Numerical Algorithms (2024) 95:1079–1102
https://doi.org/10.1007/s11075-023-01600-2

ORIGINAL PAPER

Rational QZ steps with perfect shifts

Nicola Mastronardi[1] · Marc Van Barel[2] · Raf Vandebril[2] · Paul Van Dooren[3]

Received: 19 March 2023 / Accepted: 13 June 2023 / Published online: 9 July 2023

Abstract

In this paper we analyze the stability of the problem of performing a rational QZ step with a shift that is an eigenvalue of a given regular pencil $H - \lambda K$ in unreduced Hessenberg–Hessenberg form. In exact arithmetic, the backward rational QZ step moves the eigenvalue to the top of the pencil, while the rest of the pencil is maintained in Hessenberg–Hessenberg form, which then yields a deflation of the given shift. But in finite-precision the rational QZ step gets "blurred" and precludes the deflation of the given shift at the top of the pencil. In this paper we show that when we first compute the corresponding eigenvector to sufficient accuracy, then the rational QZ step can be constructed using this eigenvector, so that the exact deflation is also obtained in finite-precision.

Keywords RQZ algorithm · Generalized eigenvalues · Perfect shift

1 Introduction

Computing all eigenvalues of a small to medium-sized matrix pencil $H - \lambda K$ is nowadays a routine task that shows up in many applications. The method of choice is

Marc Van Barel, Raf Vandebril, and Paul Van Dooren contributed equally to this work.

✉ Nicola Mastronardi
n.mastronardi@ba.iac.cnr.it

Marc Van Barel
marc.vanbarel@kuleuven.be

Raf Vandebril
raf.vandebril@kuleuven.be

Paul Van Dooren
vandooren.p@gmail.com

[1] Istituto Applicazioni Calcolo, CNR, Via Amendola 122, Bari 70126, Italy

[2] Department of Computer Science, KULeuven, Celestijnenlaan 200A, Heverlee 3001, Belgium

[3] Department of Mathematical Engineering, UCLouvain, Av. Lemaitre 4, Louvain-la-Neuve 1348, Belgium

the QZ algorithm which uses implicit QZ-type steps, implementing a bulge chasing technique. On the other hand, projection methods are often used to compute a subset of the eigenvalues of sparse, large-scale eigenproblems, and Krylov subspace methods are probably among the most used methods within this class. Even though the algorithms are totally different and they target different problems, Krylov and QZ-methods are intimately linked; theoretical support for the convergence and interpreting the QZ can be done entirely relying on Krylov theory. The rational QZ algorithm (which we will abbreviate as RQZ) is a numerical scheme that extends the ideas of the QZ algorithm and links to rational Krylov methods. It has been shown to be quite competitive with the QZ algorithm [3] because of the enhanced convergence behavior. It uses so-called RQZ steps which are pole swapping techniques on a Hessenberg-Hessenberg pencil, and not only look like bulge chasing, but also incorporate rational Krylov subspace ideas [2, 3].

The perfect shift strategy for Hessenberg pencils arises naturally in the downdating setting of orthogonal rational functions as described by Van Buggenhout, Van Barel, and Vandebril [10]. Consider a given finite discrete inner product

$$\langle f, g \rangle_m := \sum_{i=1}^{m} |w_i|^2 \overline{g(z_i)} f(z_i), \tag{1}$$

with nodes z_i and weights w_i. One wishes to construct a set of orthogonal rational functions, with prescribed poles, for this inner product. Instead of constructing the orthogonal rational functions, it is often numerically more reliable to store the recurrences for generating these functions. These recurrences are stored in a Hessenberg pencil $H - \lambda K$, satisfying

$$VH = \Lambda V K, \quad V^H V = I, \quad Ve_1 = w/\|w\|, \tag{2}$$

where w contains the weights w_i, Λ is a diagonal matrix with nodes z_i on the diagonal, and $H - \lambda K$ is a Hessenberg pencil where the ratio of the subdiagonals equals the poles. The relations (2) express that the rows of V are the left eigenvectors of the pencil $H - \lambda K$ and that the diagonal elements of Λ are their corresponding eigenvalues. The chosen nodes, weights, and poles are of course problem specific and could possibly change when, for instance, the problem changes over time. To add or remove nodes, weights, and poles, we refer to the work of Van Buggenhout, Van Barel, and Vandebril [8–10]. For removing nodes, one downdates the problem. Say we want to remove node z_j, for $j \in \{1, \ldots, m\}$. Then we need to construct unitary transformations, Z and Q such that the transformed relations

$$(VZ)(Z^H HQ) = \Lambda(VZ)(Z^H KQ)$$

allow to deflate an eigenvalue in the upper left corner[1] of the pencil $Z^H HQ - \lambda Z^H KQ$. The remaining lower right $(n-1) \times (n-1)$ part $\tilde{H} - \lambda\tilde{K}$, satisfies the relation

$$\tilde{V}\tilde{H} = \tilde{\Lambda}\tilde{V}\tilde{K}, \quad \tilde{V}^H\tilde{V} = I, \quad \tilde{V}e_1 = \tilde{w}/\|\tilde{w}\|,$$

providing the recurrences for the inner product

$$\langle f, g\rangle_{m-1} := \sum_{i=1, i\neq j}^{m} |w_i|^2 \overline{g(z_i)} f(z_i), \tag{3}$$

where $\tilde{\Lambda}$ and $\tilde{w}$ have node z_j and weight w_j removed. The exact deflation of the removed eigenvalue corresponds to the problem of deflating a perfect shift using a backward rational QZ step.

We consider only real matrix pencils and the deflation of a real eigenvalue or of a pair of complex conjugate eigenvalues. Using complex arithmetic avoids the problems of treating complex conjugate pairs together and is thus simpler. The extension to complex pencils is therefore not treated here. We will use the following notations. Matrices and submatrices are denoted by capital letters, i.e., A, B, H. The entry (i, j) of the matrix H is denoted by the lowercase letter $h_{i,j}$. Vectors are denoted by bold letters, i.e., $\mathbf{a}$, $\mathbf{b}$, ... The identity matrix of order n is denoted by I_n and its i–th column by $\mathbf{e}_i^{(n)}$, or, if there is no ambiguity, simply by I and $\mathbf{e}_i$, respectively. Generic entries different from zero in matrices or vectors are denoted by "×." The machine precision is denoted by ϵ_M. We denote a Givens rotation between two adjacent rows or columns i and $i+1$, by

$$G_i = \begin{bmatrix} I_{i-1} & & & \\ & c & -s & \\ & s & c & \\ & & & I_{n-i-1} \end{bmatrix}, \quad \begin{bmatrix} c & -s \\ s & c \end{bmatrix}\begin{bmatrix} c & -s \\ s & c \end{bmatrix}^T = I_2.$$

The rest of the paper is organized as follows. In Sect. 2, we discuss the special form of a Hessenberg-Hessenberg pencil, which is the basis for performing a perfect shift RQZ-step. In Sect. 3, we give the main result of this paper: we derive a more robust method for implementing the RQZ step so that the perfect shift can be deflated at the top of the pencil. In Sects. 4 and 5, we look at two important aspects of our algorithm, namely how to improve the accuracy of an eigenvalue/eigenvector pair and how to scale the pencil in order to improve the residual of this approximation. In Sect. 6, we illustrate the performance of our algorithm with several numerical experiments.

[1] It is compulsory to have the deflation in the upper left corner to maintain the relation between the weight vector and the matrix VZ; details can be found in [8]

2 Preliminary Hessenberg–Hessenberg form

The rational QZ algorithm for the generalized eigenvalue problem of a regular pencil $A - \lambda B$ assumes that one first reduces the pencil to a Hessenberg–Hessenberg form. This form can be obtained using orthogonal transformations U and V such that the transformed pencil $H - \lambda K := V^T(A - \lambda B)U$ consists of two Hessenberg matrices :

$$H - \lambda K := \begin{bmatrix} h_{1,1} & \dots & \dots & h_{1,n} \\ h_{2,1} & \ddots & & \vdots \\ & \ddots & \ddots & \vdots \\ & & h_{n,n-1} & h_{n,n} \end{bmatrix} - \lambda \begin{bmatrix} k_{1,1} & \dots & \dots & k_{1,n} \\ k_{2,1} & \ddots & & \vdots \\ & \ddots & \ddots & \vdots \\ & & k_{n,n-1} & k_{n,n} \end{bmatrix}. \qquad (4)$$

Such a form can be obtained by direct construction or by running a rational Krylov algorithm [2, 4]. These will be called HH pencils, and the rational QZ algorithm will be abbreviated as RQZ. The fact that the pencil $H - \lambda K$ is unreduced is equivalent to asking that the subpencil

$$H_p - \lambda K_p := \begin{bmatrix} 0 & I_{n-1} \end{bmatrix} (H - \lambda K) \begin{bmatrix} I_{n-1} \\ 0 \end{bmatrix}$$

is regular, or that the scalar pencils $h_{i+1,i} - \lambda k_{i+1,i}$ are regular for $1 \le i < n$. The subpencil $H_p - \lambda K_p$ is called the “pole pencil" of $H - \lambda K$, as its eigenvalues are the poles of the RQZ algorithm [3]. We will analyze in the next section the construction of a backward RQZ step and compare different ways to compute such a step. We go over a number of assumptions that are used in our analysis.

- Assumption (A1): $\det(H - \lambda K) \neq 0$ for almost all λ. This is well-known to hold generically and is necessary and sufficient for the definition of the generalized eigenvalues of $H - \lambda K$. Such a pencil $H - \lambda K$ is said to be *regular*.
- Assumption (A2): $\det(H_p - \lambda K_p) \neq 0$ for almost all λ, meaning that the “pole pencil" $H_p - \lambda K_p$ is regular which also holds generically and is necessary and sufficient for the definition of the poles of the HH pencil. We call such a HH pencil *unreduced*.
- Assumption (A3): $H - \lambda K$ is *proper*, meaning that the subpencil

  $$\begin{bmatrix} h_{n,n-1} & h_{n,n} \end{bmatrix} - \lambda \begin{bmatrix} k_{n,n-1} & k_{n,n} \end{bmatrix}$$

  has no zeros. Again, this holds generically.
- Assumption (A4): The perfect shift λ_0 is not a pole of $H - \lambda K$, i.e., $\det(H_p - \lambda_0 K_p) \neq 0$. This also holds generically.

Assumptions (A1) and (A2) will be assumed throughout the paper, since this is needed for the definition of generalized eigenvalues and poles of the HH pencil.

A possible extension of the above Hessenberg-Hessenberg structure occurs when the pole pencil $H_p - \lambda K_p$ is *block* upper triangular, with diagonal sub-blocks of dimensions $k_i \times k_i$. Such a block structure will be called a block-HH form. It will be discussed later on, but only for the case that the diagonal blocks have sizes $k_i = 1$ or 2.

3 Perfect shift of an unreduced *HH* pencil

In this section, we consider the case that the pole pencil $H_p - \lambda K_p$ has all block-sizes k_i equal to 1. This is the simplest case and it allows us to compare the standard RQZ approach with the eigenvector method presented in this paper.

3.1 Deflating a real eigenvalue λ_0

We assume here that we are given a regular pencil $H - \lambda K$ that is already in HH form, and that it is unreduced. If not, the operations described below can be applied to each unreduced subpencil of a general HH pencil. We also assume that assumptions (A3) and (A4) hold.

Let λ_0 be a real eigenvalue of $H - \lambda K$, then we represent it as

$$\lambda_0 := \alpha_0/\beta_0, \quad \alpha_0^2 + \beta_0^2 = 1, \quad \beta_0 \geq 0 \quad (\alpha_0 = 1, \beta_0 = 0 \text{ if } \lambda_0 = \infty).$$

In exact arithmetic, if we then perform one backward RQZ step with shift λ_0, the pencil

$$\hat{H} - \lambda \hat{K} := Z^T (H - \lambda K) Q$$

is still in HH form with its first column proportional to $\mathbf{e}_1$ and Q and Z are both unreduced Hessenberg matrices formed by the product of $n - 1$ Givens rotations. Unfortunately, the shift $(\alpha_0 - \lambda\beta_0)$ may finally not appear accurately in the $(1, 1)$ position because of a phenomenon known as “blurring of the shift." Therefore, we need to consider an alternative construction of the RQZ step, which we describe in the following theorem. Since we want to relate the rotations used in this theorem with those of the RQZ algorithm, we will make them unique by choosing the sign of s always positive when $s \neq 0$, and to choose $c = 1$ when $s = 0$.

Theorem 1 *Let $H - \lambda K$ be a real proper HH pencil with real eigenvalue $\lambda_0 = \alpha_0/\beta_0$ of absolute value $\mid \lambda_0 \mid$ bounded by 1 and normalized using $\alpha_0^2 + \beta_0^2 = 1$. Let λ_0 not be a pole of $H - \lambda K$ and define the Hessenberg matrix $M := (\beta_0 H - \alpha_0 K)$. Then*

1. *the pencil $H - \lambda K$ has a normalized real eigenvector $\mathbf{x}$ corresponding to $\lambda_0 = \alpha_0/\beta_0$:*

$$(\beta_0 H - \alpha_0 K)\mathbf{x} = M\mathbf{x} = 0, \quad \|\, \mathbf{x} \,\|_2 = 1,$$

which is unique up to a scale factor ± 1, and has a nonzero last component x_n; therefore, there is an “essentially unique” orthogonal transformation $Q := G_1^{(r)} \dots G_{n-1}^{(r)}$ that transforms $\mathbf{x}$ to $Q\mathbf{x} = \pm \mathbf{e}_1$;

2. *there is a corresponding "essentially unique" sequence of rotations* $G_{n-1}^{(\ell)}, \ldots, G_1^{(\ell)}$ *guaranteeing that the products*

$$Q := G_1^{(r)} G_2^{(r)} \cdots G_{n-1}^{(r)}, \quad Z := G_1^{(\ell)} G_2^{(\ell)} \cdots G_{n-1}^{(\ell)}, \tag{5}$$

are both Hessenberg and transform the triple $(H, K, \mathbf{x})$ *to an equivalent one*

$$(\hat{H}, \hat{K}, \hat{\mathbf{x}}) := (ZHQ^T, ZKQ^T, Q\mathbf{x})$$

where

$$\hat{\mathbf{x}} = \pm \mathbf{e}_1, \quad (\beta_0 \hat{H} - \alpha_0 \hat{K})\mathbf{e}_1 = 0,$$

and $\hat{H} - \lambda \hat{K}$ *is in* HH *form.*

Proof To prove item 1, we point out that the normalized eigenvector $\mathbf{x}$ is unique (up to a scaling factor ± 1) because it is the solution of $M\mathbf{x} = 0$, where M has rank $n - 1$ since it is unreduced and Hessenberg, because the pencil $H - \lambda K$ satisfies assumption (A4). For the same reason, its last component x_n is nonzero, since otherwise the whole vector $\mathbf{x}$ would be zero. The reduction of $\mathbf{x}$ to $\hat{\mathbf{x}} = Q\mathbf{x} = \pm \mathbf{e}_1$ then requires a sequence of Givens rotations

$$G_{i-1}^{(r)} \in \mathbb{R}^{n \times n}, \quad i = n, n-1, \ldots, 2,$$

in order to eliminate the entries $x_i, \; i = n, n-1, \ldots, 2$ of the vector $\mathbf{x}$. By choosing the sign of s in these Givens rotations positive, we make them unique.

For item 2, we point out that after the first transformation $G_{n-1}^{(r)}$ we have an updated pole pencil

$$\tilde{H}_p - \lambda \tilde{K}_p = \begin{bmatrix} 0 & I_{n-1} \end{bmatrix} (H - \lambda K) G_{n-1}^{(r)T} \begin{bmatrix} I_{n-1} \\ 0 \end{bmatrix} \tag{6}$$

that is still in generalized Schur form, but its last column has been changed and has the shift λ_0 as new pole in the bottom position. This follows from (6) which implies

$$\begin{bmatrix} \tilde{h}_{n,n-1} - \lambda_0 \tilde{k}_{n,n-1} & \tilde{h}_{n,n} - \lambda_0 \tilde{k}_{n,n} \end{bmatrix} \begin{bmatrix} x_{n-1} \\ 0 \end{bmatrix} = 0, \quad \text{where} \quad x_{n-1} \neq 0.$$

It can also be viewed as a special case of Lemma 3 with $k = 1$ and $n = 1$. Each subsequent pair of rotations $G_{i+1}^{(\ell)}$ and $G_i^{(r)T}$ moves then the perfect shift λ_0 one position up in the pole pencil $\tilde{H}_p - \lambda \tilde{K}_p$. First $G_i^{(r)}$ moves the trailing nonzero element of $\mathbf{x}$ one position up. Then $G_i^{(r)T}$ is applied to the columns of the pencil, creating a bulge in the Hessenberg matrices H and K, which is then annihilated by the left Givens transformation $G_{i+1}^{(\ell)}$. The fact that the Hessenberg form is restored in both H and K follows from Lemma 4 with $k = 1$ and $n = 1$. Therefore the pole pencil

$$\begin{bmatrix} 0 & I_{n-1} \end{bmatrix} G_2^{(\ell)} \cdots G_{n-1}^{(\ell)} (H_p - \lambda K_p) G_{n-1}^{(r)T} \cdots G_1^{(r)T} \begin{bmatrix} I_{n-1} \\ 0 \end{bmatrix}$$

has the perfect shift in its top diagonal, and then the final left rotation $G_1^{(\ell)}$ moves it to the top diagonal position of the pencil $\hat{H} - \lambda\hat{K}$ (see Lemma 5 with $n = 1$). Therefore, all the poles moved one position down, and the bottom one disappeared. All these transformations are "essentially" unique, since they implement the swapping of the eigenvalue λ_0 with one of the eigenvalues of $\hat{H}_p - \lambda\hat{K}_p$. □

The reduction described in Theorem 1, transforming an eigenvector $\mathbf{x}$ corresponding to a real eigenvalue λ_0 into a multiple of $\mathbf{e}_1$, and modifying the matrices H and K, is graphically depicted in Fig. 1, for $n = 6$. We display the evolution of the triple $(H, K, \mathbf{x})$.

In particular, a generic nonzero entry is denoted by "×," an entry to be annihilated by "⊗" and the entries becoming zero, as a consequence of the multiplication by a Givens matrix, by "⊠."

Remark 1 The implicit Q theorem for regular HH pencils is closely related to Theorem 1. It implies that the transformations Q and Z can also be determined from the first rotation $G_{n-1}^{(r)}$ that computes

$$\left[m_{n,n-1},\, m_{n,n}\right] G_{n-1}^{(r)T} = \left[0\ \times\right] \tag{7}$$

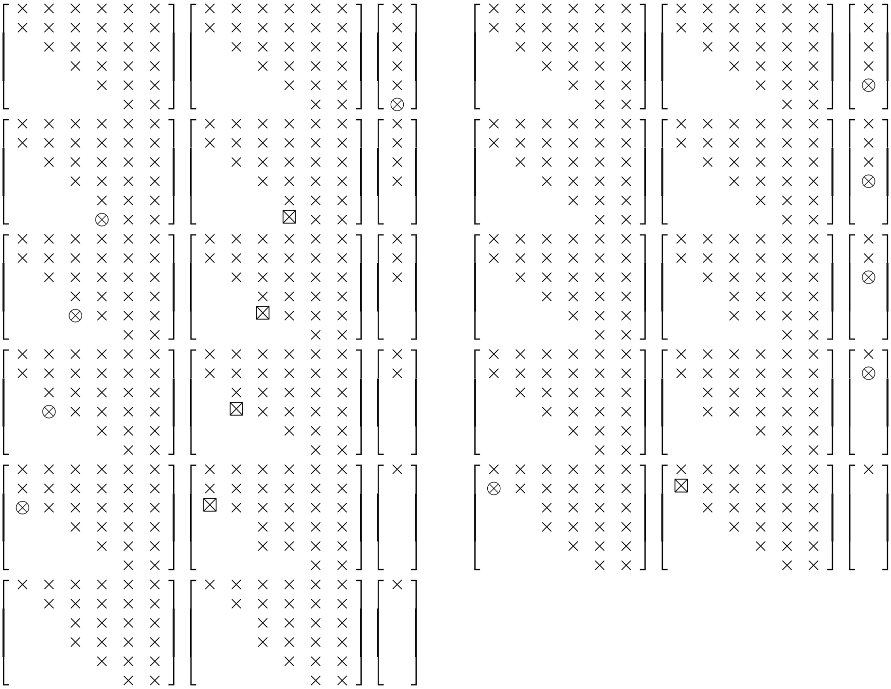

Fig. 1 Graphical description of the reduction of an eigenvector $\mathbf{x}$ to a multiple of $\mathbf{e}_1$. The matrices H and K were scaled to have norm 1, and ϵ_M-small elements were put equal to zero

and from the fact that the pair (ZHQ^T, ZKQ^T) is still in HH form. This is known as “swapping the poles" and corresponds to “chasing the bulge” [12] in the QZ algorithm.

Remark 2 Theorem 1 gives an alternative way to determine the sequences of right rotations $Q := G_1^{(r)} G_2^{(r)} \cdots G_{n-1}^{(r)}$ and left rotations $Z := G_1^{(\ell)} G_2^{(\ell)} \cdots G_{n-1}^{(\ell)}$ to implement an implicit RQZ-step. First one determines Q from $Q\mathbf{x} = \pm\mathbf{e}_1$, and then one determines Z from the restoration of the Hessenberg form of K if $\mid \lambda_0 \mid \leq 1$ and of H if $\mid \lambda_0 \mid > 1$, as indicated in Lemma 4. These particular choices are made to ensure numerical stability, as will be shown later on.

Although these different approaches are equivalent under exact arithmetic, their numerical behavior is different. We refer for this to Example 2.1 of [6], where a 3 × 3 Hessenberg matrix H of a standard eigenvalue problem is given which can be seen as a special case of a Hessenberg–Hessenberg pencil $H - \lambda K$ with $K = I$ and all its poles at infinity. The RQZ algorithm then reduces to the standard QR and will yield the same results. It was shown in [6] that the eigenvector approach is then the more reliable method for implementing the perfect shift.

3.2 Importance of the assumptions

In this subsection, we give two examples to illustrate the differences between the RQZ and the eigenvector method. The first example shows that when assumption (A4) is dropped, these two methods are not equivalent anymore.

Example 1 Consider the pencil

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix} - \lambda \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

and the shift $\lambda_0 = 0$. Its eigenvalues are 0, 0, 1 and 2 and the two eigenvalues 0 belong to one Jordan block. Therefore, λ_0 has only one eigenvector $\mathbf{x} = \mathbf{e}_4$. Assumption (A3) is satisfied, but assumption (A4) not. The eigenvector method will then use three adjacent permutations to transform $\mathbf{e}_4$ to $\mathbf{e}_1$ yielding the transformed HH pencil (where $c = \sqrt{2}/2$)

$$\left[\begin{array}{c|ccc} 0 & -c & -c & -2c \\ \hline 0 & c & c & -2c \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] - \lambda \left[\begin{array}{c|ccc} 2c & 0 & 0 & -c \\ \hline 0 & 0 & 0 & -c \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right].$$

The RQZ method, on the other hand, will obtain after the first column rotation $G_3^{(r)}$ the pencil

$$\left[\begin{array}{c|cc|c} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{array}\right] - \lambda \left[\begin{array}{c|cc|c} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right]$$

and has to perform swapping on the 2×2 subpencil $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, which is an ill-posed problem and has non-unique solutions. Therefore the RQZ method does not have a unique way to proceed further when Assumption (A4) does not hold.

Remark 3 When the pencil $H - \lambda K$ has one or more poles coalescent with the shift λ_0, the matrix $M := \beta_0 H - \alpha_0 K$ is no longer unreduced, and the proof that x_n is nonzero does not hold anymore. But it is easy to verify that if the eigenvector $\mathbf{x}$ is unique, then one (and only one) of the unreduced Hessenberg blocks of M, is singular, and that $x_n \neq 0$ if and only if this is the last block. In the above example, this was indeed the case. But even when $x_n = 0$, the eigenvector method would still work, when starting with the unreduced Hessenberg block that is singular, since the eigenvector corresponding to that subblock will have a trailing nonzero component.

In the next example, the assumption (A3) is dropped and again these two methods are not equivalent anymore.

Example 2 Consider the pencil

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix} - \lambda \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Its eigenvalues are still 0, 0, 1 and 2 and the two eigenvalues 0 belong to one Jordan block. Therefore, the perfect shift $\lambda_0 = 0$ has a single eigenvector $\mathbf{x} = \mathbf{e}_4$. The eigenvector method will then use three adjacent permutations to transform $\mathbf{e}_4$ to $\mathbf{e}_1$ yielding the transformed HH pencil

$$\left[\begin{array}{c|ccc} 0 & -1 & -1 & 0 \\ \hline 0 & 0 & 0 & -2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] - \lambda \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right].$$

The RQZ method, on the other hand, will obtain after the first column rotation $G_3^{(r)}$ the pencil

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 2 \end{array}\right] - \lambda \left[\begin{array}{ccc|c} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \end{array}\right].$$

Since Assumption (A3) does not hold, the RQZ method will not be able to introduce the shift properly in order to move it to the top. Therefore, it will rather attempt to do an early deflation of the bottom eigenvalue 2.

This example shows that the eigenvector method still works when assumption (A3) fails, provided $x_n \neq 0$.

3.3 Deflating a complex conjugate pair $(\lambda_0, \overline{\lambda}_0)$

We assume now that we are given two complex conjugate eigenvalues $(\lambda_0, \overline{\lambda}_0)$ of a regular and unreduced pencil $H - \lambda K$ that is in HH form and therefore has only real poles. This implies that assumption (A4) holds. Let us represent the eigenvalues and eigenvectors by their real and imaginary parts : $\alpha_0 \pm \imath\beta_0$ and $\mathbf{x} \pm \imath\mathbf{y}$. Then the eigenvector/eigenvalue equations $(H - (\alpha_0 \pm \imath\beta_0)K)(\mathbf{x} \pm \imath\mathbf{y}) = 0$ can be expressed as

$$HX = KX\Lambda, \quad \text{where} \quad \Lambda := \begin{bmatrix} \alpha_0 & \beta_0 \\ -\beta_0 & \alpha_0 \end{bmatrix}, \quad X := \begin{bmatrix} \mathbf{x} \, \mathbf{y} \end{bmatrix} \tag{8}$$

indicating that X spans a two-dimensional deflating subspace of the pencil $H - \lambda K$. When multiplying X with an invertible matrix R, the new basis XR can be made orthonormal and the matrix Λ then becomes $R^{-1}\Lambda R$, which preserves its eigenvalues, as expected.

The following theorem extends essentially the ideas of Theorem 1 to the case of a complex conjugate pair of shifts. Therefore, we restricted the proof to the issues that are different in the two proofs.

Theorem 2 *Let $H - \lambda K$ be a real, regular, proper and unreduced HH pencil with two complex conjugate eigenvalues $(\alpha_0 \pm \imath\beta_0)$ of absolute value $\mid \lambda_0 \mid$ bounded by 1. Then the following holds:*

1. *There exists an "essentially unique" basis X of the two-dimensional deflating subspace of the eigenvalue pair $\alpha_0 \pm \imath\beta_0$ such that*

$$X^T X = I_2, \quad X = \begin{bmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_{n-1} & y_{n-1} \\ 0 & y_n \end{bmatrix}, \quad \text{where} \quad x_{n-1} \neq 0, \;\; y_n \neq 0.$$

Moreover there exists a matrix $Q := G^{(r,2)}G^{(r,1)}$, consisting of 2 essentially unique sequences of Givens rotations

$$G^{(r,i)} = G_i^{(r,i)} \dots G_{n+i-3}^{(r,i)}, \quad i = 1, 2$$

such that their product Q gives the QR factorization $X = Q^T R$;

2. *There is a matrix $Z := G^{(\ell,2)}G^{(\ell,1)}$, consisting of 2 essentially unique sequences of Givens rotations*

$$G^{(\ell,i)} = G_i^{(\ell,i)} \dots G_{n+i-3}^{(\ell,i)}, \quad i = 1, 2,$$

and a matrix $Q := G^{(r,2)}G^{(r,1)}$*, consisting of 2 essentially unique sequences of Givens rotations*

$$G^{(r,i)} = G_i^{(r,i)} \dots G_{n+i-3}^{(r,i)}, \quad i = 1, 2,$$

such that the triple (H, K, X)*, is transformed into an equivalent one*

$$(\hat{H}, \hat{K}, \hat{X}) := (ZHQ^T, ZKQ^T, QX),$$

where $\hat{X}$ *is upper triangular, and* $(\hat{H} - \lambda\hat{K})$ *is in HH form with a leading* 2×2 *block* $[I_2, 0](\hat{H} - \lambda\hat{K})[I_2, 0]^T$ *that is decoupled and contains the eigenvalues* $\alpha_0 \pm \imath\beta_0$*.*

Proof Clearly the complex eigenvector $\mathbf{x} + \imath\mathbf{y}$ has a non-zero last component because $H - \lambda K$ is unreduced Hessenberg; and hence, the last row of X is nonzero. After the normalization, this is still the case; and hence, there exists a rotation on the columns of X that annihilates x_n. The fact that x_{n-1} is then non-zero follows from the properness assumption (A3): if $x_{n-1} = 0$, then there exists a rotation such that $G_{n-1}\begin{bmatrix} 0 & y_{n-1} \\ 0 & y_n \end{bmatrix} = \begin{bmatrix} 0 & \hat{y}_{n-1} \\ 0 & 0 \end{bmatrix}$ implying

$$\begin{bmatrix} h_{n,n-1} & h_{n,n} \end{bmatrix} G_{n-1}^T \begin{bmatrix} 0 & \hat{y}_{n-1} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} k_{n,n-1} & k_{n,n} \end{bmatrix} G_{n-1}^T \begin{bmatrix} 0 & \hat{y}_{n-1} \\ 0 & 0 \end{bmatrix} \Lambda.$$

So both $\begin{bmatrix} h_{n,n-1} & h_{n,n} \end{bmatrix}$ and $\begin{bmatrix} k_{n,n-1} & k_{n,n} \end{bmatrix}$ are parallel to the last row of G_{n-1} and this violates assumption (A3). The only degree of freedom left over is a scaling of the columns of X with ±1. Once the properties of X are established, the existence of the sequences of Givens rotations $G^{(r,i)} = G_i^{(r,i)}, \dots, G_{n-i-1}^{(r,i)}$, for $i = 1, 2$, follow: these are the rotations needed for the classical QR factorization of X. This then completes the proof of Item 1.

The proof of Item 2 is very similar to that of Item 2 in Theorem 1, except that $n = 2$ when using Lemma 3, 4 and 5, and that we need two rotations $G_{i+1}^{(r,2)}G_i^{(r,1)}$ to annihilate the two bottom positions of the matrix $\hat{X}$, and then two rotations $G_{i+1}^{(\ell,2)}G_i^{(\ell,1)}$ to restore the Hessenberg form of K if $\mid \lambda_0 \mid \leq 1$ and of H, otherwise. □

The reduction described in Theorem 2, transforming an orthogonal basis of the real and the imaginary part of an eigenvector $\mathbf{x}$ corresponding to a complex conjugate eigenvalue λ_0 into a multiple of $[\mathbf{e}_1, \mathbf{e}_2]$ and modifying the matrices H and K, is graphically depicted in Fig. 2, for $n = 6$. We display the evolution of the triple $(H, K, \mathbf{x})$.

3.4 Perfect shift of a block HH pencil

Let us now consider the case of complex conjugate poles in the pencil $H - \lambda K$. We then have a real block-HH pencil where the diagonal blocks of the pole pencil $H_p - \lambda K_p$ are 1×1 or 2×2. Again, we describe the method for a shift λ_0 of modulus

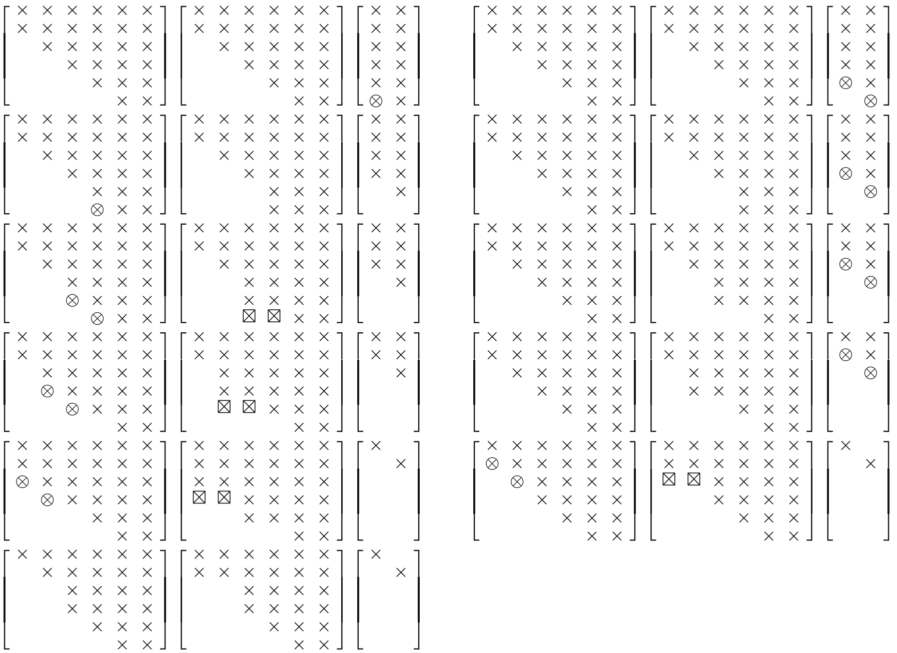

Fig. 2 Graphical description of the reduction of the real and the imaginary parts of an eigenvector corresponding to a complex conjugate eigenvalue to a multiple of $[\mathbf{e}_1, \mathbf{e}_2]$. The matrices H and K were scaled to have norm 1, and ϵ_M-small elements were put equal to zero

smaller or equal to 1. In that case, we assume K_p to be upper-triangular (and hence K is Hessenberg), while H_p is block triangular (and hence H is block Hessenberg) :

$$H = \left[\begin{array}{cccccc|c} H_{1,1} & H_{1,2} & \dots & & \dots & H_{1,n-1} & H_{1,n} \\ \hline H_{2,1} & H_{2,2} & \dots & & \dots & H_{2,n-1} & H_{2,n} \\ & H_{3,2} & \ddots & & & H_{3,n-1} & H_{3,n} \\ & & \ddots & & \ddots & \vdots & \vdots \\ & & & H_{n-1,n-2} & & H_{n-1,n-1} & H_{n-1,n} \\ & & & & & H_{n,n-1} & H_{n,n} \end{array}\right]$$

Theorems 1 and 2 are still valid, except that the HH form is now replaced by a block HH form for the pencil $H - \lambda K$. We briefly discuss the differences of the algorithm for both the case of a real shift and a complex conjugate pair. The proofs of our arguments follow from Lemmas 3, 4, and 5.

3.5 A single real shift

If the bottom block $H_{n,n-1}$ is 1×1 then a single Givens rotation $G_{n-1}^{(r)T}$ will rotate the shift λ_0 to position $(n, n-1)$. If, on the other hand, the bottom block $H_{n,n-1}$ is 2×2, two Givens transformations $G_{n-1}^{(r)T}$ and $G_{n-2}^{(r)T}$ and one Givens rotation $G_{n-1}^{(\ell)}$ have to be applied to $H - \lambda K$ to move the shift to position $(n-1, n-2)$.

After this preliminary step, λ_0 is swapped with the next block on the diagonal of the pole pencil. If this block is 1×1 a rotation $G_{i-1}^{(r)T}$ followed by a rotation $G_{i}^{(\ell)}$ moves the shift one position up. If this block is 2×2, two rotations $G_{i-1}^{(r)T}$ and $G_{i-2}^{(r)T}$ followed by 2 rotations $G_{i}^{(\ell)}$ and $G_{i-1}^{(\ell)}$ moves the shift two positions up.

The RQZ step is finalized by a single rotation $G_{1}^{(\ell)}$ moving the shift to position $(1, 1)$ in $H - \lambda K$.

3.6 Two complex conjugate shifts

Here again there are two different starting scenarios. If the bottom block $H_{n,n-1}$ is 2×2 or if the two bottom blocks are both 1×1, then a pair of Givens rotations $G_{n-2}^{(r,1)T} G_{n-1}^{(r,2)T}$ will move the shifts $(\lambda_0, \overline{\lambda}_0)$ to a new 2×2 block $H_{n,n-1}$. If this is not the case, two pairs of Givens rotations $G_{n-2}^{(r,1)T} G_{n-1}^{(r,2)T}$ and $G_{n-3}^{(r,1)T} G_{n-2}^{(r,2)T}$ and one pair of Givens rotations $G_{n-2}^{(\ell,2)} G_{n-1}^{(\ell,1)}$ have to be applied to $H - \lambda K$ to move the pair $(\lambda_0, \overline{\lambda}_0)$ to position $(n-1, n-2)$.

After this preliminary step, the pair $(\lambda_0, \overline{\lambda}_0)$ is swapped with the next block on the diagonal of the pole pencil. If this block is 1×1, a pair of rotations $G_{i-2}^{(r,1)T} G_{i-1}^{(r,2)T}$ followed by a pair of rotation $G_{i-1}^{(\ell,2)} G_{i}^{(\ell,1)}$ moves the shift one position up. If this block is 2×2, two such pairs of rotations are used to move the pair $(\lambda_0, \overline{\lambda}_0)$ two positions up.

The RQZ step is finalized by a single pair of rotations $G_{1}^{(\ell,1)}\ G_{2}^{(\ell,2)}$ moving the pair $(\lambda_0, \overline{\lambda}_0)$ to the $(1, 1)$ block in $H - \lambda K$.

The graphical description of the former reduction, transforming an orthogonal basis of the real and the imaginary parts of an eigenvector $\mathbf{x}$ corresponding to a complex conjugate eigenvalue λ_0 into a multiple of $[\mathbf{e}_1, \mathbf{e}_2]$ and modifying the matrices H and K, is depicted in Fig. 3, for $n = 7$. The evolution of the triple $(H, K, \mathbf{x})$ is displayed in that figure.

4 Approximation of eigenvalue/eigenvector pair

In this section, we show how to find or improve an approximation $(\tilde{\lambda}, \tilde{\mathbf{x}})$ to an exact eigenpair $(\lambda, \mathbf{x})$ of a pencil $H - \lambda K$ that is in proper HH form. This section applies to both real and complex eigenvalues. The eigenvalue $\tilde{\lambda}$ is given as the ratio

$$\widetilde{\lambda} = \widetilde{\alpha}/\widetilde{\beta} \text{ with } |\widetilde{\alpha}|^2 + \left|\widetilde{\beta}\right|^2 = 1$$

and the eigenvector $\tilde{\mathbf{x}}$ is supposed to have norm $\| \tilde{\mathbf{x}} \|_2 = 1$. We want to improve this approximation by reducing the norm $\|\mathbf{r}\|_2$ of the residual $\mathbf{r}$ defined by

$$(\tilde{\alpha} H - \tilde{\beta} K)\tilde{\mathbf{x}} =: \mathbf{r}, \tag{9}$$

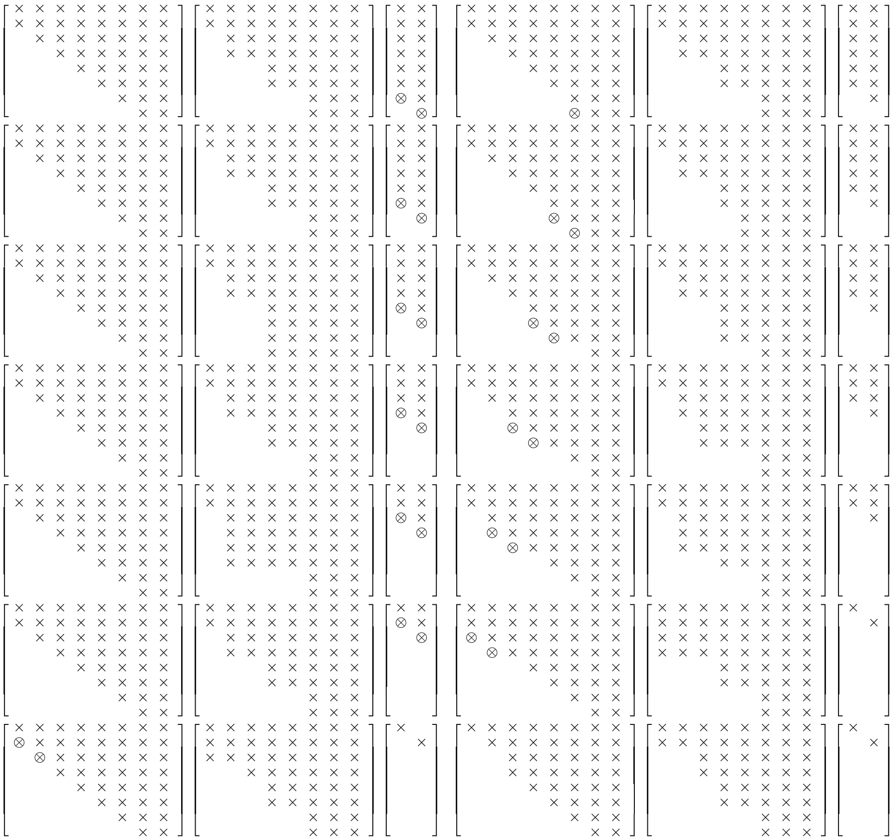

Fig. 3 Graphical description of the reduction of the real and the imaginary parts of an eigenvector corresponding to a complex conjugate eigenvalue to a multiple of $[\mathbf{e}_1, \mathbf{e}_2]$, with K in Hessenberg form and H in block Hessenberg form. The matrices H and K were scaled to have norm 1, and ϵ_M-small elements were put equal to zero

where $\mathbf{r}$ is assumed to be small, but nonzero. If the vector $\tilde{\mathbf{x}}$ is given, then the minimization of $\|\mathbf{r}\|_2$ is equivalent to

$$\min_{\left\|\begin{bmatrix}\widetilde{\alpha}\\ \widetilde{\beta}\end{bmatrix}\right\|_2=1} \left\| [H\widetilde{x} - K\widetilde{x}] \begin{bmatrix}\widetilde{\alpha}\\ \widetilde{\beta}\end{bmatrix}\right\|_2,$$

which is a total least squares problem [1] that can be solved by choosing $\begin{bmatrix}\widetilde{\alpha}\\ \widetilde{\beta}\end{bmatrix} = \mathbf{v}_2$ using the right singular vector $\mathbf{v}_2$ of the singular value decomposition of the matrix

$$\left[\, H\tilde{\mathbf{x}} \; -K\tilde{\mathbf{x}} \,\right] = \left[\, \mathbf{u}_1 \; \mathbf{u}_2 \,\right] \begin{bmatrix}\sigma_1 & 0\\ 0 & \sigma_2\end{bmatrix} \left[\, \mathbf{v}_1 \; \mathbf{v}_2 \,\right]^H .$$

If the vectors $H\tilde{\mathbf{x}}$ and $K\tilde{\mathbf{x}}$ are not parallel, this update is guaranteed to decrease the norm $\|\mathbf{r}\|_2$ (see [1]).

Let us now suppose that $\tilde{\lambda}$ is given, then the best choice for $\tilde{\mathbf{x}}$ to reduce the residual $\|\mathbf{r}\|_2$ in (9) is the n-th singular vector $\mathbf{v}_n$ of the singular value decomposition of $\tilde{M} := (\tilde{\alpha}H - \tilde{\beta}K)$, but this may be too expensive when incorporated in an iteration where $\tilde{\lambda}$ and $\tilde{\mathbf{x}}$ are updated recursively. A simpler scheme is to apply inverse iteration

$$\mathbf{z} := \tilde{M}^{-1}(\tilde{M}^{-1})^H\tilde{\mathbf{x}}, \quad \tilde{\mathbf{x}}_{new} := \mathbf{z}/\|\mathbf{z}\|_2,$$

which is again guaranteed to decrease the norm of the residual if the singular values of $\tilde{M}$ are distinct (see [5]).

The procedure explained in this section, to refine the pair $(\lambda, \mathbf{x})$ to $(\tilde{\lambda}, \tilde{\mathbf{x}})$, is primarily aimed at improving the residual of a scaled eigenvalue problem, as explained in the next section.

5 Improving the scaled residual

Let us suppose now that the pair $(\tilde{\lambda}, \tilde{\mathbf{x}})$ yields a residual (9) that is of the order of $\epsilon_M\|(H, K)\|_F$. The backward error analysis of [6, 7] then shows that we need the *stronger* bounds

$$| r_{i+1} | \le \epsilon_M\|(H, K)\|_F\|\tilde{\mathbf{x}}(i:n)\|_2, \quad i = 1, \ldots, n-1 \tag{10}$$

to ensure that the *structured backward error*[2] of the RQZ step with perfect shift is also of the order of $\epsilon_M\|(H, K)\|_F$. This can be achieved as follows. We first update the eigenvalue using the procedure explained in Sect. 4. This will already reduce the residual. For simplicity, we do not change the notation for this simple step. Then define the vector $\mathbf{d}$ as

$$d_1 = 1, \quad d_{i+1} = 2^{\text{round} \log_2 \|\tilde{\mathbf{x}}(i:n)\|_2},$$

then $d_{i+1}/\sqrt{2} \le \|\tilde{\mathbf{x}}(i:n)\|_2 \le d_{i+1}\sqrt{2}$ and the vector $\mathbf{d}$ is non-increasing (i.e. $d_{i+1} \le d_i$) since $\|\tilde{\mathbf{x}}(i:n)\|_2 \le \|\tilde{\mathbf{x}}(i-1:n)\|_2$. Also the pencil matrices

$$H_d := D^{-1}HD, \quad K_d := D^{-1}KD, \quad \text{with} \quad D := \text{diag}(d_1, \ldots, d_n)$$

satisfy the bounds

$$\|(H_d, K_d)\|_F \le \gamma\|(H, K)\|_F, \quad \text{where} \quad \gamma := \max_{1\le i\le n-1} d_i/d_{i+1} \ge 1,$$

and the equation

$$(\tilde{\alpha}H_d - \tilde{\beta}K_d)D^{-1}\tilde{\mathbf{x}} = D^{-1}\mathbf{r}.$$

[2] A *structured* backward error of an HH pencil is one that has zero elements where the pencil has zero elements.

The scaled subvectors of $\tilde{\mathbf{x}}_d := D^{-1}\tilde{\mathbf{x}}$ then have approximately the same norm (see Appendix Scaling) :

$$\frac{1}{\gamma\sqrt{2}} \leq \|\tilde{\mathbf{x}}_d(n-1:n)\|_2 \leq \ldots \leq \|\tilde{\mathbf{x}}_d(1:n)\|_2 \leq \sqrt{2n},$$

which implies that the norm of $\tilde{\mathbf{x}}_d$ is of the order of 1. After performing one step of inverse iteration on $\tilde{\mathbf{x}}_d$ to improve that computed eigenvector, we obtain a scaled residual $\mathbf{r}_{d,new} = (\tilde{\alpha} H_d - \tilde{\beta} K_d)\tilde{\mathbf{x}}_{d,new}$ satisfying (for a moderate value of c)

$$\|\mathbf{r}_{d,new}\|_2 \leq c\epsilon_M \|(H_d, K_d)\|_F \leq c\gamma\epsilon_M \|(H, K)\|_F.$$

Multiplying the above equation with D yields in the original coordinate system

$$\tilde{\mathbf{x}}_{new} := D\tilde{\mathbf{x}}_{d,new}, \quad \tilde{\mathbf{r}}_{new} := D\mathbf{r}_{d,new}, \quad (\tilde{\alpha} H - \tilde{\beta} K)\tilde{\mathbf{x}}_{new} = \tilde{\mathbf{r}}_{new}.$$

The $(i+1)$-th element of $\tilde{\mathbf{r}}_{new}$ then satisfies the required bound since

$$\mathbf{e}_{i+1}^T \tilde{\mathbf{r}}_{new} = d_{i+1}\mathbf{e}_{i+1}^T \mathbf{r}_{d,new} \leq d_{i+1}\|\mathbf{r}_{d,new}\|_2 \leq c\gamma d_{i+1}\epsilon_M \|(H, K)\|_F.$$

If the constant factor $c\gamma$ is large, the scaled refinement step may not yield the expected error bound (10) and an additional refinement step may be needed. In the numerical experiments section, we show that one step of refinement often yields a satisfactory result.

The above method can also be applied to complex eigenvectors, but its impact of the properties of a real deflating subspace X used in the case of complex conjugate pairs of eigenvalues is not clear.

The efficacity of the above approximations, and their use for complex conjugate pairs, is verified in the "Numerical results" section.

6 Numerical results

In this section, we report some numerical experiments. All the computations were performed with `Matlab ver. R2022a` with machine precision $\epsilon_M \approx 2.22 \times 10^{-16}$.

We consider 10,000 HH matrix pencils $(H^{(i)}, K^{(i)})$, of size 100, with pseudo-random values drawn from the standard normal distribution (generated by the function `randn` of `Matlab`) as entries, and scaled such that $\|H^{(i)}\|_2 = \|K^{(i)}\|_2 = 1$. For each matrix, we randomly pick a real and a complex conjugate eigenpair $(\lambda^{(i)}, \mathbf{x}^{(i)})$, and apply the perfect shift technique to deflate that particular eigenvalue from the matrix pencil obtaining the new HH matrix pencils $(\tilde{H}^{(i)}, \tilde{K}^{(i)})$. Furthermore, we also apply the improved scaled residual approach, described in Sect. 5, to compute a better approximation of the eigenpair, obtaining $(\hat{\lambda}^{(i)}, \hat{\mathbf{x}}^{(i)})$, and deflate it by means of the perfect shift technique obtaining the HH matrix pencils $(\hat{H}^{(i)}, \hat{K}^{(i)})$.

We define $b^{(i)} = 1/\sqrt{1+\lambda^{(i)2}}$, $a^{(i)} = \lambda^{(i)}\, b^{(i)}$, $\hat{b}^{(i)} = 1/\sqrt{1+\hat{\lambda}^{(i)2}}$, $\hat{a}^{(i)} = \hat{\lambda}^{(i)}\, \hat{b}^{(i)}$. Moreover, we adopt the Matlab function tril(F, k) to denote the lower triangular part of the matrix F below the kth subdiagonal.

The results are depicted in the histograms displayed in the following pictures. In each figure, the histogram to the left refers to the matrices $(\tilde{H}^{(i)}, \tilde{K}^{(i)})$, while the one to the right refers to the HH matrix pencils $(\hat{H}^{(i)}, \hat{K}^{(i)})$.

The first five figures concern the deflation of a real eigenpair, while the last three figures refer to the complex conjugate case.

In Fig. 4, the histograms of $\log_{10}\sqrt{\tilde{c}_1^{(i)} + \tilde{c}_2^{(i)}}$ (left), with $\tilde{c}_1^{(i)} = \|\texttt{tril}(\tilde{H}^{(i)}, -2)\|_2^2 + \|\texttt{tril}(\tilde{K}^{(i)}, -2)\|_2^2$, and $\tilde{c}_2^{(i)} = \mid \tilde{H}_{2,1}^{(i)} \mid^2 + \mid \tilde{K}_{2,1}^{(i)} \mid^2$, and $\log_{10}\sqrt{\hat{c}_1^{(i)} + \hat{c}_2^{(i)}}$ (right), with $\hat{c}_1^{(i)} = \|\texttt{tril}(\hat{H}^{(i)}, -2)\|_2 + \|\texttt{tril}(\hat{K}^{(i)}, -2)\|_2$ and $\hat{c}_2^{(i)} = \mid \hat{H}_{2,1}^{(i)} \mid^2 + \mid \hat{K}_{2,1}^{(i)} \mid^2$, are displayed. It can be noticed that if the improved scaled residual approach is not applied, the part below the first subdiagonal of the computed HH matrices often gets blurred.

In Fig. 5, the histograms of $\log_{10} \mid b\tilde{H}_{1,1} - a\tilde{K}_{1,1} \mid$ (left) and $\log_{10} \mid \hat{b}\hat{H}_{1,1} - \hat{a}\hat{K}_{1,1} \mid$ (right), are displayed.

In Fig. 6, the histograms of the logarithms of the residuals $\log_{10}\|(a\tilde{K} - b\tilde{H})\mathbf{x}\|_2$ (left) and $\log_{10}\|(\hat{a}\hat{K} - \hat{b}\hat{H})\hat{\mathbf{x}}\|_2$ (right), are displayed.

The histograms in Fig. 7 show the accuracy of the poles in the HH matrices after deflation. In particular, using the definition $p_j = H_{j+1,j}/K_{j+1,j}$, $\tilde{p}_j =$

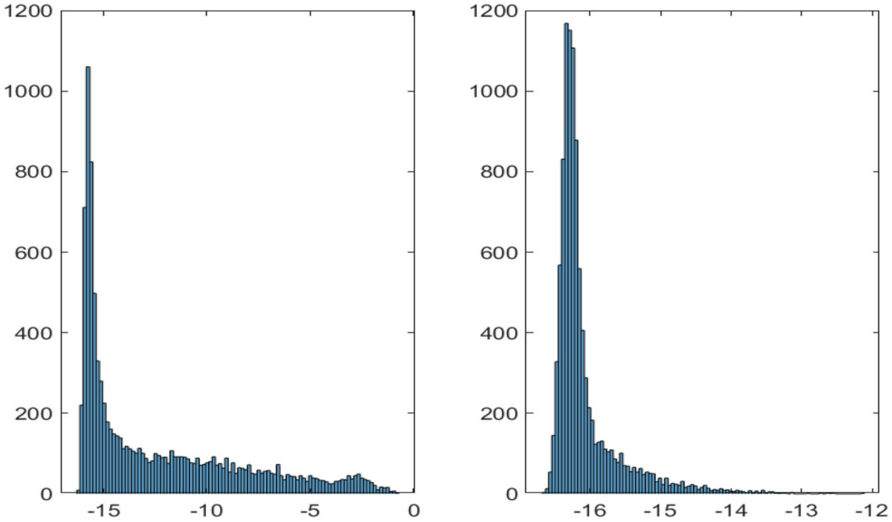

Fig. 4 Histograms of $\log_{10}\sqrt{\tilde{c}_1^{(i)} + \tilde{c}_2^{(i)}}$ (left), with $\tilde{c}_1^{(i)} = \|\texttt{tril}(\tilde{H}^{(i)}, -2)\|_2^2 + \|\texttt{tril}(\tilde{K}^{(i)}, -2)\|_2^2$, and $\tilde{c}_2^{(i)} = \mid \tilde{H}_{2,1}^{(i)} \mid^2 + \mid \tilde{K}_{2,1}^{(i)} \mid^2$, and $\log_{10}\sqrt{\hat{c}_1^{(i)} + \hat{c}_2^{(i)}}$ (right), with $\hat{c}_1^{(i)} = \|\texttt{tril}(\hat{H}^{(i)}, -2)\|_2 + \|\texttt{tril}(\hat{K}^{(i)}, -2)\|_2$ and $\hat{c}_2^{(i)} = \mid \hat{H}_{2,1}^{(i)} \mid^2 + \mid \hat{K}_{2,1}^{(i)} \mid^2$, real eigenvalue

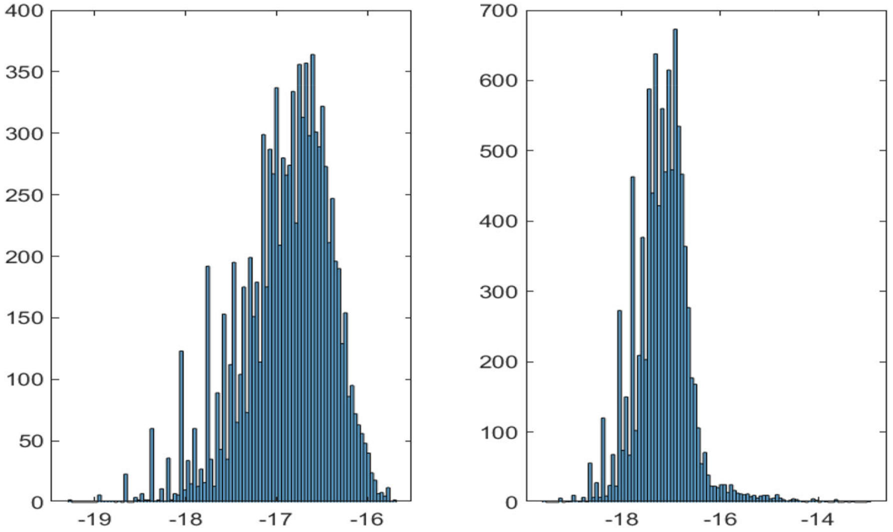

Fig. 5 Histograms of $\log_{10} \mid b\tilde{H}_{1,1} - a\tilde{K}_{1,1} \mid$ (left) and $\log_{10} \mid \hat{b}\hat{H}_{1,1} - \hat{a}\hat{K}_{1,1} \mid$ (right), real eigenvalue

$\tilde{H}_{j+2,j+1}/\tilde{K}_{j+2,j+1}$, and $\hat{p}_j = \hat{H}_{j+2,j+1}/\hat{K}_{j+2,j+1}$, $j = 1, \ldots, n-2$, the values of $\log_{10} \max_j \frac{|p_j - \tilde{p}_j|}{|p_j|}$ (left) and $\log_{10} \max_j \frac{|p_j - \hat{p}_j|}{|p_j|}$ (right), are displayed.

The next three figures report the histograms corresponding to the complex conjugate eigenvalue pair $(\lambda^{(i)}, \bar{\lambda}^{(i)})$. The histograms of $\log_{10} \sqrt{\tilde{c}_1^{(i)} + \tilde{c}_2^{(i)}}$ (left), with $\tilde{c}_1^{(i)} =$

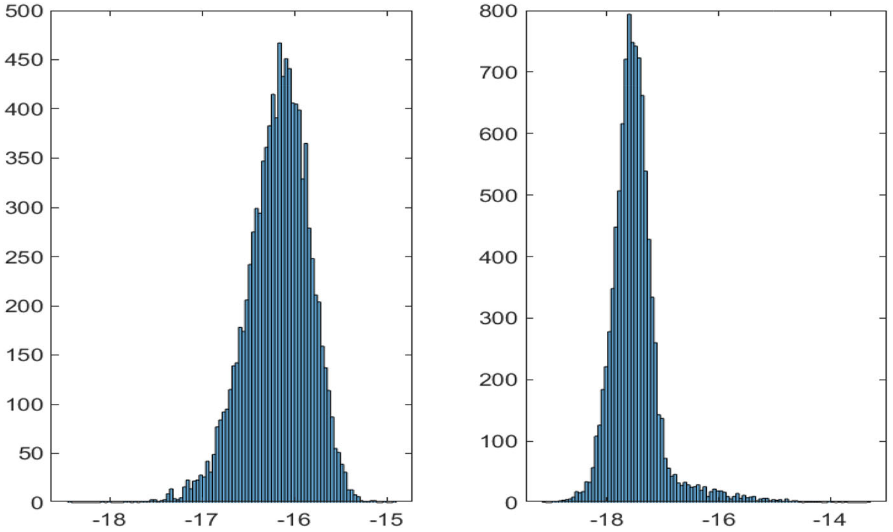

Fig. 6 Histograms of the logarithms of the residuals, $\log_{10} \|(aK - bH)\mathbf{x}\|_2$ (left) and $\log_{10} \|(\hat{a}K - \hat{b}H)\hat{\mathbf{x}}\|_2$ (right), real eigenvalue

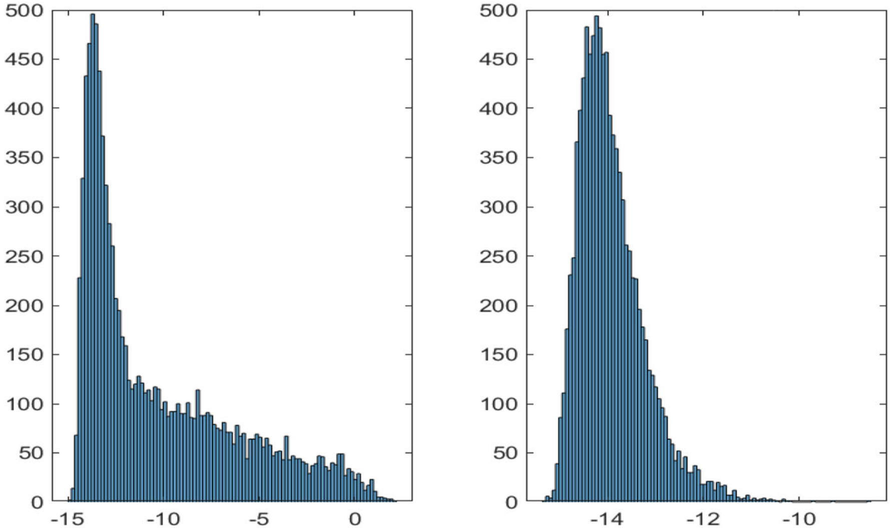

Fig. 7 Accuracy of the poles in the HH matrices after deflation, real eigenvalue

$\|\texttt{tril}(\tilde{H}^{(i)}, -2)\|_2^2 + \|\texttt{tril}(\tilde{K}^{(i)}, -2)\|_2^2$, and $\tilde{c}_2^{(i)} =\mid \tilde{H}_{2,1}^{(i)} \mid^2 + \mid \tilde{K}_{2,1}^{(i)} \mid^2$, and $\log_{10}\sqrt{\hat{c}_1^{(i)} + \hat{c}_2^{(i)}}$ (right), with $\hat{c}_1^{(i)} = \|\texttt{tril}(\hat{H}^{(i)}, -2)\|_2 + \|\texttt{tril}(\hat{K}^{(i)}, -2)\|_2$ and $\hat{c}_2^{(i)} =\mid \hat{H}_{2,1}^{(i)} \mid^2 + \mid \hat{K}_{2,1}^{(i)} \mid^2$, are displayed in Fig. 8. Similar to the real case, it can be

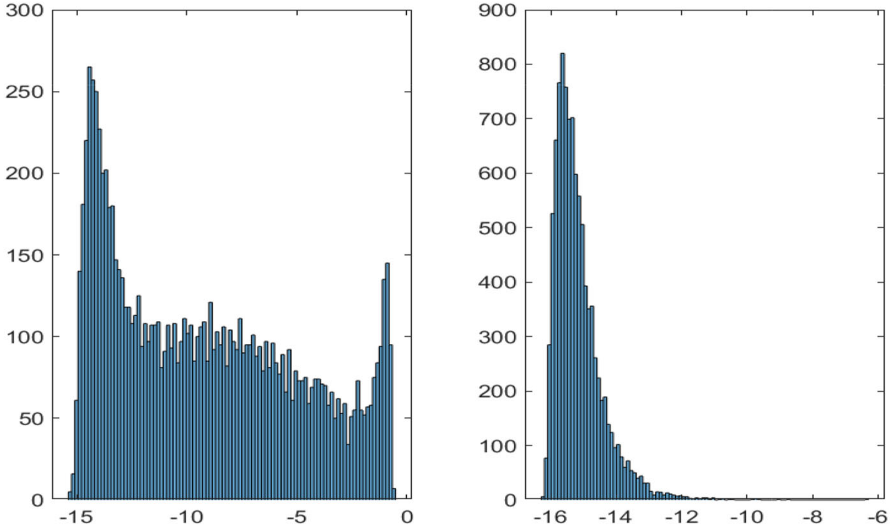

Fig. 8 Histograms of $\log_{10}\sqrt{\tilde{c}_1^{(i)} + \tilde{c}_2^{(i)}}$ (left), with $\tilde{c}_1^{(i)} = \|\texttt{tril}(\tilde{H}^{(i)}, -2)\|_2^2 + \|\texttt{tril}(\tilde{K}^{(i)}, -2)\|_2^2$, and $\tilde{c}_2^{(i)} =\mid \tilde{H}_{2,1}^{(i)} \mid^2 + \mid \tilde{K}_{2,1}^{(i)} \mid^2$, and $\log_{10}\sqrt{\hat{c}_1^{(i)} + \hat{c}_2^{(i)}}$ (right), with $\hat{c}_1^{(i)} = \|\texttt{tril}(\hat{H}^{(i)}, -2)\|_2 + \|\texttt{tril}(\hat{K}^{(i)}, -2)\|_2$ and $\hat{c}_2^{(i)} =\mid \hat{H}_{2,1}^{(i)} \mid^2 + \mid \hat{K}_{2,1}^{(i)} \mid^2$, complex conjugate eigenpair

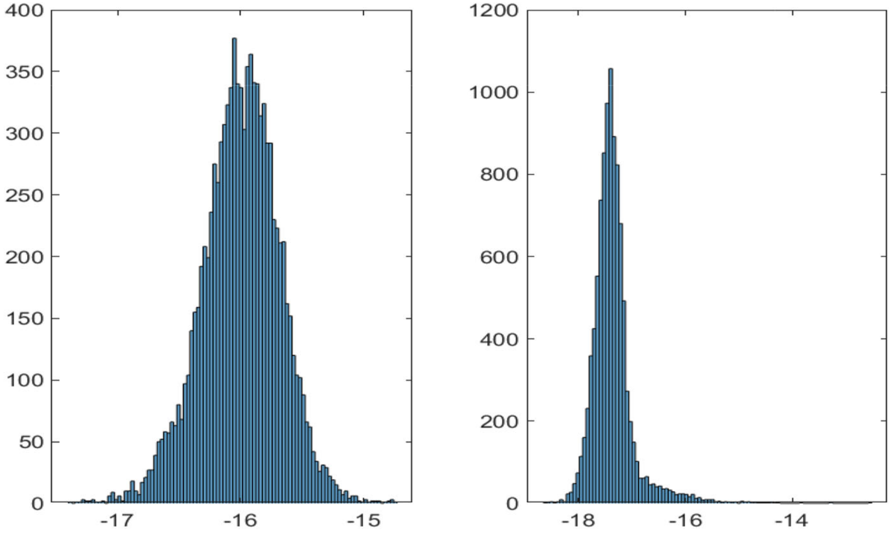

Fig. 9 Histograms of the logarithms of the residuals $\log_{10}\|(\tilde{a}K-\tilde{b}H)\mathbf{x}\|_2$ (left) and $\log_{10}\|(\hat{a}\hat{K}-\hat{b}\hat{H})\hat{\mathbf{x}}\|_2$ (right), complex conjugate eigenpair

noticed that if the improved scaled residual approach is not applied, the part below the first subdiagonal of the computed HH matrices gets blurred.

In Fig. 9, the histograms of the logarithms of the residuals $\log_{10}\|(aK-bH)\mathbf{x}\|_2$ (left) and $\log_{10}\|(\hat{a}K-\hat{b}H)\hat{\mathbf{x}}\|_2$ (right), are displayed.

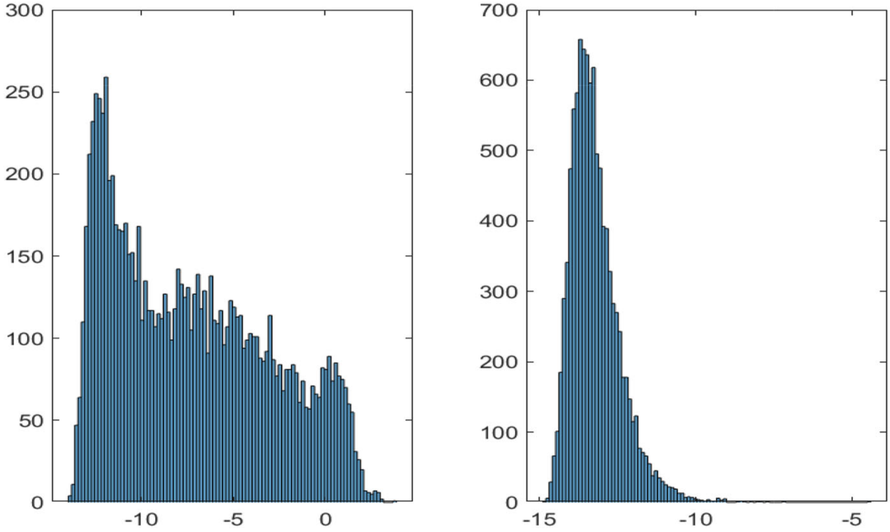

Fig. 10 Accuracy of the poles in the HH matrices after deflation, complex conjugate eigenpair

The histograms in Fig. 10 display the accuracy of the poles in the HH matrices after deflation. In particular, using the definition $p_j = H_{j+1,j}/K_{j+1,j}$, $\tilde{p}_j = \tilde{H}_{j+3,j+2}/\tilde{K}_{j+3,j+2}$, and $\hat{p}_j = \hat{H}_{j+3,j+2}/\hat{K}_{j+3,j+2}$, $j = 1, \ldots, n-3$, the values of $\log_{10} \max_j \frac{|p_j - \tilde{p}_j|}{|p_j|}$ (left) and $\log_{10} \max_j \frac{|p_j - \hat{p}_j|}{|p_j|}$ (right), are displayed.

Appendix

A. Deflations and perturbations

In this section, we assume that the pencil $H - \lambda K$ has coefficients of norm bounded by 1, that X has full column rank and has norm bounded by 1, and that Λ is a square matrix with spectral radius bounded by 1. Therefore, when applying orthonormal transformations to H, K or X, the numerical errors will be of the order of ϵ_M. We show how the perfect shift propagates in the backward RQZ step by tracking the residual of the deflating subspace equation $R := HX - KX\Lambda$. When applying the orthonormal transformations Q and Z, the residual R changes to a new residual $\hat{R} := \hat{H}\hat{X} - \hat{K}\hat{X}\Lambda$, which can be evaluated as follows. Let us superpose the errors performed during the transformations on the data X, H, and K :

$$\hat{X} := Q(X + \Delta_X), \quad \hat{K} := Z(K + \Delta_K)Q^T, \quad \hat{H} := Z(H + \Delta_H)Q^T,$$

then

$$\hat{R} = Z\left[R + \Delta_H X + H\Delta_X - (\Delta_K X + K\Delta_X)\Lambda\right] + \mathcal{O}(\epsilon_M^2)$$

which shows they are of the same order of magnitude, provided $\|\Lambda\|_2$ is of the order of 1.

Lemma 3 *Let $HX = KX\Lambda$, where the pencil $H - \lambda K$ is $k \times (k+1)$ and the matrix X is $(k+1) \times n$, where the matrix KX and hence also X, have full column rank n and the $n \times n$ matrix Λ has spectral radius bounded by 1. If Q is an orthonormal transformation satisfying $\hat{X} := QX = \begin{bmatrix} \hat{X}_1 \\ 0 \end{bmatrix}$ where $\hat{X}_1$ is $n \times n$ and invertible, and then we partition*

$$\hat{H} - \lambda\hat{K} := (H - \lambda K)Q^T = \left[\,\hat{H}_{1,1}\ \hat{H}_{1,2}\,\right] - \lambda\left[\,\hat{K}_{1,1}\ \hat{K}_{1,2}\,\right],$$

where the pencil $\hat{H}_{1,1} - \lambda\hat{K}_{1,1}$ is $n \times n$, then the resulting equation

$$\left[\,\hat{H}_{1,1}\ \hat{H}_{1,2}\,\right]\begin{bmatrix} \hat{X}_1 \\ 0 \end{bmatrix} = \left[\,\hat{K}_{1,1}\ \hat{K}_{1,2}\,\right]\begin{bmatrix} \hat{X}_1 \\ 0 \end{bmatrix}\Lambda$$

implies that the spectrum of the pencil $\hat{H}_{1,1} - \lambda\hat{K}_{1,1}$ is that of the matrix Λ.

Proof This follows immediately from $\hat{H}_{1,1}\hat{X}_1 = \hat{K}_{1,1}\hat{X}_1\Lambda$ and the fact that $\hat{K}_{1,1}\hat{X}_1$ has full rank n. □

Lemma 4 *Let $HX = KX\Lambda$ where the $(k+n)\times(k+n)$ pencil $H-\lambda K$ is regular, the $(k+n)\times n$ matrix X has full column rank n and the $n\times n$ matrix Λ has spectral radius bounded by 1. If Q is an orthonormal transformation satisfying $\hat{X} := QX = \begin{bmatrix} \hat{X}_1 \\ 0 \end{bmatrix}$ where $\hat{X}_1$ is $n\times n$ and invertible, and Z is an orthonormal matrix such that*

$$\hat{K} := ZKQ^T = \begin{bmatrix} \hat{K}_{1,1} & \hat{K}_{1,2} \\ 0 & \hat{K}_{2,2} \end{bmatrix}, \quad \hat{H} := ZHQ^T = \begin{bmatrix} \hat{H}_{1,1} & \hat{H}_{1,2} \\ \hat{H}_{2,1} & \hat{H}_{2,2} \end{bmatrix},$$

where $\hat{K}_{1,1}$ is $n\times n$, then the resulting equation

$$\begin{bmatrix} \hat{H}_{1,1} & \hat{H}_{1,2} \\ \hat{H}_{2,1} & \hat{H}_{2,2} \end{bmatrix} \begin{bmatrix} \hat{X}_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \hat{K}_{1,1} & \hat{K}_{1,2} \\ 0 & \hat{K}_{2,2} \end{bmatrix} \begin{bmatrix} \hat{X}_1 \\ 0 \end{bmatrix} \Lambda$$

implies that $\hat{H}_{2,1} = 0$ and the spectrum of the pencil $\hat{H}_{1,1} - \lambda\hat{K}_{1,1}$ is that of the matrix Λ. The correponding deflating subspace has then been transformed to the top block. If, on the other hand, there is a nonzero residual $\hat{R} := \hat{H}\hat{X} - \hat{K}\hat{X}\Lambda$ then $\|\hat{H}_{2,1}\|_2 = \mathcal{O}(\epsilon_M)$ and $\hat{H}_{2,1}$ can safely be dismissed if $\|\hat{R}\hat{X}^{-1}\|_2 = \mathcal{O}(\epsilon_M)$.

Proof The result with zero residual follows from the equation $\hat{H}_{2,1}\hat{X}_1 = 0$. The result with nonzero residual follows from $\hat{H}_{2,1}\hat{X}_1 = \begin{bmatrix} 0 & I_k \end{bmatrix} \hat{R}$, which is $\mathcal{O}(\epsilon_M)$ when $\|\hat{R}\hat{X}_1^{-1}\|_2 = \mathcal{O}(\epsilon_M)$. □

Lemma 5 *Let $HX = KX\Lambda$ where $H-\lambda K$ is a $(n+1)\times n$ pencil and the $n\times n$ matrix X has full rank n and Λ has spectral radius bounded by 1. Let Z be an orthonormal matrix such that $\hat{K} := ZK = \begin{bmatrix} \hat{K}_{1,1} \\ 0 \end{bmatrix}$, where $\hat{K}_{1,1}$ is $n\times n$, then $\hat{H} := ZH = \begin{bmatrix} \hat{H}_{1,1} \\ 0 \end{bmatrix}$ and the spectrum of $\hat{H}_{1,1} - \lambda\hat{K}_{1,1}$ is that of the matrix Λ. If, on the other hand, there is a nonzero residual $\hat{R} := \hat{H}\hat{X} - \hat{K}\hat{X}\Lambda$ then $\|\hat{H}_{2,1}\|_2 = \mathcal{O}(\epsilon_M)$ and $\hat{H}_{2,1}$ can safely be dismissed if $\|\hat{R}\hat{X}_1^{-1}\|_2 = \mathcal{O}(\epsilon_M)$.*

Proof This follows immediately from $\hat{H}_{2,1}\hat{X} = \begin{bmatrix} 0 & I_k \end{bmatrix} \hat{R}$. □

B. Scaling

Lemma 6 *Let $\tilde{x}_n \neq 0$, then the scaling $d_1 = 1,\ d_{i+1} = 2^{\mathrm{round}\log_2 \|\tilde{\mathbf{x}}(i:n)\|_2}$ and scaled vector $\tilde{\mathbf{x}}_d$ of the normalized vector $\tilde{\mathbf{x}}$, satisfies $d_1 \geq \ldots \geq d_n$ and*

$$\frac{1}{\gamma\sqrt{2}} \leq \|\tilde{\mathbf{x}}_d(n-1:n)\|_2 \leq \ldots \leq \|\tilde{\mathbf{x}}_d(1:n)\|_2 \leq \sqrt{2n}.$$

Proof Clearly, each element of $\tilde{\mathbf{x}}_d$ is upper bounded by $\sqrt{2}$ because of the rounding, and therefore $\|\tilde{\mathbf{x}}_d(i:n)\|_2 \leq \sqrt{2n}$. The smallest of the subvectors $\|\tilde{\mathbf{x}}_d(n-1:n)\|_2$ is lower bounded by

$$1/(\gamma\sqrt{2}) \leq d_n/(d_{n-1}\sqrt{2}) \leq \|\tilde{\mathbf{x}}(n-1:n)\|_2/d_{n-1} \leq \|\tilde{\mathbf{x}}_d(n-1:n)\|_2.$$

□

Acknowledgements The authors wish to thank the anonymous reviewer for his/her valuable comments.

Author Contributions The authors contributed equally to this work.

Funding Open access funding provided by Consiglio Nazionale Delle Ricerche (CNR) within the CRUI-CARE Agreement. The first and fourth authors were partly supported by Gruppo Nazionale Calcolo Scientifico (GNCS) of Istituto Nazionale di Alta Matematica (INdDAM). The second author was supported by the Research Council KU Leuven, C1-project C14/17/073 and by the Fund for Scientific Research-Flanders (Belgium), EOS Project no 30468160.The third author was supported by the Research Council KU Leuven (Belgium), project C16/21/002 and by the Fund for Scientific Research – Flanders (Belgium), project G0A9923N. The second and third author were also supported by the Fund for Scientific Research – Flanders (Belgium), project G0B0123N.

Availability of data and materials Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Code Availability The codes described in the manuscript are available from the corresponding author upon request.

Declarations

Ethics approval Not applicable.

Conflict of interest The authors declare no competing interests.

References

1. Boutry, G., Elad, M., Golub, G., Milanfar, P.: The generalized eigenvalue problem for nonsquare pencils using a minimal perturbation approach. SIAM Matr. Anal. Appl. **27**(2), 582–601 (2005)
2. Camps, D.: Pole swapping methods for the eigenvalue problem, Ph.D. Thesis Dept. Comp. Sc., KULeuven (2019)
3. Camps, D., Meerbergen, K., Vandebril, R.: A rational QZ method. SIAM Matr. Anal. Appl. **40**(3), 943–972 (2019)
4. Camps, D., Mastronardi, N., Vandebril, R., Van Dooren, P.: Swapping 2 × 2 blocks in the Schur and generalized Schur form. J. Comput. Appl. Math. **373**, 112274 (2020)

5. Ipsen, I.: Computing an eigenvector with inverse iteration. SIAM Rev. **39**(2), 254–291 (1997)
6. Mastronardi, N., Van Dooren, P.: The QR-steps with perfect shifts. SIAM J. Matr. Anal. Appl. **39**(4), 1591–1615 (2018)
7. Mastronardi, N., Van Dooren, P.: On QZ steps with perfect shifts and computing the index of a differential-algebraic equation. IMA J. Num. Anal. (2020). https://doi.org/10.1093/imanum/draa049
8. Van Barel, M., Van Buggenhout, N., Vandebril, R.: Algorithms for modifying recurrence relations of orthogonal polynomials and rational functions when changing the discrete inner product, (2023). arXiv:2302.00355
9. Van Buggenhout, N., Van Barel, M., Vandebril, R.: Non-unitary CMV-decomposition. Special Matrices **8**(1), 144–159 (2020)
10. Van Buggenhout, N., Van Barel, M., Vandebril, R.: Generation of orthogonal rational functions by procedures for structured matrices. Numerical Algorithms **89**(2), 1–32 (2021)
11. Watkins, D.S.: The transmission of shifts and shift blurring in the QR algorithm. Lin. Alg. Appl. **241–243**, 877–896 (1996)
12. Watkins, D.S.: The Matrix Eigenvalue Problem. SIAM, Philadelphia (2007)