \sqrt{1+\gamma_2^2} = 1.7300,$$

$$\sqrt{1+\gamma_2^2}\,\|x_1 - \mathcal{P}_2 x_1\| = 1.1709.$$

Again the two bounds are close to the values of the residual norms.

4.2. Bounds for the minimum distance

The formula of Theorem 7 is quite intricate when the matrix is non-normal. Let us now consider bounds for $\|(I - \mathcal{P}_k)\, x_1\|^2$ for this case. Bellalij, Jbilou and Sadok [2] proved the following result.

Lemma 8. *Let E and F be two matrices of sizes $N \times (k+1)$ and $N \times N$ respectively, $k \leq N-1$. If the matrix E is of full rank then*

$$\frac{\sigma_{\min}(F)^2}{e_1^T (E^*E)^{-1} e_1} \leq \frac{1}{e_1^T (E^*(F^*F)E)^{-1} e_1} \leq \frac{\sigma_{\max}(F)^2}{e_1^T (E^*E)^{-1} e_1}, \tag{18}$$

where $\sigma_{\min,\max}(F)$ are the smallest and largest singular values of F.

Lemma 8 can be used straightforwardly (using $E = X$ and $F = D_\alpha W_{k+1}$) together with Corollary 3 to obtain bounds as follows.

Theorem 9. *Let A be a diagonalizable matrix satisfying hypothesis H. The distance of the eigenvector x_1 to its orthogonal projection on the Krylov subspace $\mathcal{K}_k(A, v)$ is bounded by*

$$\sigma_{\min}(X)^2 \frac{N_b}{D_b} \leq \|(I - \mathcal{P}_k)\, x_1\|^2 \leq \|X\|^2 \frac{N_b}{D_b}, \tag{19}$$

where

$$N_b = \sum_{\hat{I}_{k+1}} \left[\prod_{j=2}^{k+1} |\alpha_{i_j}|^2 \right] \prod_{\substack{1<i_2\leq i_p<i_q\leq i_{k+1} \\ i_p,i_q\in\hat{I}_{k+1}}} |\lambda_{i_q} - \lambda_{i_p}|^2,$$

$$D_b = \sum_{I_k} \left[\prod_{j=1}^{k} |\alpha_{i_j}|^2 \right] \prod_{\substack{i_1\leq i_p<i_q\leq i_k \\ i_p,i_q\in I_k}} |\lambda_{i_q} - \lambda_{i_p}|^2,$$

where the summation sets of indices are defined in Lemmas 5 and 6 and $\alpha = X^{-1}v$.

The bounds of Theorem 9 are simpler than what we had in Theorem 7 since we have only one summation in the numerator and denominator like in the normal case. Moreover, the interest of these bounds is that when $\sigma_{\min}(X)$ is not too much different from $\sigma_{\max}(X)$, the distance $\|(I-\mathcal{P}_k)\, x_1\|$ is close to what we would have for a normal matrix. However, the norm of X must be larger than one and therefore the upper bound may be larger than one.

5. Study of the worst case

Obviously, we have

$$\|(I - \mathcal{P}_k)\, x_1\| \leq 1.$$

We are interested in knowing if, for a given matrix A, we can find starting vectors v such that $\|(I - \mathcal{P}_k)\, x_1\| = 1$, where $\mathcal{P}_k$ is the orthogonal projector onto the Krylov subspace $\mathcal{K}_k(A, v)$.

We first remark that $\|(I - \mathcal{P}_k)\, x_1\|^2$ is a decreasing function of k. This can be seen by using the orthonormal basis $v_1, \ldots, v_k$ of the Krylov subspace. Let $\omega_j = v_j^* x_1$. Then

$$\|(I - \mathcal{P}_k)\, x_1\|^2 = 1 - \sum_{j=1}^{k} |\omega_j|^2. \tag{20}$$

Therefore, if we find a vector $v = v_1$ such that $\|(I - \mathcal{P}_k)\, x_1\| = 1$ then, $\|(I - \mathcal{P}_j)\, x_1\| = 1$ for all $1 \leq j \leq k$ because $|\omega_j| = 0$ for $j \leq k$.

We first answer to the question above when A is normal.

5.1. A normal

Relation (17) showed that,

$$\|(I-\mathcal{P}_k)\,x_1\|^2 = \frac{S_{\mathcal{I}_k}}{S_{\mathcal{I}_1}+S_{\mathcal{I}_k}} = \frac{1}{1+\frac{S_{\mathcal{I}_1}}{S_{\mathcal{I}_k}}},$$

the last equality occurring if $S_{\mathcal{I}_k} \neq 0$. The sum $S_{\mathcal{I}_1}$ was defined in relation (16). Having $S_{\mathcal{I}_1} = 0$ is a necessary and sufficient condition to have $\|(I-\mathcal{P}_k)\,x_1\| = 1$. We have a factor $|\alpha_1|^2$ in $S_{\mathcal{I}_1}$. Therefore $S_{\mathcal{I}_1} = 0$ if $|\alpha_1| = 0$; but this case has been excluded in Corollary 3. Nevertheless, we can take $|\alpha_1|^2 = \varepsilon$ and let $\varepsilon \to 0$. Then, since $S_{\mathcal{I}_k}$ is not small,

$$\|(I-\mathcal{P}_k)\,x_1\|^2 = \frac{1}{1+\frac{\mathcal{O}(\varepsilon)}{S_{\mathcal{I}_k}}},$$

can be as close to 1 as we wish. If $\alpha_1 \neq 0$ then all the other terms in the sum $S_{\mathcal{I}_1}$ have to be zero. This means that for each term either the coefficient β_i is zero or the product of the distances between eigenvalues is zero.

To illustrate this, let us consider the first example of section 4. We now choose a starting vector v such that

$$X^*v = \begin{pmatrix} 5.774\ 10^{-5} \\ 0.5774 \\ 0.5774 \\ 0.5774 \end{pmatrix}.$$

It yields

$$S_{\mathcal{I}_1} = 6.6669\ 10^{-9}, \quad S_{\mathcal{I}_2} = 6.6667\ 10^{-5},$$

and the distance is

$$\|(I-\mathcal{P}_2)\,x_1\| = 0.99995.$$

At iteration 2 the two Ritz values are $2.0081+0.99995i$, $1.9918+0.99995i$. They are close to the cluster of eigenvalues, but not to the eigenvalue 1.

5.2. A non normal but satisfying hypothesis H

The case $k = 1$ is trivial. From relation (20) we need to have $\omega_1 = v^*x_1 = 0$. Therefore we just have to find a vector orthogonal to the eigenvector x_1. This can be done by taking any vector and orthogonalizing it against x_1.

When $k > 1$, we need to have x_1 orthogonal to the subspace generated by $v_1, \ldots, v_k$ that is, the Krylov subspace $\mathcal{K}_k(A, v)$. We can express this condition using the natural basis $v, Av, \ldots, A^{k-1}v$. It yields,

$$x_1^* A^j v = 0,\ j = 0, \ldots, k-1.$$

Using $A = X\Lambda X^{-1}$ and $v = X\alpha$, we obtain

$$x_1^* X \Lambda^j \alpha = 0,\ j = 0, \ldots, k-1. \tag{21}$$

This gives k linear equations for N unknowns which are the components of α. We have assumed in hypothesis H that A has at least k distinct eigenvalues (including λ_1) denoted as $\lambda_1, \ldots, \lambda_k$. Then, we can choose arbitrarily the components $\alpha_{k+1}, \ldots, \alpha_N$. The other ones are obtained by solving a linear system

$$\hat{\mathcal{V}}_k^T D_k^* \hat{\alpha} = -\tilde{\mathcal{V}}_{N-k}^T D_{N-k}^* \tilde{\alpha},$$

where $\hat{\alpha} = (\alpha_1, \cdots, \alpha_k)^T$, $\tilde{\alpha} = (\alpha_{k+1}, \cdots, \alpha_N)^T$, D_k (resp. D_{N-k}) is a diagonal matrix whose diagonal entries are the components 1 to k (resp. $k+1$ to N) of $y = X^* x_1$ and we have the Vandermonde matrices

$$\hat{\mathcal{V}}_k = \begin{pmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{k-1} \\ 1 & \lambda_2 & \cdots & \lambda_2^{k-1} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_k & \cdots & \lambda_k^{k-1} \end{pmatrix}, \quad \tilde{\mathcal{V}}_{N-k} = \begin{pmatrix} 1 & \lambda_{k+1} & \cdots & \lambda_{k+1}^{k-1} \\ 1 & \lambda_{k+2} & \cdots & \lambda_{k+2}^{k-1} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_N & \cdots & \lambda_N^{k-1} \end{pmatrix}.$$

If we assume that the first k components of y are different from zero, the matrix of the linear system is non singular. Then $\alpha = (\hat{\alpha}, \tilde{\alpha})^T$ that we can eventually normalize and $v = X\alpha$. This will give $\|(I - \mathcal{P}_j)\, x_1\| = 1$, $j \leq k$. Unfortunately, this technique yields starting vectors whose components are complex numbers even when the matrix A is real. Of course, the problem is that to construct such starting vectors we have to know the eigenvalues and eigenvectors of A. Therefore, this construction has just a theoretical interest. Note that when A is normal we have $y = e_1$, the Vandermonde matrix on the left-hand side is singular and we find $\alpha_1 = 0$.

Let us consider a small example with a non-normal matrix,

$$A = \begin{pmatrix} 0.5377 & 0.3188 & 3.578 & 0.7254 \\ 1.834 & -1.308 & 2.769 & -0.06305 \\ -2.259 & -0.4336 & -1.35 & 0.7147 \\ 0.8622 & 0.3426 & 3.035 & -0.205 \end{pmatrix}.$$

The eigenvalues are

$$\left(0.05302 + 2.337i, \quad 0.05302 - 2.337i, \quad -1.215 + 0.1047i, \quad -1.215 - 0.1047i\right).$$

The vector $y = X^* x_1$ is

$$y = \begin{pmatrix} 1 \\ 0.6969 - 0.07591i \\ 0.2454 - 0.2571i \\ 0.279 + 0.09219i \end{pmatrix}.$$

Let us choose arbitrarily $\alpha_3 = 1, \alpha_4 = 1$. The linear system to solve for α_1, α_2 has a matrix

$$\begin{pmatrix} 1 & 0.6969 + 0.07591i \\ 0.05302 + 2.337i & 0.2144 - 1.625i \end{pmatrix},$$

and a right-hand side

$$-\begin{pmatrix} 0.2454 + 0.2571i & 0.279 - 0.09219i \\ -0.3252 - 0.2868i & -0.3488 + 0.08284i \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

The solution is

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} -0.2167 - 0.2326i \\ -0.4259 + 0.1435i \end{pmatrix}.$$

It yields an unnormalized starting vector

$$v = \begin{pmatrix} 0.2568 - 0.05427i \\ -1.573 - 0.02843i \\ -0.16 + 0.09976i \\ 0.8892 - 0.03279i \end{pmatrix}.$$

Running the Arnoldi algorithm and computing the projection matrix we find that $\|(I - \mathcal{P}_2)\, x_1\| = 1$ as requested.

6. Conclusion

In this paper we derived closed-form expressions for the distance of an eigenvector of a diagonalizable matrix A to the Krylov subspace generated from A and a given starting vector v. This distance was given in terms of the eigenvalues and eigenvectors of A as well as the components of v in the eigenvector basis. Unfortunately, the formulas are quite intricate and deserve further study to fully understand how the distance decreases with the number of iterations in the Arnoldi method. Nevertheless, our results provide bounds for the norm of the residual of an exact eigenpair of A in the approximate problems generated by the Arnoldi method. They can also be used in a more general setting for the study of convergence of the harmonic Ritz values; see [8], [9], [10], [11]. One can extend these results to the non-diagonalizable case using the Jordan canonical form $A = SJS^{-1}$ where J is made of Jordan blocks and S is the matrix of eigenvectors and principal vectors. However, in that case determinants appear which are not known analytically (at least up to our knowledge) and not Vandermonde determinants as in the diagonalizable case. This makes the study much more difficult.

Acknowledgments

The authors thank the referees for their comments and remarks that help improving the paper.

Unfortunately our good friend, colleague and co-author Mohammed Bellalij passed away on October 19, 2015. May he rest in peace. G.M. and H.S.

References

[1] W.E. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., v 9 (1951), pp. 17–29.

[2] M. Bellalij, K. Jbilou and H. Sadok, *New convergence results on the global GMRES method for diagonalizable matrices*, J. Comp. Appl. Math., v 219 (2008), pp. 350–358.

[3] M. Bellalij, Y. Saad and H. Sadok, *Further analysis of the Arnoldi process for eigenvalue problems*, SIAM J. Numer. Anal., v 48 n 2 (2010), pp. 393–407.

[4] G. Chen and Z. Jia, *An analogue of the results of Saad and Stewart for harmonic Ritz vectors*, J. Comp. Appl. Math., v 167, n 2 (2004), pp. 493–498.

[5] J. Duintjer Tebbens, G. Meurant, H. Sadok and Z. Strakoš, *On investigating GMRES convergence using unitary matrices*, Linear Algebra Appl., v 450 (2014), pp. 83–107.

[6] J. Duintjer Tebbens and G. Meurant, *On the convergence of Q-OR and Q-MR Krylov methods for solving linear systems*, accepted in BIT Numerical Mathematics (2015).

[7] F.R. Gantmacher, *The theory of matrices*, Vol. 1, AMS Chelsea Publishing, Providence, RI, (1959).

[8] Z. Jia, *The convergence of generalized Lanczos methods for large unsymmetric eigenproblems*, SIAM J. Matrix Anal. Appl., v 16 n 3 (1995), pp. 843–862.

[9] Z. Jia, *Composite orthogonal projection methods for large matrix eigenproblems*, Science in China, v 42 n 6 (1999), pp. 577–585.

[10] Z. Jia, *The convergence of harmonic Ritz values, harmonic Ritz vectors, and refined harmonic Ritz vectors*, Math. Comp., v 74 n 251 (2005), pp. 1441–1456.

[11] Z. Jia, *Some theoretical comparisons of refined Ritz vectors and Ritz vectors*, Science in China Ser. A Mathematics, v 47 (2004), pp. 222–233.

[12] G. Meurant and J. Duintjer Tebbens, *The role eigenvalues play in forming GMRES residual norms with non-normal matrices*, Numer. Algor., v 68, n 1 (2015), pp. 143–165.

[13] Y. Saad, *Variations on Arnoldi's method for computing eigenelements of large nonsymmetric matrices*, Linear Algebra Appl., v 34 (1980), pp. 269–295.

[14] Y. Saad, *Projection methods for solving large sparse eigenvalue problems*, Matrix Pencils, Proceedings Pitea Havsbad, B.K. Kågström and A. Ruhe, eds., Lecture Notes in Math., Springer-Verlag, Berlin, v 973, (1983), pp. 121–144.

[15] Y. Saad, *Numerical methods for large eigenvalue problems*, Halstead Press, (1992). Revised edition, SIAM (2011).