SIAM J. MATRIX ANAL. APPL.
Vol. 40, No. 3, pp. 943–972

A RATIONAL QZ METHOD*

DAAN CAMPS†, KARL MEERBERGEN†, AND RAF VANDEBRIL†

Abstract. We propose a *rational QZ* method for the solution of the dense, unsymmetric generalized eigenvalue problem. This generalization of the classical QZ method operates implicitly on a Hessenberg, Hessenberg pencil instead of on a Hessenberg, triangular pencil. Whereas the QZ method performs nested subspace iteration driven by a polynomial, the rational QZ method allows for nested subspace iteration driven by a rational function; this creates the additional freedom of selecting poles. In this article we study Hessenberg, Hessenberg pencils, link them to rational Krylov subspaces, propose a direct reduction method to such a pencil, and introduce the implicit rational QZ step. The link with rational Krylov subspaces allows us to prove essential uniqueness (implicit Q theorem) of the rational QZ iterates as well as convergence of the proposed method. In the proofs, we operate directly on the pencil instead of rephrasing it all in terms of a single matrix. Numerical experiments are included to illustrate competitiveness in terms of speed and accuracy with the classical approach. Two other types of experiments exemplify new possibilities. First we illustrate that good pole selection can be used to deflate the original problem during the reduction phase, and second we use the rational QZ method to implicitly filter a rational Krylov subspace in an iterative method.

Key words. generalized eigenvalues, implicit, rational QZ, rational Krylov

AMS subject classifications. 65F15, 15A18

DOI. 10.1137/18M1170480

1. Introduction.

The numerical computation of the eigenvalues of a regular[1] matrix pair $A, B \in \mathbb{C}^{n\times n}$ is the principal problem studied in this paper. The set of eigenvalues of (A, B) is denoted as Λ and defined by

$$\Lambda = \{\lambda = \alpha/\beta \in \bar{\mathbb{C}} : \det(\beta A - \alpha B) = 0\} \tag{1}$$

with $\bar{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. If $\beta \neq 0$, the eigenvalue is equal to $\lambda = \alpha/\beta$, while for $\beta = 0$ the eigenvalue is located at ∞. When there are no infinite eigenvalues B is invertible and the eigenvalues of the pencil are equal to those of $B^{-1}A$ or AB^{-1} (see, e.g., the monographs [10, 33]).

The QZ method, originally introduced by Moler and Stewart [16], is presumably the method of choice for the solution of this problem for general small to medium-size matrix pairs. The original pencil (A, B) is transformed via unitary equivalences to generalized Schur form (S, T), where both S and T are upper triangular. The eigenvalues of (A, B) are readily found as the ratios of the diagonal elements of the pencil (S, T). The method consists conceptually of 2 phases, just like the QR algorithm:

1. A direct reduction of the matrix pair (A, B) to an equivalent Hessenberg, triangular matrix pair (H, R).

*Received by the editors February 12, 2018; accepted for publication (in revised form) May 21, 2019; published electronically August 6, 2019.
https://doi.org/10.1137/18M1170480
Funding: The work of the authors was partially supported by the Research Council KU Leuven projects C14/16/056 (Inverse-free Rational Krylov Methods: Theory and Applications) and OT/14/074 (Numerical algorithms for large scale matrices with uncertain coefficients).
†Department of Computer Science, KU Leuven, University of Leuven, 3001 Leuven, Belgium (daan.camps@kuleuven.be, karl.meerbergen@kuleuven.be, raf.vandebril@kuleuven.be).
[1]Regular means that the characteristic polynomial differs from zero.

2. An iterative phase during which deflating subspaces of the matrix pair (H, R) are determined and the matrix pair is essentially reduced to the triangular, triangular pair (S, T).

Various modifications and additions to the original algorithm have been proposed after its original introduction. Kaufman [13] added a deflation strategy and Ward [30] further refined various aspects of the method. Watkins and Elsner [35] generalized the QZ algorithm to a class of GZ iterations, and more recently, Kågström and Kressner [11] incorporated an aggressive early deflation strategy into a multishift QZ iteration. For more information we refer the reader to the monographs of Watkins [33] and Kressner [14].

Vandebril and Watkins [29] proposed a generalization beyond the Hessenberg, upper triangular pair. Their QZ like method reduces the matrix pair (A, B) to *condensed* form and iterates directly on the condensed matrix pair. A matrix pair (A, B) is said to be a condensed pair if both matrices are Hessenberg matrices and if there is exactly one nonzero element for every subdiagonal position which can be either at A or B. The classical Hessenberg, triangular format used in the QZ method is a special instance of a condensed matrix pair which maintains all zero subdiagonal elements at the side of B.

In this paper we propose a further generalization of the QZ method beyond condensed pairs. We will call this method the *rational QZ* (RQZ) method as it links strongly to rational Krylov subspaces [2]. As we will demonstrate in detail in the remainder of the paper, Hessenberg pairs and the associated rational Krylov subspaces are determined by *poles* that can be exploited to improve the convergence of the method. Both the original QZ algorithm [16] and the condensed QZ algorithm [29] turn out to be special instances of the RQZ method determined by a specific choice of poles. An implementation of the RQZ method is publicly available at http://numa.cs.kuleuven.be/software/rqz.

This article is closely related to the article by Berljafa and Güttel [2]. Starting from a rational Krylov subspace and the linked Hessenberg pair, their article discusses a way to change the poles by operating solely on the Hessenberg pair. We will see in this article that their way of introducing poles and moving them is related to introducing a shift and chasing it, like in typical QR algorithms. We will extend these results and formulate an implicit QZ algorithm that executes nested subspace iteration driven by a rational function. Moreover, in the theoretical analysis we directly rely on the pair (A, B) instead of rephrasing the relations in terms of a single matrix AB^{-1} or $B^{-1}A$ as is usually done.

This paper is organized as follows. The notion of a Hessenberg pair is formally defined in section 2, its properties are studied subsequently, and two types of operations on Hessenberg pairs are discussed: the introduction of a new pole and the *swapping* of poles. Section 3 proposes a method to reduce a general matrix pair to a Hessenberg pencil by means of unitary equivalence transformations. This is the RQZ analogue of the initial reduction phase in the QZ algorithm. The generalization of the iterative phase is presented in section 4. It is illustrated how an RQZ step with a single shift can be performed implicitly and numerical experiments illustrate the speed and accuracy. An implicit Q theorem for Hessenberg pairs is stated and used to prove that the RQZ iteration implicitly performs nested subspace iteration driven by a set of rational functions in sections 5 and 6. In section 7 we apply the RQZ method to filter a rational Krylov subspace in an iterative method. We conclude in section 8.

In this article we adopt the following notational conventions. Scalars $\alpha, \beta, \ldots$ are denoted with Greek letters and matrices $A, B, \ldots$ with capital Latin letters. Vectors

$\boldsymbol{a}, \boldsymbol{b}, \ldots$ are denoted in lowercase boldface Latin letters. The entry on row i and column j of A is denoted as a_{ij}, and column i of A as $\boldsymbol{a}_i$. The colon notation of MATLAB is sometimes used to denote part of a matrix: $A_{i:j,:}$ stands for rows i to j of A. I is the identity matrix, and $\boldsymbol{e}_i$ is its ith column. A^* is the Hermitian conjugate of A, and $\mathcal{R}(A)$ is the column space of A. $\mathcal{E}_k = \mathcal{R}(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k)$ is the subspace spanned by the k first canonical basis vectors. $\mathcal{K}_k(A, \boldsymbol{v}) = \mathcal{R}(\boldsymbol{v}, A\boldsymbol{v}, \ldots, A^{k-1}\boldsymbol{v})$ is the Krylov subspace of order k generated by A from $\boldsymbol{v}$. The complex plane extended with the point at infinity, $\mathbb{C} \cup \{\infty\}$, is denoted as $\bar{\mathbb{C}}$. For all nonzero scalars $\alpha \in \mathbb{C}$, we define $\alpha/0 = \infty$ and $\alpha/\infty = 0$.

2. Hessenberg pairs and their poles. In this section we repeat necessities from the literature and introduce some basic concepts linked to Hessenberg pairs. These pairs appear naturally in the context of the rational Krylov method introduced and studied by Ruhe [18, 19, 20, 21]. We will elaborate on this connection in section 7.

2.1. Proper Hessenberg pairs. A matrix H is of Hessenberg form if all its elements below the first subdiagonal are zero. A *proper* or *irreducible Hessenberg* matrix has all its subdiagonal elements different from zero. Being proper ensures that there are no obvious *deflations* allowing us to split the Hessenberg matrix into block upper triangular form with smaller submatrices. For a pair of Hessenberg matrices there is a subtlety, as there are two less obvious possibilities for deflation.

DEFINITION 2.1 (proper Hessenberg pair). *A pair of Hessenberg matrices $A, B \in \mathbb{C}^{n\times n}$ is said to be proper (or irreducible) if the following two conditions are met:*

I. *There is no i in $1, \ldots, n-1$ so that $a_{i+1,i}$ and $b_{i+1,i}$ are simultaneously zero.*
II. *The first columns of A and B are linearly independent, as are the last rows of A and B.*

For a proper Hessenberg pair we define its ordered pole tuple as $\Xi = (\xi_1, \ldots, \xi_{n-1})$, $\xi_i \in \bar{\mathbb{C}}$, where $a_{i+1,i}/b_{i+1,i} = \xi_i$ for all i from 1 to $n-1$.

The ratios of the subdiagonal elements of A over the subdiagonal elements of B are thus called the poles of the proper Hessenberg pair. Since we set division by zero equal to ∞ in $\bar{\mathbb{C}}$, a pole is located at ∞ if the respective subdiagonal element of B is zero.

The first condition of being proper means that all poles are well-defined over $\bar{\mathbb{C}}$, so there is no $0/0$. Just like in the classical case $a_{i+1,i} = b_{i+1,i} = 0$ allows us to deflate the problem into two independent subproblems.

The second condition is less obvious, but it is simple to deflate an eigenvalue if it is not met. Construct a rotation Q_1, acting on the first two rows such that Q_1^* maps the first column of A and B in the direction of $\boldsymbol{e}_1$, and then the pair $Q_1^*(A, B)$ allows for a deflation. Similarly we can construct a rotation Z_{n-1} to transform $(A, B)Z_{n-1}$ to a deflatable format in the case the last rows are linearly dependent. If condition II does not hold, then the pair can be transformed into an equivalent pair for which condition I does not hold in the first or last subdiagonal position.

We remark that even if condition II of the definition of a proper Hessenberg pair were not met, we still define the first pole ξ_1 and last pole ξ_{n-1} as in Definition 2.1, provided they are in $\bar{\mathbb{C}}$. Suppose there is some scalar γ such that $\boldsymbol{a}_1 = \gamma \boldsymbol{b}_1$ with $\boldsymbol{a}_1$ and $\boldsymbol{b}_1$ the first columns of A and B, respectively. This means that γ is both the first pole, $\xi_1 = a_{21}/b_{21} = \gamma$, and an eigenvalue, $A\boldsymbol{e}_1 = \gamma B\boldsymbol{e}_1$. Similarly the last pole ξ_{n-1} is an eigenvalue if the last rows of A and B are linearly dependent.

Properness of the Hessenberg matrix ensures *essential uniqueness* of the QR iterates, which is crucial in the design of an *implicit QR* algorithm [8, 9] for the standard eigenvalue problem. We will prove in section 5 that also proper Hessenberg pairs inherit a type of essential uniqueness allowing for the design of an implicit method, which is the implicit Q theorem for Hessenberg pairs.

The other pencils for which QZ algorithms were designed fit in Definition 2.1. Pairs in Hessenberg, triangular form [16] are proper with poles $\Xi = (\infty, \infty, \ldots, \infty)$; a pair of matrices in condensed form [29] is also a proper Hessenberg pair with poles being either 0 or ∞.

The properties of proper Hessenberg pairs discussed in the next lemma are frequently used throughout the paper.

LEMMA 2.2. *Let $(A, B) \in \mathbb{C}^{n\times n}$ be a proper Hessenberg pair with poles $\Xi = (\xi_1, \ldots, \xi_{n-1})$. Then the following statements hold:*

I. *For $\mu, \nu \in \mathbb{C}$, such that $\mu/\nu \notin \Xi$, we have that $(\nu A - \mu B)$ is a proper Hessenberg matrix.*

II. *For $\mu, \nu \in \mathbb{C}$, such that μ/ν is equal to a certain pole ξ_k $(1 \leq k \leq n-1)$, we have that $N = (\nu A - \mu B)$ is block upper triangular,*

$$N = \begin{bmatrix} N_{11} & N_{12} \\ & N_{22} \end{bmatrix},$$

where N_{11} and N_{22} are Hessenberg matrices of sizes $k\times k$ and $(n-k)\times(n-k)$, respectively.

III. *For $\mu, \nu, \alpha, \beta \in \mathbb{C}$, such that $\mu\beta \neq \alpha\nu$, we have that*

$$(M, N) = (\beta A - \alpha B, \nu A - \mu B)$$

is a proper Hessenberg pair with poles $\frac{\beta a_{k+1,k} - \alpha b_{k+1,k}}{\nu a_{k+1,k} - \mu b_{k+1,k}}$ for $k = 1, \ldots, n-1$.

IV. *For $k = 1, \ldots, n-1$ we have that $\mathcal{R}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k) \neq \mathcal{R}(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k)$.*

Proof. Statements I and II are trivial. The pencil of statement III satisfies the definition of a proper Hessenberg pair: M and N are clearly upper Hessenberg matrices, and their kth subdiagonal elements are

$$\begin{bmatrix} m_{k+1,k} \\ n_{k+1,k} \end{bmatrix} = \begin{bmatrix} \beta & -\alpha \\ \nu & -\mu \end{bmatrix} \begin{bmatrix} a_{k+1,k} \\ b_{k+1,k} \end{bmatrix}.$$

The vector on the left is different from zero since the matrix is nonsingular and the vector on the right is nonzero. The first column of M is also linear independent from the first column of N because the same nonsingular matrix is used in the transformation. The same holds for the last row. The proof of statement IV is by induction and contradiction. The case $k = 1$ follows from the definition of a proper Hessenberg pair. Suppose the statement holds up to column k. We assume now, by contradiction, that it breaks down at column $k+1$, and thus $\mathcal{R}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k+1}) = \mathcal{R}(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{k+1})$. The equality implies the existence of a $(k+1)\times(k+1)$ matrix C such that

$$[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k+1}] = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{k+1}] \begin{bmatrix} c_{11} & \cdots & c_{1,k+1} \\ \vdots & \ddots & \vdots \\ c_{k+1,1} & \cdots & c_{k+1,k+1} \end{bmatrix}. \tag{2}$$

It follows from the induction hypothesis that there is a j with $1 \leq j \leq k$ such that

$\boldsymbol{a}_j \notin \mathcal{R}(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k)$. Therefore $c_{k+1,j} \neq 0$. By the Hessenberg structure,

$$0 = a_{k+2,j} = \sum_{i=1}^{k+1} b_{k+2,i}\, c_{i,j} = b_{k+2,k+1} c_{k+1,j}.$$

This implies that $b_{k+2,k+1}$ must be zero. Equation (2) consequently implies that also $a_{k+2,k+1} = 0$. These two values being simultaneously zero contradicts the properness. ∎

2.2. Manipulating the poles of a Hessenberg pair. In this section we will revisit two operations for manipulating the poles of a Hessenberg pair, namely, changing the first or the last pole and swapping poles (see also Berljafa and Güttel [2]).

Changing poles at the boundaries. Let $A, B \in \mathbb{C}^{n\times n}$ be a proper Hessenberg pair and assume the first pole ξ_1 different from the eigenvalues of (A, B). The pole ξ_1 can be changed to another pole $\hat{\xi}_1 \in \bar{\mathbb{C}}$ by multiplying (A, B) from the left with a unitary transformation Q_1^*, where $Q_1^*\boldsymbol{x} = \alpha\boldsymbol{e}_1$ and

$$\begin{aligned} \boldsymbol{x} &= \hat{\gamma}\,(\hat{\beta}_1 A - \hat{\alpha}_1 B)(\beta_1 A - \alpha_1 B)^{-1}\boldsymbol{e}_1 \\ &= \gamma\,(A - \hat{\xi}_1 B)(A - \xi_1 B)^{-1}\boldsymbol{e}_1 \end{aligned} \tag{3}$$

with γ and $\hat{\gamma}$ convenient scaling factors, and $\hat{\alpha}_1, \hat{\beta}_1, \alpha_1, \beta_1 \in \mathbb{C}$ are chosen to satisfy the new pole $\hat{\xi}_1 = \hat{\alpha}_1/\hat{\beta}_1$ and the old pole $\xi_1 = \alpha_1/\beta_1$. The notation with α and β to denote $(\beta A - \alpha B)$ is factually the most correct one. For notational simplicity, however, we will often use the shorthand notation $(A - \xi B)$, where $\xi = \alpha/\beta$ instead. As $\hat{\xi}_1 \neq \xi_1$, otherwise nothing needs to be done, $\boldsymbol{x}$ must be a vector with only the two leading elements nonzero and thus Q_1 is always well defined and can, for example, be chosen as a rotation matrix.

If Q_1 is used to compute $(\hat{A}, \hat{B}) = Q_1^*(A, B)$, then $\hat{\xi}_1$ will become the first pole of $(\hat{A}, \hat{B})$ because the first subdiagonal element of $(\hat{A} - \hat{\xi}_1\hat{B})$ is zero:

$$\begin{aligned} (\hat{A} - \hat{\xi}_1\hat{B})\boldsymbol{e}_1 &= Q_1^*(A - \hat{\xi}_1 B)\boldsymbol{e}_1 \\ &= \tilde{\gamma}\, Q_1^*(A - \hat{\xi}_1 B)(A - \xi_1 B)^{-1}\boldsymbol{e}_1 = \frac{\tilde{\gamma}}{\gamma}\, Q_1^*\boldsymbol{x} = \frac{\alpha\,\tilde{\gamma}}{\gamma}\boldsymbol{e}_1. \end{aligned}$$

Theoretically, under the assumption that B is nonsingular, we could equally well define $\boldsymbol{x} = \gamma(AB^{-1} - \hat{\xi}_1 I)(AB^{-1} - \xi_1 I)^{-1}\boldsymbol{e}_1$. Practically, however, to avoid the nonsingularity assumption of B, and for reasons of numerical stability, we stick to (3).

Remark 2.3. As $(A - \xi_1 B)^{-1}\boldsymbol{e}_1$ is a scalar multiple of $\boldsymbol{e}_1$ there is no need to compute this in practice. Moreover, even if ξ_1 is an eigenvalue, a scalar multiple of $\boldsymbol{e}_1$ is always a solution of $(A - \xi B)\boldsymbol{y} = \boldsymbol{e}_1$. The inverse factor is included to emphasize the rational function used to update the pole and moreover, it is consistent with the analysis of Vandebril and Watkins [28, 29], where it does play a role in the multishift setting. In practice we compute $\boldsymbol{x} = \gamma(A - \hat{\xi}_1 B)\boldsymbol{e}_1$ in $O(1)$ operations.

We can compute an equivalence transformation to change the last pole, by operating from the right on the Hessenberg pair in a comparable way. Assume ξ_{n-1} different from the eigenvalues of (A, B). We can change the pole ξ_{n-1} to $\hat{\xi}_{n-1} \in \bar{\mathbb{C}}$. If we consider the row vector $\boldsymbol{x}^T = \gamma\boldsymbol{e}_n^T(A - \xi_{n-1}B)^{-1}(A - \hat{\xi}_{n-1}B)$, with γ a convenient scaling factor, and a transformation Z_{n-1} that introduces a zero in the penultimate position of $\boldsymbol{x}^T$, $\boldsymbol{x}^T Z_{n-1} = \alpha\boldsymbol{e}_n^T$, then the last pole in the Hessenberg pair $(\hat{A}, \hat{B}) = (A, B)Z_{n-1}$ is changed to $\hat{\xi}_{n-1}$.

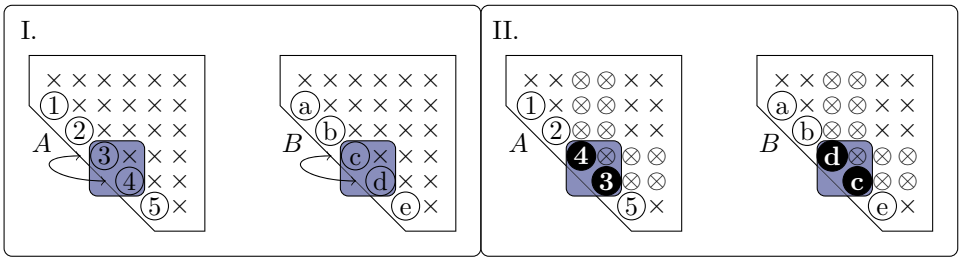


FIG. 1. *Swapping poles in a Hessenberg pair:* (I) *before swap,* (II) *after swap.*

Again, the system $\boldsymbol{e}_n^T(A-\xi_{n-1}B)^{-1}$ is never solved in practice as the solution is a scalar multiple of $\boldsymbol{e}_n^T$, but is only included for theoretical purposes.

Swapping poles. Any two consecutive poles ξ_i and ξ_{i+1} in a proper Hessenberg pair (A, B) can be *swapped* via a unitary equivalence on (A, B). We assume both poles to be different; otherwise nothing needs to be done. This procedure is illustrated in Figure 1, where poles $\xi_3 = ③/ⓒ$ and $\xi_4 = ④/ⓓ$ are swapped. The swapping is achieved by computing unitary matrices Q_4 and Z_3 that change the order of the eigenvalues in the 2×2 blocks $A_{4:5,3:4}$ and $B_{4:5,3:4}$. These blocks are indicated with the shaded region in Figure 1. The equivalence transformation affects all elements marked with $\otimes$ in pane II of Figure 1. Note that the ratios ❹/❹ and ❸/❸ are preserved under swapping but the subdiagonal values themselves can change. Swapping diagonal elements in an upper triangular matrix is a classical problem, also used to reorder eigenvalues in the (generalized) Schur form. It can be solved as the solution of a coupled Sylvester equation [12] or by direct computations [26]. Its solution is unique if ξ_i differs from ξ_{i+1}.

Details and solutions are found, e.g., in Watkins [31], Kågström and Poromaa [12], and Van Dooren [26]. In [26] it is also proved that the problem can be solved in a backward stable manner.

3. Direct reduction to a proper Hessenberg pair. The rational QZ algorithm we propose in section 4 operates on a proper Hessenberg pair. If we are given an arbitrary matrix pencil (A, B) not yet in (proper) Hessenberg form, we first need to transform it to this form. We use equivalences since we are interested in the eigenvalues and, for reasons of numerical stability, we will stick to unitary equivalences. At the end of the section we will illustrate with a numerical experiment that good pole selection can lead to deflations, already in the reduction process.

3.1. The reduction algorithm. We will transform an $n\times n$ matrix pair (A, B) to a unitary equivalent Hessenberg pair with a prescribed tuple of poles $\Xi = (\xi_1, \ldots, \xi_{n-1})$. The algorithm proceeds similarly to the direct reduction to Hessenberg, triangular form, with the difference that a pole is introduced at every step.

As in the classical reduction to a Hessenberg, upper triangular pair we commence with computing a QR factorization of $B = QR$ and updating the matrix pair to (Q^*A, Q^*B). The matrix Q^*B is now already in upper triangular form. This is shown in pane I of Figure 2 for our running example matrix pair of size 5×5. Moreover, we assume in the remainder of this section that all zeros on the diagonal of B—infinite eigenvalues—are removed [33].

We will now bring the first column of A to Hessenberg form. In pane II, a zero is introduced in position $(5, 1)$ of matrix A by operating on the last two rows. This

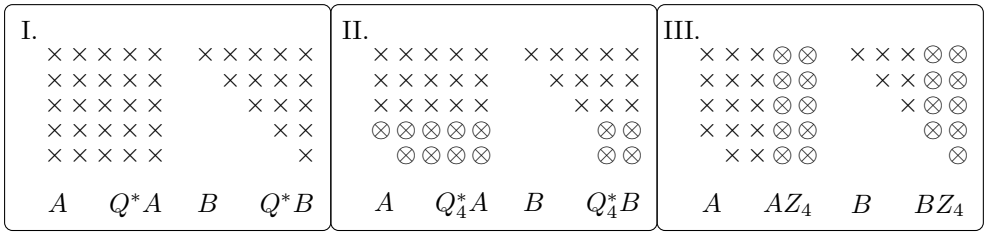


FIG. 2. *Reduction to a Hessenberg pencil. First part.*

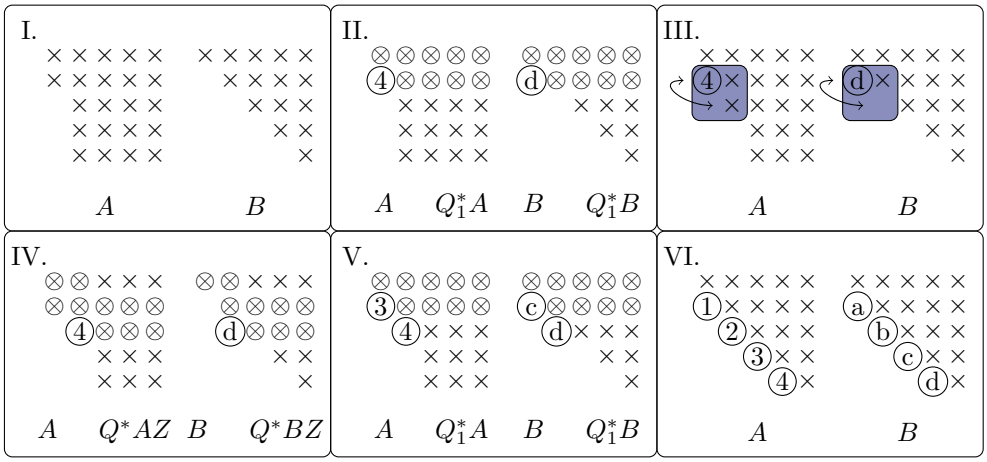


FIG. 3. *Reduction to a Hessenberg pencil. Second part.*

destroys the upper triangular shape in the last two rows of B. The upper triangular shape can be restored by acting on columns 4 and 5 as shown in pane III without destroying the newly created zero in A.

The process of introducing zeros in the first column of A by acting on the rows and maintaining the upper triangular structure in B by acting on the columns can be repeated until the first column of A is brought to upper Hessenberg shape. This coincides with the standard reduction to a Hessenberg, triangular pair [33]. We have arrived at pane I of Figure 3. The first column of (A, B) is now already in the correct form, but has a pole at ∞. We replace ∞ by another pole using the techniques from subsection 2.2 applied to the first column of (A, B) which is in Hessenberg form. This is always possible, except when there is an obvious deflation in the top left corner. This does not pose any problems: deflate and continue. We start by introducing the last pole $\xi_4 = ④/ⓓ$ first; as in the following steps of the reduction procedure this pole will move down to end up at the correct position at the bottom of the subdiagonal. The current state of the pair is visualized in pane II of Figure 3.

The second column has been brought to Hessenberg, triangular form in pane III of Figure 3 via the classical procedure of introducing zeros in the second column of A and maintaining the upper triangular structure in B. This procedure does not affect the existing pole ξ_4. At this stage, the first pole equals ξ_4, while the second pole is ∞. The poles in the shaded region of pane III are now swapped via the techniques

from subsection 2.2, which moves the pole at ∞ to the top of the matrix pair in pane IV. The swapping technique can be used, as the two leading columns of (A, B) are in Hessenberg form at this stage of the reduction algorithm. The swapping is always well defined, even if there is a succession of identical poles. The pole ξ_4 has moved one row down and one column to the right. The pair is now ready for the introduction of pole ξ_3 as shown in pane V. This entire process of creating zeros, swapping poles, and introducing a new pole can be repeated until the end result of pane VI is obtained, and the matrix is in the desired Hessenberg form.

After the reduction process, the matrix does not necessarily need to be in proper Hessenberg form. Possibly the pole ξ_{n-1} coincides with an eigenvalue, allowing for deflation in the lower right corner. In this case one deflates ξ_{n-1} and checks whether ξ_{n-2} leads to a deflation, and so forth, until the matrix has become proper. It can also happen that during the reduction any of the interior poles deflate. In this case the reduction can be continued on the separated parts of the pencil. This situation is studied in the example of subsection 3.2.

The introduction of the poles takes an additional $6n^3$ flops on top of the $8n^3$ operations required to reduce a pencil to Hessenberg, triangular form [10].

3.2. Numerical experiment. We study two matrix pairs from the magnetohydrodynamics (MHD) dataset available in the Matrix Market collection [3]. The matrices are of sizes 416 and 1280, respectively, and are known to be ill-conditioned. They originate from a Galerkin finite element discretization of the underlying MHD problem. Their spectrum consists of a tail along the negative real axis and a set of eigenvalues close to the imaginary axis. In this numerical experiment we determine deflating subspaces for the two regions of eigenvalues already during the reduction phase. The tests were run in MATLAB R2017b.

The idea is to introduce poles that make up a rational filter that mainly affects one region of eigenvalues. To achieve this effect, the poles are chosen on a contour Γ in the complex plane that contains the eigenvalues along the negative real axis. This approach is inspired by the link between contour integration methods [17, 22] and rational filtering techniques [24, 25]. In section 6 we explain in full detail how introducing and swapping poles implicitly applies a rational filter.

The poles are chosen on an elliptical contour $\Gamma = e(c, r_x, r_y, \theta)$, where c is the center of the ellipse, r_x is the radius in x-direction, r_y is the radius in the y-direction, and θ is the angle over which the axes of the ellipse are rotated. For the smaller problem, Γ is selected as $e(-1.3, 1.5, 3, 0)$ and discretized in 120 nodes. For the larger problem, $\Gamma = e(-25, 27, 6, 0)$ and it is discretized in 400 nodes. These nodes are the poles introduced during the reduction to Hessenberg form. The aim is to get the pair improper, enforcing thereby a middle deflation separating the two regions. In the case of a *middle* deflation we continue introducing poles on the separated parts.

The results are presented in Figures 4 to 6. Figure 4 shows an overview of the spectrum of both matrix pairs. The two regions of eigenvalues are indicated with different markers. The box in Figure 4 marks the area in which Figure 5 will zoom in; it shows where the regions meet in detail, together with the poles of the Hessenberg pair.

Figure 6 displays the magnitude of the subdiagonal elements $|a_{i+1,i}|+|b_{i+1,i}|$. All poles which are considered numerically zero and thus lead to a deflation are emphasized in a shaded rectangle. Typically some of the first and last poles are deflated, but more important is the presence of interior deflations. This happens at poles 103 to 106 after 160 poles have been introduced in the pair of size 416. For the larger pair,

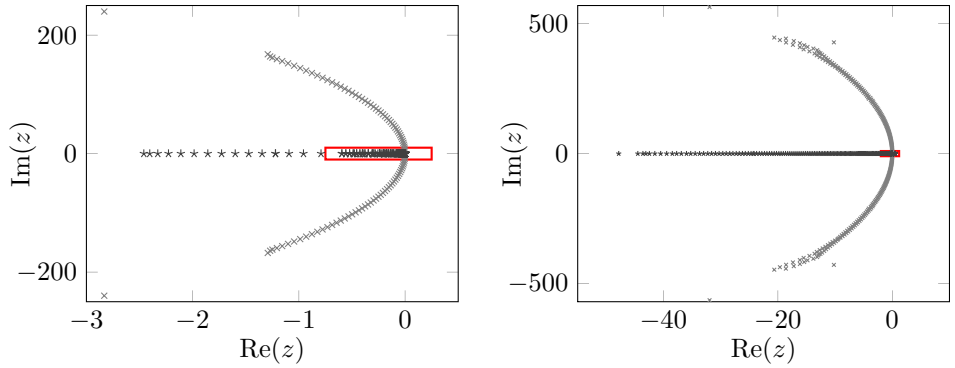


FIG. 4. *Eigenvalues in region* 1 (×; *bow-shape) and region* 2 (⋆; *close to the real axis). On the left we have the problem of size* 416 *and on the right* 1280.

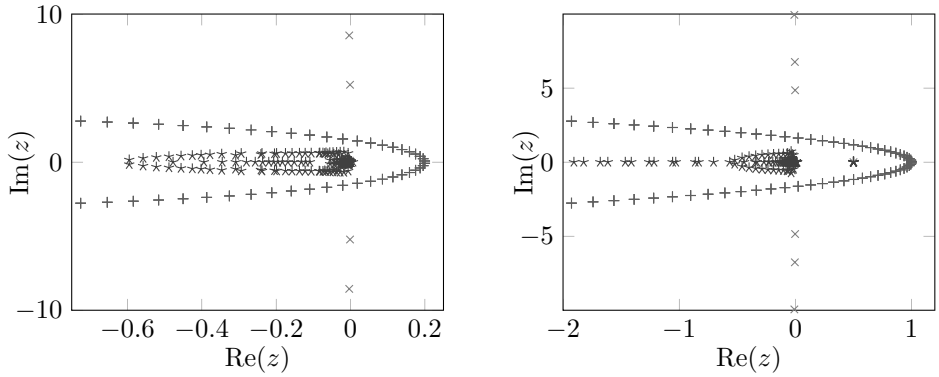


FIG. 5. *Close-up of the central part where the regions meet for the problem of size* 416 *and* 1280. *The legend is identical to the one of Figure* 4 *extended with the poles* (+; *on the ellipse around the real axis).*

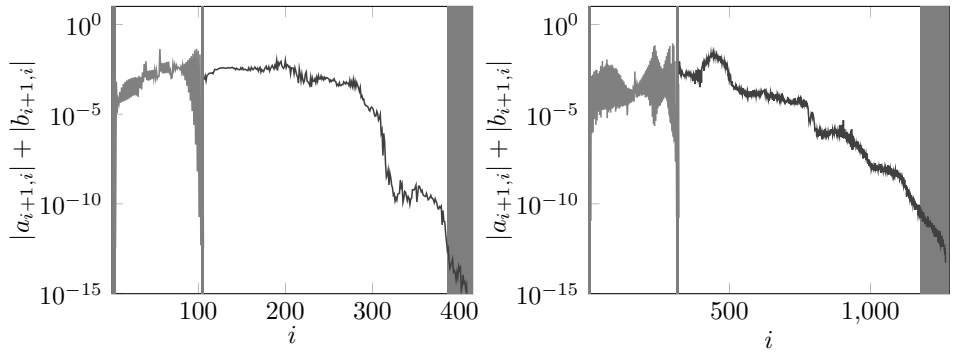


FIG. 6. *Magnitudes of the subdiagonal elements in the matrix pair after the Hessenberg reduction for the problem of size* 416 *(left) and* 1280 *(right).*

poles 317 to 321 are deflated after 621 poles have been introduced. The eigenvalues outside Γ are located in the top left part of the Hessenberg pair, and those inside Γ appear after the interior deflation.

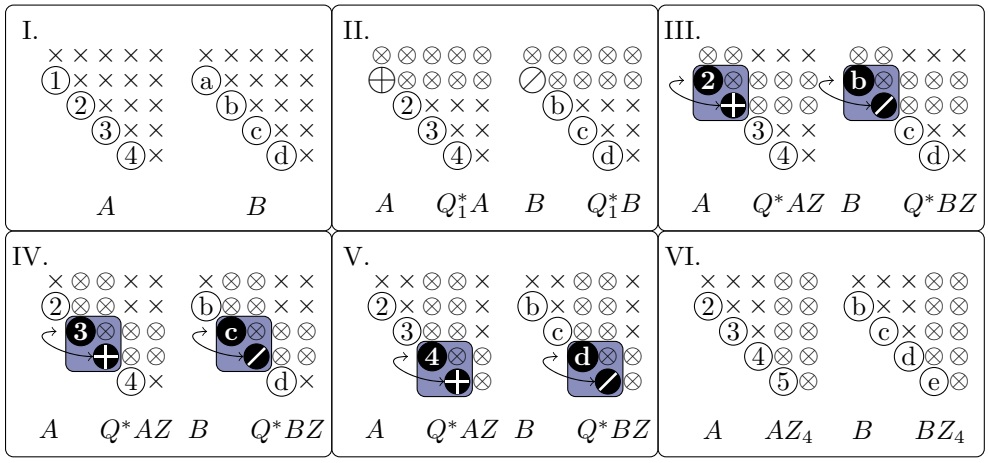


FIG. 7. *Single shifted implicit RQZ step on a* 5×5 *Hessenberg pair with shift* ϱ.

This numerical experiment shows that deflating subspaces containing regions of eigenvalues can be found already during the reduction to Hessenberg form. We would like to stress that deflation is obtained without any of the poles converging toward an eigenvalue, but by choosing poles on a contour such that they construct an effective rational filter.

4. Implicitly single shifted rational QZ step. In this section we present the implicit RQZ step for a Hessenberg pair. Numerical experiments are included at the end of this section to illustrate the performance and accuracy of the algorithm.

The algorithm operates on proper Hessenberg pairs. These pairs could be the result of the reduction procedure presented in section 3 or they could be given directly, e.g., as coming from an iterative rational Krylov method, where one would like to compute the eigenvalues of the projected Hessenberg pair. These eigenvalues are approximations to the eigenvalues of the original problem and are called the Ritz values if the final pole is at ∞ or Harmonic Ritz values for a final pole at 0 [4, 21].

4.1. The algorithm. Before describing the algorithm we would like to comment on the nomenclature. We use both the terms *poles* ξ and *shifts* ϱ to refer to elements on the subdiagonal of a Hessenberg pair. In fact our shifts are poles as well, but we typically consider poles as subdiagonal elements that are sustained in the Hessenberg pair, while shifts are introduced and removed in a single implicit RQZ step. A shift is pushed in at the top, chased to the bottom, and pulled out at the end.

We introduce the RQZ procedure with an example. Given a 5×5 Hessenberg pair (A, B) with poles $\xi_1 = ①/ⓐ$, $\xi_2 = ②/ⓑ$, $\xi_3 = ③/ⓒ$, $\xi_4 = ④/ⓓ \in \bar{\mathbb{C}}$, the RQZ step consists of three stages, similar to all algorithms of implicit QR type. These are an initialization, a chasing, and a finalization phase.

Initialization. Suppose we are given a shift $\varrho = \oplus/\oslash \in \bar{\mathbb{C}}$, for instance, the Wilkinson shift. Pane I in Figure 7 shows the Hessenberg pair in its initial state. The shift[2] is introduced in pane II by changing the first pole with a transformation Q_1.

[2]A shift equal to a pole will not result in a breakdown, but leads to slow or no convergence at all (see section 6). In practice shifts should be taken differently from the poles.

Chasing. Panes III–V show how the shift is relocated from the first position on the subdiagonal to position $n-1$ by repeatedly swapping it with the poles of the Hessenberg pair. The shift is chased to the bottom. The matrix elements that are changed in every step are marked with an $\otimes$.

During this procedure the shift will move from the top-left to the bottom-right and all poles will move up one position in the direction of the top-left corner. The assumption that the shift differs from the poles $\varrho \neq \xi_i$ for all i ensures that none of the swapping operations equals an identity; otherwise the downward movement of the shift will undo the upward movement of the corresponding pole.

Finalization. Finally, in pane VI, one last operation can be performed where we have the possibility to remove the shift ϱ and introduce any new pole $\hat{\xi}_4 = ⑤/ⓔ \in \bar{\mathbb{C}}$, via the procedure described in subsection 2.2.

The reader familiar with the classical QZ algorithm [16] or the condensed QZ algorithm [29] can verify that the algorithm described here generalizes both methods. In the QZ algorithm [33] one chases a bulge and in the final step the new pole was always put to ∞ thereby restoring the upper triangular form of B. In the condensed QZ algorithm [29] a rotation was chased and in the final step one allowed for a pole to be at 0 or ∞.

In the rational QZ algorithm we chase a shift instead of a bulge or a rotation. However, the shift is encoded in the rotation and bulge as well, as it is found as an eigenvalue of Watkins' bulge pencil [32, section 5], [33]; the other eigenvalue in the bulge pencil is ∞. If we consider the same bulge pencil in the rational QZ case we see that the eigenvalue at ∞ is replaced by a pole of the pencil. Moreover, also the pole swapping technique is nothing other than the bulge exchange interpretation of Watkins [31].

4.2. Shifts, poles, and deflation. In order to implement the RQZ algorithm and in particular a single RQZ step, we need good strategies to select the shift, the new pole introduced at the very end, and a procedure to check if there are deflations.

For the shifts we typically take the Wilkinson shift. This is the eigenvalue of the trailing 2×2 block that is closest to a_{nn}/b_{nn}. For the poles there are several options: one could as well consider a Wilkinson strategy determined by the 2×2 block in the upper-left corner or one could use other techniques such as poles on a contour to do filtering (see, e.g., subsection 3.2). Optimal pole selection is a difficult issue and very problem specific, and this is beyond the scope of this manuscript; in the numerical experiments we will test some straightforward options.

The deflation criterion for the poles $\xi_2, \ldots, \xi_{n-2}$ is obvious. If one of these is not in $\bar{\mathbb{C}}$, the problem can be split into smaller, independent problems. This means in fact that for a certain i, two subdiagonal elements $a_{i+1,i}$ and $b_{i+1,i}$ are simultaneously zero. To numerically check this we use the classical relative criterion taking the sizes of the neighboring elements into consideration [10], i.e., $|a_{i+1,i}| \leq c\epsilon_m(|a_{i,i}|+|a_{i+1,i+1}|)$ and $|b_{i+1,i}| \leq c\epsilon_m(|b_{i,i}|+|b_{i+1,i+1}|)$, with ϵ_m the machine precision and c a small constant.

The situation for the exterior poles ξ_1 and ξ_{n-1} is more peculiar. Whereas the interior poles are fixed, the exterior ones can be altered. Instead of changing ξ_1 or ξ_{n-1} to another pole, we would like to know whether it is possible to move them outside of $\bar{\mathbb{C}}$: we would like to deflate an eigenvalue. To this end we need to create two zeros with a single operation such that the pair is no longer proper. We discuss the situation at the bottom-right; the top-left corner proceeds similarly. Suppose we would like to simultaneously introduce zeros in the final subdiagonal positions, which are $a_{n,n-1}$ and $b_{n,n-1}$. This is possible if the matrix $\left[\begin{smallmatrix} a_{n,n-1} & a_{nn} \\ b_{n,n-1} & b_{nn} \end{smallmatrix}\right]$ is of rank 1. If this

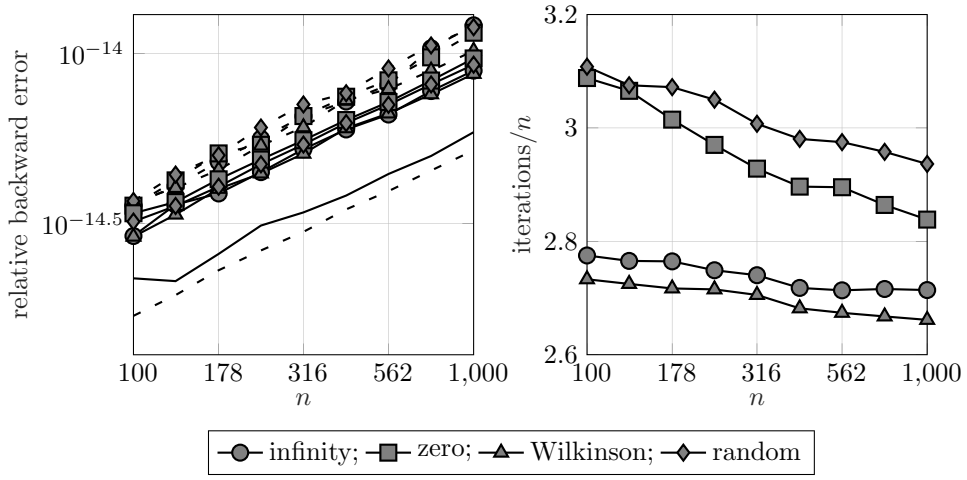


FIG. 8. *On the left the relative backward errors related to the reduction to a Hessenberg pair (no markers) and to the Schur form (with markers) are demonstrated. The error on A is represented with a dashed line and the error on B with a full line. On the right we see the average number of iterations per eigenvalue for the four different pole strategies.*

is the case, we can simultaneously annihilate the subdiagonal elements by operating on the final two columns of (A, B). In our experiments we assume the matrix to be of rank 1 if $\sigma_{\min}/\sigma_{\max} < \epsilon_m$.

4.3. Numerical experiment. We apply the RQZ method on two sets of problems: random matrix pairs and two problems from fluid dynamics. We are interested in the accuracy and speed.

Random matrix pairs. We test the single shift RQZ algorithm on 9 randomly generated, complex-valued matrix pairs with sizes ranging from 100 to 1000. The results are averaged over 10 runs. The pairs are first reduced to Hessenberg pairs with all poles at infinity, implying that no additional computational work has been done compared to the reduction phase of the QZ method. The shift is always taken as the Wilkinson shift. The poles are selected according to four different strategies: poles at infinity, poles at zero, random poles, and poles chosen as the Wilkinson shift from the upper-left 2×2 block. The last choice is called the Wilkinson pole.

The results are summarized in Figures 8 and 9. The left pane of Figure 8 shows the relative backward errors $\|\hat{A} - Q^*AZ\|_2/\|A\|_2$ and $\|\hat{B} - Q^*BZ\|_2/\|B\|_2$ for the reduction to a Hessenberg pair (lines without markers) and the backward error on the Schur form for the four different pole strategies. The backward error is small in all cases. The right pane shows the average number of iterations per eigenvalue. Clearly, the Wilkinson pole requires the least number of iterations per eigenvalue. It requires on average 1.5% less iterations than the classic choice of poles at infinity. Random poles and poles at zero perform the worst.

Figure 9 shows the total number of pole swaps scaled with n^2. The scaling factor is used since the number of pole swaps per iteration is $O(n)$ and the number of iterations is also $O(n)$. This measure of performance depends heavily on the positions where deflations occur and as such gives a much better view on the algorithmic behavior. The order of the four strategies remains the same, but the gains with Wilkinson poles increase up to 4%. This signals the occurrence of deflations at

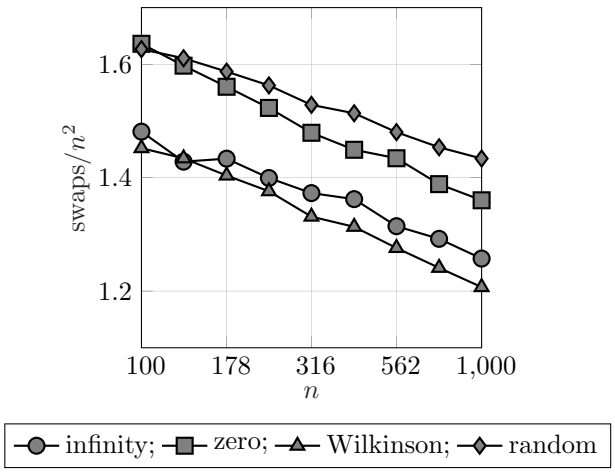


FIG. 9. *The total number of swaps scaled by n^2 for the computation of the Schur form for the four different pole strategies.*

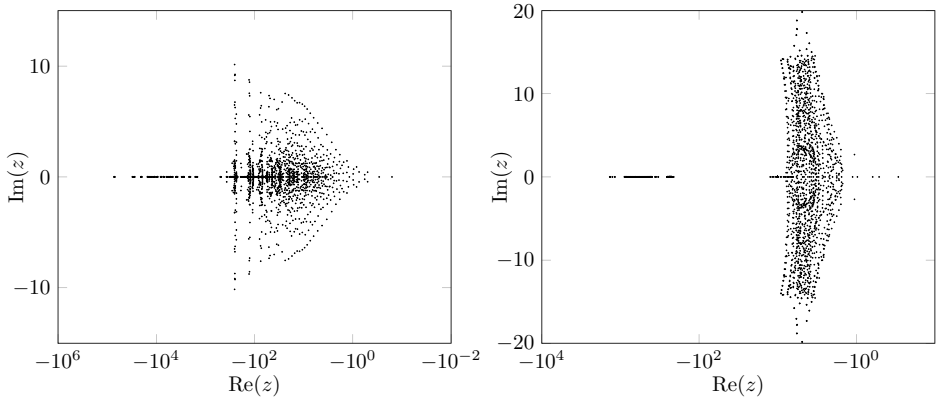


FIG. 10. *On the left the spectrum of the cavity flow problem and on the right the spectrum of the obstacle flow problem are shown.*

spots other than only in the lower-right corner as is typically the case in the classical setting.

IFISS problems. In this experiment we apply the RQZ method on two problems from fluid dynamics generated with IFISS [6, 7]. The first problem originates from a model for the flow in a unit-square cavity, and the second problem comes from a model for the flow around an obstacle. Both models are discretized, resulting in two real, generalized eigenvalue problems. The *cavity flow* problem is of size 2467, and the *obstacle flow* problem of size 2488. We applied the single shift RQZ method after initial reduction to Hessenberg form with poles at infinity. Wilkinson shifts are employed in all cases. We used poles at infinity and Wilkinson poles. The spectra of the matrices is shown in Figure 10.

The results of the experiment are summarized in Table 1. It lists the relative backward error on the Schur form for both A and B for both problems and the two pole strategies. The backward error is very good in all cases. The table also lists the average iterations per eigenvalue and how this compares relatively to the result of

TABLE 1

Results of the RQZ method on the IFISS problems. The first column lists the problem, and the second column the pole strategy. Columns 3 *and* 4 *present the backward error on* A *and* B*, columns* 5 *and* 6 *the average number of iterations and performance compared to QZ, and columns* 7 *and* 8 *the total number of swaps and the performance compared to QZ.*

Problem	pole	error A	error B	it/n	%	swaps/n^2	%
Cavity flow	∞	$7.5 \cdot 10^{-15}$	$4.4 \cdot 10^{-15}$	2.49	100	0.446	100
	Wilk.	$7.8 \cdot 10^{-15}$	$4.1 \cdot 10^{-15}$	2.34	94.2	0.443	99.3
Obstacle flow	∞	$9.2 \cdot 10^{-15}$	$7.8 \cdot 10^{-15}$	2.54	100	0.617	100
	Wilk.	$8.8 \cdot 10^{-15}$	$7.8 \cdot 10^{-15}$	2.36	93.0	0.595	96.3

poles at infinity. We observe that the average number of swaps and iterations when employing Wilkinson poles is always below the numbers generated by the classical approach.

4.4. Tightly packed shifts. The single shifted RQZ method is, just like the classical QZ method, sequential in nature and not very cache efficient. To enhance cache performance one can go for *multishift* and chase m shifts simultaneously or one can chase m single shifts as close as possible after each other. Since the theory in this manuscript is not suited for a multishift setting we will confine ourselves to a description and numerical experiment for *tightly packed shifts*.

Assume we would like to chase m tightly packed shifts, which are typically the eigenvalues of the bottom-right $m \times m$ block of (A, B). These shifts are introduced one after another in the Hessenberg pair. The first shift is introduced and swapped down one row. Next the second shift is introduced and both shifts need to be swapped down a single row, starting with the lower-right one first. As a result there is space to introduce the third shift, and the procedure continues. After having introduced the shifts, the first m subdiagonal elements of the pair (A, B) link to these shifts.

In order to chase the block of m shifts one needs to swap all shifts down one row, starting again with the one in lower-right corner first. In total there are m equivalence transformations which should be accumulated to update the necessary parts of the matrices in a cache efficient manner.

The finalization phase commences when the shifts occupy the last subdiagonal positions in the Hessenberg pair. We can now introduce m new poles. The first new pole is introduced in the final subdiagonal element and swapped up m positions thereby swapping all remaining shifts down. The second new pole is now introduced and this course of action continues until the new poles occupy the last m subdiagonal elements.

We test the tightly packed RQZ method on randomly generated matrix pairs of size 1000 that are first reduced to Hessenberg pairs with poles at infinity. We run the RQZ method for shift batches of sizes $m = 2, 4, 8, 16, 32$. The results are averaged over 2 runs. The poles are selected following three criteria: always at infinity (classical QZ), m times the Wilkinson pole of the leading 2×2 block, or as the eigenvalues of the leading $m \times m$ block, the *Rayleigh* poles.

Figure 11 displays the performance in terms of the average number of iterations per eigenvalue (left) and total number of swaps scaled with n^2 (right) in function of the batch size m for the three types of poles. We observe that the number of iterations remains constant up to a batch size of 16 but increases significantly for $m = 32$. This effect is most pronounced with the Wilkinson and Rayleigh poles. Also in terms of the number of swaps the poles at infinity are the most efficient choice for $m = 32$. We

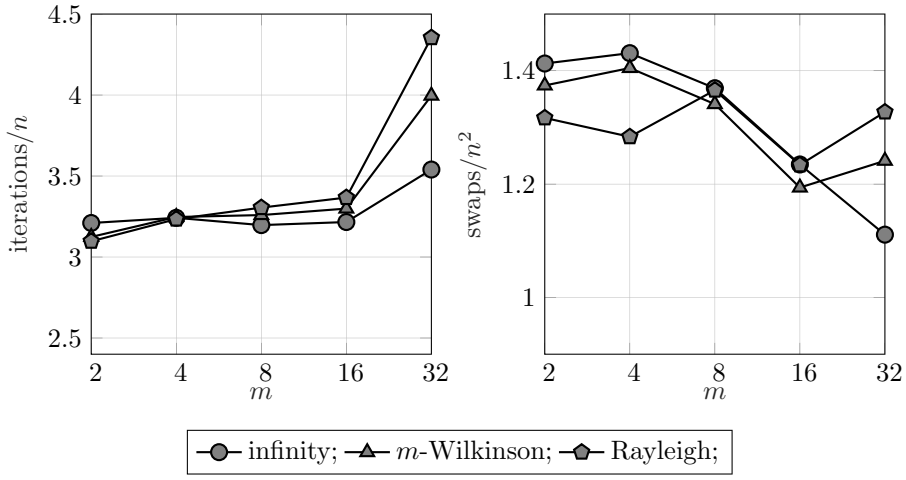


FIG. 11. *On the left the average number of iterations per eigenvalue is depicted in function of batch size m for three different pole strategies. On the right the average number of swaps scaled with n^2 in function of the batch size m. These results are for the random problem.*

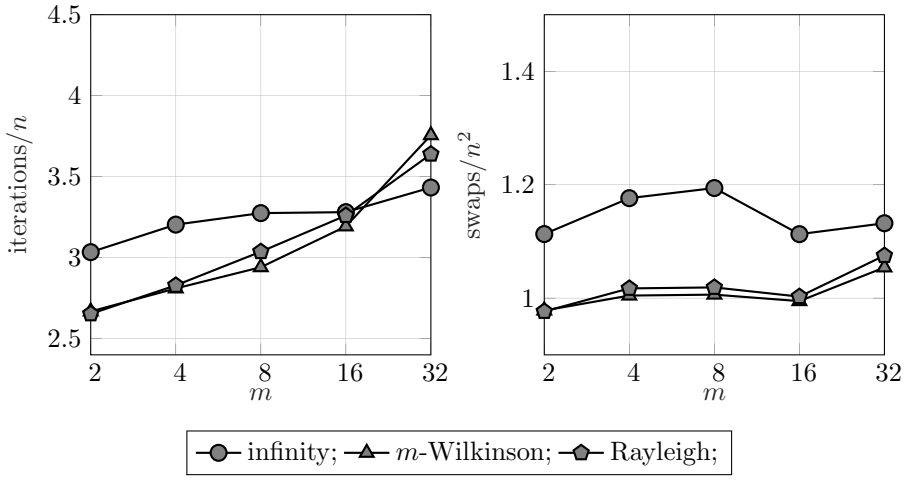


FIG. 12. *On the left the average number of iterations per eigenvalue is depicted in function of batch size m for three different pole strategies. On the right the average number of swaps scaled with n^2 is depicted in function of the batch size m. These are the results for the random problems with two clusters.*

attribute this effect to the spectrum of the randomly generated problems. All, except typically one, of the eigenvalues are located in one cluster around zero. Likely, due to the increased batch size, some of the Wilkinson and Rayleigh poles will somehow be too close to each other, thereby deteriorating the convergence.

Therefore, we have repeated this experiment with randomly generated matrix pairs of size 1000 having two equally sized clusters of eigenvalues centered around 0 and 10. The results are shown in Figure 12. Now the Wilkinson and Rayleigh poles outperform the poles at infinity in terms of total number of swaps for all batch sizes.

We conclude that we can pack the shifts tightly without a significant degradation in convergence behavior. The advantages of allowing pole selection remain but become

more problem specific. A cache efficient implementation as well as a good criterion to pick the poles is, however, future work.

5. Implicit Q theorem. In this section we prove the following implicit Q theorem for proper Hessenberg pairs justifying the implicit approach since the result of a rational QZ step is uniquely determined.

THEOREM 5.1 (implicit Q theorem for proper Hessenberg pairs). *Let (A, B) be a regular matrix pair and let $\hat{Q}$, $\check{Q}$, $\hat{Z}$, $\check{Z}$ be unitary matrices with $\hat{Q}\boldsymbol{e}_1 = \sigma\check{Q}\boldsymbol{e}_1$, $|\sigma| = 1$, such that*

$$(\hat{A}, \hat{B}) = \hat{Q}^*(A, B)\hat{Z} \quad \textit{and} \quad (\check{A}, \check{B}) = \check{Q}^*(A, B)\check{Z}$$

are both proper Hessenberg pairs having both the same pole tuple $\Xi = (\xi_1, \ldots, \xi_{n-1})$, $\xi_i \in \bar{\mathbb{C}}$, with the poles different from the spectrum of the pair.

Then the pairs $(\hat{A}, \hat{B})$ and $(\check{A}, \check{B})$ are essentially identical*, meaning that*

$$\hat{A} = D_1^*\check{A}D_2 \quad \textit{and} \quad \hat{B} = D_1^*\check{B}D_2 \tag{4}$$

with D_1 and D_2 unitary diagonal matrices.

The implicit Q theorem guarantees that the unitary equivalence transformations, which are implicitly applied in the direct reduction to a Hessenberg pair and in a rational QZ step, are essentially unique. Once the reduction or the rational QZ step is initiated, the outcome is determined.

The remainder of this section contains all ingredients to prove this theorem. Various related implicit Q theorems already exist. Vandebril, Van Barel, and Mastronardi [27] provide one for semiseparable plus diagonal matrices linked to rational Krylov spaces. Pranic, Mach, and Vandebril [15] formulate a variant for extended Hessenberg plus diagonal matrices linked to general rational Krylov subspaces as did Berljafa and Güttel [2] for (rectangular) Hessenberg pairs.

The proof we provide here significantly differs from the one by Berljafa and Güttel, who rely on direct computations and utilize the invertibility of B to formulate the theory for the single matrix setting. We make use of the properties of the associated Krylov matrices, as done by Watkins for the classical case [33]; this allows us to easily prove that the rational QZ algorithm performs nested subspace iteration driven by rational functions.

5.1. Rational Krylov matrices and subspaces. We define *rational Krylov matrices* generated by a matrix pair (A, B), a vector $\boldsymbol{v}$, and a driving rational function determined by shifts P and poles Ξ. These rational Krylov matrices span Krylov subspaces, which, for consistency, we will name *rational Krylov subspaces*. The description holds for regular matrix pairs, so the matrices do not need to be of Hessenberg form.

For the aim of a concise notation we introduce two *elementary rational matrices* generated from the pair (A, B) with shift $\varrho = \mu/\nu \in \bar{\mathbb{C}}$ and pole $\xi = \alpha/\beta \in \bar{\mathbb{C}}$:

$$\begin{aligned} M(\varrho, \xi) &= (\nu A - \mu B)(\beta A - \alpha B)^{-1}, \\ N(\varrho, \xi) &= (\beta A - \alpha B)^{-1}(\nu A - \mu B). \end{aligned} \tag{5}$$

We assume, throughout the remainder of the text, that the shift is different from the pole $\varrho \neq \xi$ and since we take inverses, the pole may not be an eigenvalue $\xi \notin \Lambda$. Notice that $M(\varrho, \xi)$ and $N(\varrho, \xi)$ represent an entire class of matrices generated by

parameters that result in the correct shift and pole. These are all scalar multiples of one another and as the theory remains scale invariant, every nontrivial representative is fine.

We remark that in the case B is invertible the following relations hold:

$$\begin{aligned} M(\varrho,\xi) &= (\nu AB^{-1} - \mu I)(\beta AB^{-1} - \alpha I)^{-1}, \\ N(\varrho,\xi) &= (\beta B^{-1}A - \alpha I)^{-1}(\nu B^{-1}A - \mu I). \end{aligned} \tag{6}$$

This could be helpful to link this analysis to existing theorems of Berljafa and Güttel [2] and Watkins [33].

The elementary rational matrices are used to define rational Krylov matrices.

DEFINITION 5.2 (rational Krylov matrices). *Let $(A,B) \in \mathbb{C}^{n\times n}$ be a regular matrix pair, $\boldsymbol{v} \in \mathbb{C}^n$ a nonzero vector, $\Xi = (\xi_1,\ldots,\xi_{k-1})$, $\xi_i \in \bar{\mathbb{C}}$, the pole tuple with the poles different from the spectrum, and $P = (\varrho_1,\ldots,\varrho_{k-1})$, $\varrho_i \in \bar{\mathbb{C}}$, the tuple of shifts distinct from the poles, with $k \le n$. The corresponding rational Krylov matrices are defined as*

$$\begin{aligned} K_k^{rat}(A,B,\boldsymbol{v},\Xi,P) &= \left[\boldsymbol{v}, M(\varrho_1,\xi_1)\boldsymbol{v}, M(\varrho_2,\xi_2)M(\varrho_1,\xi_1)\boldsymbol{v}, \ldots, \prod_{i=1}^{k-1} M(\varrho_i,\xi_i)\boldsymbol{v}\right], \\ L_k^{rat}(A,B,\boldsymbol{v},\Xi,P) &= \left[\boldsymbol{v}, N(\varrho_1,\xi_1)\boldsymbol{v}, N(\varrho_2,\xi_2)N(\varrho_1,\xi_1)\boldsymbol{v}, \ldots, \prod_{i=1}^{k-1} N(\varrho_i,\xi_i)\boldsymbol{v}\right]. \end{aligned} \tag{7}$$

The following properties of the elementary rational matrices are frequently used in the remainder of the text.

LEMMA 5.3. *The elementary rational matrices* (5) *satisfy the following:*

I. *Commutativity: For shifts $\varrho,\hat{\varrho}$ different from the poles $\xi,\hat{\xi}$,*

$$\begin{aligned} M(\varrho,\xi)\, M(\hat{\varrho},\hat{\xi}) &= M(\hat{\varrho},\hat{\xi})\, M(\varrho,\xi), \\ N(\varrho,\xi)\, N(\hat{\varrho},\hat{\xi}) &= N(\hat{\varrho},\hat{\xi})\, N(\varrho,\xi). \end{aligned} \tag{8}$$

II. *Repositioning shifts: For shifts $\varrho,\hat{\varrho}$ different from the poles $\xi,\hat{\xi}$,*

$$\begin{aligned} M(\varrho,\xi)\, M(\hat{\varrho},\hat{\xi}) &= M(\hat{\varrho},\xi)\, M(\varrho,\hat{\xi}), \\ N(\varrho,\xi)\, N(\hat{\varrho},\hat{\xi}) &= N(\hat{\varrho},\xi)\, N(\varrho,\hat{\xi}). \end{aligned} \tag{9}$$

III. *Inverse: If the shift is not an eigenvalue, $\varrho \notin \Lambda$, and different from the pole, $\varrho \neq \xi$, then*

$$\begin{aligned} M(\varrho,\xi)^{-1} &= M(\xi,\varrho), \\ N(\varrho,\xi)^{-1} &= N(\xi,\varrho). \end{aligned} \tag{10}$$

IV. *Shift invariance: For every nonzero vector $\boldsymbol{v} \in \mathbb{C}^n$, and parameters $\varrho,\hat{\varrho} \neq \xi$,*

$$\begin{aligned} \mathcal{R}(\boldsymbol{v},\ M(\varrho,\xi)\boldsymbol{v}) &= \mathcal{R}(\boldsymbol{v},\ M(\hat{\varrho},\xi)\boldsymbol{v}), \\ \mathcal{R}(\boldsymbol{v},\ N(\varrho,\xi)\boldsymbol{v}) &= \mathcal{R}(\boldsymbol{v},\ N(\hat{\varrho},\xi)\boldsymbol{v}). \end{aligned} \tag{11}$$

Proof. If B is invertible, properties I and II of the lemma follow from (6) and the property that every matrix commutes with its shifted inverse. For singular B the

same result follows from an elementary continuity argument. Property III is trivial. For property IV, we consider first the case that $\xi \neq \infty$. Assuming $\varrho \neq \infty$, it follows from (5) that

$$
\begin{aligned}
M(\varrho,\xi) &= \frac{\nu}{\beta}(I + (\xi-\varrho)B(A-\xi B)^{-1}) \equiv I + (\xi-\varrho)\, M(\infty,\xi),\\
N(\varrho,\xi) &= \frac{\nu}{\beta}(I + (\xi-\varrho)(A-\xi B)^{-1}B) \equiv I + (\xi-\varrho)\, N(\infty,\xi).
\end{aligned} \tag{12}
$$

The second part of the equation is considered as an equivalence: both sides belong to the same class of rational matrices but differ by a finite, nonzero scalar factor. It is clear that both $\mathcal{R}(\boldsymbol{v}, M(\varrho,\xi)\boldsymbol{v}) = \mathcal{R}(\boldsymbol{v}, M(\infty,\xi)\boldsymbol{v})$ and $\mathcal{R}(\boldsymbol{v}, N(\varrho,\xi)\boldsymbol{v}) = \mathcal{R}(\boldsymbol{v}, N(\infty,\xi)\boldsymbol{v})$. This equality holds trivially in the case $\varrho = \infty$. Consequently, the shift invariance property is satisfied for $\xi \neq \infty$. In the case $\xi = \infty$, assuming $\varrho \neq 0$, (5) reads

$$
M(\varrho,\infty) \equiv \varrho I - M(0,\infty),\ N(\varrho,\infty) \equiv \varrho I - N(0,\infty), \tag{13}
$$

and the shift invariance also follows for a pole at ∞ as the shift can always be moved to zero. This is trivial in the case $\varrho = 0$. ∎

THEOREM 5.4 (nested shift invariance). *For every nonzero vector $\boldsymbol{v} \in \mathbb{C}^n$, all shifts ϱ_i different from all poles ξ_j for i,j from 1 to $k-1$, and an alternative shift $\hat{\varrho}$ different from all poles, $k \leq n$,*

$$
\begin{aligned}
\mathcal{R}\left(\boldsymbol{v}, M(\varrho_1,\xi_1)\boldsymbol{v}, \ldots, \prod_{i=1}^{k-1} M(\varrho_i,\xi_i)\boldsymbol{v}\right) &= \mathcal{R}\left(\boldsymbol{v}, M(\hat{\varrho},\xi_1)\boldsymbol{v}, \ldots, \prod_{i=1}^{k-1} M(\hat{\varrho},\xi_i)\boldsymbol{v}\right),\\
\mathcal{R}\left(\boldsymbol{v}, N(\varrho_1,\xi_1)\boldsymbol{v}, \ldots, \prod_{i=1}^{k-1} N(\varrho_i,\xi_i)\boldsymbol{v}\right) &= \mathcal{R}\left(\boldsymbol{v}, N(\hat{\varrho},\xi_1)\boldsymbol{v}, \ldots, \prod_{i=1}^{k-1} N(\hat{\varrho},\xi_i)\boldsymbol{v}\right).
\end{aligned} \tag{14}
$$

Proof. We prove the first relation of (14) by induction. The proof for the second relation proceeds similarly. The case $k = 1$ is trivial, and the case $k = 2$ is equal to the shift invariance property IV of Lemma 5.3. Assume now Theorem 5.4 holds up to index k and denote this subspace as $\mathcal{U}_k$. We remark that (14) also implies that

$$
\mathcal{R}\left(\boldsymbol{v}, M(\varrho_1,\xi_1)\boldsymbol{v}, \ldots, \prod_{i=1}^{k-1} M(\varrho_i,\xi_i)\boldsymbol{v}\right) = \mathcal{R}\left(\boldsymbol{v}, M(\hat{\varrho}_1,\xi_1)\boldsymbol{v}, \ldots, \prod_{i=1}^{k-1} M(\hat{\varrho}_i,\xi_i)\boldsymbol{v}\right) \tag{15}
$$

for arbitrary shifts $\hat{\varrho}_i$ different from all poles. The subspace $\mathcal{U}_{k+1}$ is equal to

$$
\mathcal{U}_{k+1} = \mathcal{R}\left(\boldsymbol{v},\ M(\varrho_1,\xi_1)\boldsymbol{v}, \ldots, \prod_{i=1}^{k} M(\varrho_i,\xi_i)\boldsymbol{v}\right) = \mathcal{U}_k + \mathcal{R}\left(\prod_{i=1}^{k} M(\varrho_i,\xi_i)\boldsymbol{v}\right).
$$

By the induction hypothesis, the result holds for $\mathcal{U}_k$. We now modify the additional

term in the subspace $\mathcal{U}_{k+1}$ to prove the result:

$$\begin{aligned}
\mathcal{U}_{k+1} &= \mathcal{U}_k + M(\varrho_k, \xi_k)\,\mathcal{R}\left(\prod_{i=1}^{k-1} M(\varrho_i, \xi_i)\boldsymbol{v}\right) \\
&= \mathcal{U}_k + M(\hat{\varrho}, \xi_k)\,\mathcal{R}\left(\prod_{i=1}^{k-1} M(\varrho_i, \xi_i)\boldsymbol{v}\right) \\
&= \mathcal{U}_k + M(\varrho_{k-1}, \xi_k)\,\mathcal{R}\left(M(\hat{\varrho}, \xi_{k-1})\prod_{i=1}^{k-2} M(\varrho_i, \xi_i)\boldsymbol{v}\right) \\
&= \mathcal{U}_k + M(\hat{\varrho}, \xi_k)\,\mathcal{R}\left(M(\hat{\varrho}, \xi_{k-1})\prod_{i=1}^{k-2} M(\varrho_i, \xi_i)\boldsymbol{v}\right) \\
&= \mathcal{U}_k + M(\varrho_{k-2}, \xi_k)\,\mathcal{R}\left(M(\hat{\varrho}, \xi_{k-2})M(\hat{\varrho}, \xi_{k-1})\prod_{i=1}^{k-3} M(\varrho_i, \xi_i)\boldsymbol{v}\right) \\
&= \cdots \\
&= \mathcal{U}_k + \mathcal{R}\left(\prod_{i=1}^{k} M(\hat{\varrho}, \xi_i)\boldsymbol{v}\right).
\end{aligned}$$

The second equality above applies the shift invariance property IV of Lemma 5.3 to change ϱ_k to $\hat{\varrho}$. This is permitted as $\prod_{i=1}^{k-1} M(\varrho_i, \xi_i)\boldsymbol{v}$ is a vector in $\mathcal{U}_k$. In the third equality the shifts $\hat{\varrho}$ and ϱ_{k-1} are interchanged based on property II of Lemma 5.3. The fourth equality again applies the shift invariance property IV to change ϱ_{k-1} to $\hat{\varrho}$. This is again permitted: $M(\hat{\varrho}, \xi_{k-1})\prod_{i=1}^{k-2} M(\varrho_i, \xi_i)\boldsymbol{v}$ is a vector in $\mathcal{U}_k$ based on the induction hypothesis and (15). This reasoning is continued in the fifth equality, where $\hat{\varrho}$ and ϱ_{k-2} are interchanged, until eventually all shifts are changed to $\hat{\varrho}$ in the final equality. ∎

We can now define the *rational Krylov subspaces* as the column spaces of the rational Krylov matrices from Definition 5.2. It follows directly from the nested shift invariance property of Theorem 5.4 that these subspaces are independent of the choice of P.

DEFINITION 5.5 (rational Krylov subspaces). *We define the rational Krylov subspaces $\mathcal{K}_k^{rat}$ and $\mathcal{L}_k^{rat}$, $k \leq n$, associated with the regular pair $(A, B) \in \mathbb{C}^{n\times n}$, a vector $\boldsymbol{v} \in \mathbb{C}^n$, and pole tuple $\Xi = (\xi_1, \ldots, \xi_{k-1})$, assuming the poles are different from the eigenvalues as*

$$\begin{aligned}
\mathcal{K}_k^{rat}(A, B, \boldsymbol{v}, \Xi) &= \mathcal{R}(K_k^{rat}(A, B, \boldsymbol{v}, \Xi, \mathrm{P})), \\
\mathcal{L}_k^{rat}(A, B, \boldsymbol{v}, \Xi) &= \mathcal{R}(L_k^{rat}(A, B, \boldsymbol{v}, \Xi, \mathrm{P})),
\end{aligned} \tag{16}$$

where the shift tuple P is freely chosen, assuming all shifts are different from all poles.

The two rational Krylov subspaces reduce to the same subspace if B is the identity matrix which is in agreement with earlier definitions. The rational Krylov subspaces satisfy the following elementary properties.

LEMMA 5.6 (properties of rational Krylov subspaces). *The rational Krylov subspaces $\mathcal{K}^{rat}$ and $\mathcal{L}^{rat}$ generated from $(A, B) \in \mathbb{C}^{n\times n}$, $\boldsymbol{v} \in \mathbb{C}^n$, and $\Xi = (\xi_1, \ldots, \xi_{n-1})$, assuming all poles are different from the eigenvalues, satisfy the following properties.*

I. *They form a sequence of nested subspaces,*

$$\mathcal{K}_1^{rat} \subseteq \mathcal{K}_2^{rat} \subseteq \cdots \subseteq \mathcal{K}_n^{rat} \quad \text{and} \quad \mathcal{L}_1^{rat} \subseteq \mathcal{L}_2^{rat} \subseteq \cdots \subseteq \mathcal{L}_n^{rat}. \tag{17}$$

II. *For $k = 1, \ldots, n-1$, with the shift $\hat{\varrho}$ different from all eigenvalues and poles, and an alternative shift $\check{\varrho} \neq \hat{\varrho}$, we get*

(18)

$$\mathcal{K}_k^{rat}(A, B, \boldsymbol{v}, \Xi) = \prod_{i=1}^{k-1} M(\hat{\varrho}, \xi_i)\ \mathcal{K}_k(M(\check{\varrho}, \hat{\varrho}), \boldsymbol{v}) = \mathcal{K}_k\left(M(\check{\varrho}, \hat{\varrho}), \prod_{i=1}^{k-1} M(\hat{\varrho}, \xi_i)\ \boldsymbol{v}\right),$$

$$\mathcal{L}_k^{rat}(A, B, \boldsymbol{v}, \Xi) = \prod_{i=1}^{k-1} N(\hat{\varrho}, \xi_i)\ \mathcal{K}_k(N(\check{\varrho}, \hat{\varrho}), \boldsymbol{v}) = \mathcal{K}_k\left(N(\check{\varrho}, \hat{\varrho}), \prod_{i=1}^{k-1} N(\hat{\varrho}, \xi_i)\ \boldsymbol{v}\right),$$

which connects rational Krylov subspaces with regular Krylov subspaces.

III. *For $k = 1, \ldots, n-1$, and $\varrho_k \notin \Xi$,*

$$\begin{aligned} M(\varrho_k, \xi_k)\ \mathcal{K}_k^{rat}(A, B, \boldsymbol{v}, \Xi) &\subseteq \mathcal{K}_{k+1}^{rat}(A, B, \boldsymbol{v}, \Xi), \\ N(\varrho_k, \xi_k)\ \mathcal{L}_k^{rat}(A, B, \boldsymbol{v}, \Xi) &\subseteq \mathcal{L}_{k+1}^{rat}(A, B, \boldsymbol{v}, \Xi). \end{aligned} \tag{19}$$

IV. *If, for $k < n$, $\mathcal{K}_k^{rat} = \mathcal{K}_{k+1}^{rat}$ or $\mathcal{L}_k^{rat} = \mathcal{L}_{k+1}^{rat}$, the subspaces become, respectively, M- or N-invariant.*

Proof. The nestedness follows directly from the definition. To prove the second property we rely on Theorem 5.4,

$$\begin{aligned} \mathcal{K}_k^{\text{rat}}(A, B, \boldsymbol{v}, \Xi) &= \mathcal{R}\left(\boldsymbol{v}, M(\hat{\varrho}, \xi_1)\,\boldsymbol{v}, \ldots, \prod_{i=1}^{k-1} M(\hat{\varrho}, \xi_i)\,\boldsymbol{v}\right) \\ &= \prod_{i=1}^{k-1} M(\hat{\varrho}, \xi_i)\ \mathcal{R}\left(\prod_{i=1}^{k-1} M(\xi_i, \hat{\varrho})\,\boldsymbol{v}, \prod_{i=2}^{k-1} M(\xi_i, \hat{\varrho})\,\boldsymbol{v}, \ldots,\ \boldsymbol{v}\right) \\ &= \prod_{i=1}^{k-1} M(\hat{\varrho}, \xi_i)\ \mathcal{R}\left(\prod_{i=1}^{k-1} M(\check{\varrho}, \hat{\varrho})\,\boldsymbol{v}, \prod_{i=2}^{k-1} M(\check{\varrho}, \hat{\varrho})\,\boldsymbol{v}, \ldots,\ \boldsymbol{v}\right) \\ &= \prod_{i=1}^{k-1} M(\hat{\varrho}, \xi_i)\ \mathcal{K}_k(M(\check{\varrho}, \hat{\varrho}), \boldsymbol{v}). \end{aligned}$$

The first equality is the definition with $\mathrm{P} = (\hat{\varrho}, \ldots, \hat{\varrho})$. The second equality extracts the last rational term. The third equality applies the nested shift invariance property of Theorem 5.4 to change all shifts ξ_i to $\check{\varrho}$. We end up with a Krylov subspace in the last equality. The result for $\mathcal{L}^{\text{rat}}$ is proved in a similar way. The third property follows from the second property and the nestedness of Krylov subspaces, setting $\hat{\varrho} = \varrho_k$. The fourth property follows from (19) by imposing $\mathcal{K}_k^{\text{rat}} = \mathcal{K}_{k+1}^{\text{rat}}$ or $\mathcal{L}_k^{\text{rat}} = \mathcal{L}_{k+1}^{\text{rat}}$. ∎

We remark that property II states that rational Krylov subspaces are nothing other than Krylov subspaces whose starting vector is modified by a rational function determined by the poles Ξ.

5.2. Proper Hessenberg pairs and rational Krylov. In the previous section (A, B) could be any regular pair. Now we'll see that if (A, B) is a proper Hessenberg pair, the rational Krylov subspaces and matrices have a special structure.

THEOREM 5.7. *Let $(A, B) \in \mathbb{C}^{n\times n}$ be a proper Hessenberg pair having poles $\Xi = (\xi_1, \ldots, \xi_{n-1})$ distinct from the eigenvalues, Then for k from 1 to n,*

$$\mathcal{K}_k^{rat}(A, B, \boldsymbol{e}_1, (\xi_1, \ldots, \xi_{k-1})) = \mathcal{E}_k, \tag{20}$$

while for k from 1 to $n-1$,

$$\mathcal{L}_k^{rat}(A, B, \boldsymbol{e}_1, (\xi_2, \ldots, \xi_k)) = \mathcal{E}_k. \tag{21}$$

Proof. We prove the results by induction on the subspace dimension. The case $k = 1$ is trivial for both statements. To prove (20), assume the result holds up to dimension $k \leq n-1$,

$$\mathcal{K}_k^{\text{rat}}(A, B, \boldsymbol{e}_1, (\xi_1, \ldots, \xi_{k-1})) = \mathcal{E}_k.$$

From the nestedness of rational Krylov subspaces, we have by induction

$$\mathcal{E}_k \subseteq \mathcal{K}_{k+1}^{\text{rat}}(A, B, \boldsymbol{e}_1, (\xi_1, \ldots, \xi_k)).$$

It remains to be shown that $\boldsymbol{e}_{k+1} \in \mathcal{K}_{k+1}^{\text{rat}}(A, B, \boldsymbol{e}_1, (\xi_1, \ldots, \xi_k))$. From (19) and the induction hypothesis we deduce

$$M(\varrho_k, \xi_k)\ \mathcal{E}_k \subseteq \mathcal{K}_{k+1}^{\text{rat}}(A, B, \boldsymbol{e}_1, (\xi_1, \ldots, \xi_k)) \tag{22}$$

for $\varrho_k \notin \Xi$. Now consider the vector $\boldsymbol{k}_k = (\beta_k A - \alpha_k B)\boldsymbol{e}_k$ with $\alpha_k/\beta_k = \xi_k$. As $\beta_k A - \alpha_k B$ is an upper Hessenberg matrix with a zero in position $(k+1, k)$, $\boldsymbol{k}_k \in \mathcal{E}_k$. It follows that

$$\boldsymbol{k}_{k+1} = M(\varrho_k, \xi_k)\ \boldsymbol{k}_k = (\nu_k A - \mu_k B)(\beta_k A - \alpha_k B)^{-1}\ \boldsymbol{k}_k = (\nu_k A - \mu_k B)\ \boldsymbol{e}_k$$

is a vector in $\mathcal{E}_{k+1}$ with $k_{k+1} \neq 0$ and by (22), $\boldsymbol{k}_{k+1} \in \mathcal{K}_{k+1}^{\text{rat}}$. This proves the first result.

In order to prove (21), we can start in a similar way. Assume the result holds up to dimension $k < n-1$.[3] We get from the nestedness of rational Krylov subspaces and the induction hypothesis that

$$\mathcal{E}_k \subseteq \mathcal{L}_{k+1}^{\text{rat}}(A, B, \boldsymbol{e}_1, (\xi_2, \ldots, \xi_{k+1})).$$

From (19) and the induction hypothesis we deduce

$$N(\varrho_{k+1}, \xi_{k+1})\ \mathcal{E}_k \subseteq \mathcal{L}_{k+1}^{\text{rat}}(A, B, \boldsymbol{e}_1, (\xi_2, \ldots, \xi_{k+1}))$$

for $\varrho_{k+1} \notin \Xi$. To complete the proof, we need to show as before that there exists a pair of vectors $\boldsymbol{\ell}_k, \boldsymbol{\ell}_{k+1}$, with $\boldsymbol{\ell}_k \in \mathcal{E}_k$ and $\boldsymbol{\ell}_{k+1} \in \mathcal{E}_{k+1}$ whose $(k+1)$st element $\ell_{k+1} \neq 0$, that are related as

$$\boldsymbol{\ell}_{k+1} = N(\varrho_{k+1}, \xi_{k+1})\ \boldsymbol{\ell}_k = (\beta_{k+1} A - \alpha_{k+1} B)^{-1}(\nu_{k+1} A - \mu_{k+1} B)\ \boldsymbol{\ell}_k. \tag{23}$$

An explicit construction is not possible in this case. Nonetheless, by (23) we have that $(\boldsymbol{\ell}_k, \boldsymbol{\ell}_{k+1})$ must satisfy

$$(\beta_{k+1} A - \alpha_{k+1} B)\ \boldsymbol{\ell}_{k+1} = (\nu_{k+1} A - \mu_{k+1} B)\ \boldsymbol{\ell}_k.$$

[3]For $\mathcal{K}_k^{\text{rat}}$, $k+1$ can be as large as n since (20) goes up to ξ_{k-1}. For $\mathcal{L}_k^{\text{rat}}$, $k+1$ is limited to $n-1$ as we don't want to run out of poles.

From properties I and II of Lemma 2.2, we have that the matrix $\beta_{k+1}A - \alpha_{k+1}B$ is an upper Hessenberg matrix that admits a block upper triangular partition with a leading block of size $(k+1)\times(k+1)$, while the matrix $\nu_{k+1}A-\mu_{k+1}B$ is a proper upper Hessenberg matrix since the shift ϱ_{k+1} is different from all the poles. Observe that all vectors $\boldsymbol{\ell}_k \in \mathcal{E}_k$ would lead to a vector $\boldsymbol{\ell}_{k+1}$ with element $\ell_{k+1} = 0$ if and only if the first k columns of $(\beta_{k+1}A-\alpha_{k+1}B)$ would span the same subspace as the first k columns of $(\nu_{k+1}A-\mu_{k+1}B)$. It follows from properties III and IV of Lemma 2.2 that this cannot be true. We conclude that a valid pair $(\boldsymbol{\ell}_k, \boldsymbol{\ell}_{k+1})$ must exist. ∎

A direct corollary of the theorem considers the structure of rational Krylov matrices generated from proper Hessenberg pairs.

COROLLARY 5.8. *Let $(A, B) \in \mathbb{C}^{n\times n}$ be a proper Hessenberg pair with poles $\Xi = (\xi_1, \ldots, \xi_{n-1})$ different from the eigenvalues of (A, B) and let $(\varrho_1, \ldots, \varrho_{n-1})$ be a shift tuple different from the poles. Then, for k from 1 to n,*

$$K_k^{rat}(A, B, \boldsymbol{e}_1, (\xi_1, \ldots, \xi_{k-1}), (\varrho_1, \ldots, \varrho_{k-1})),$$

and, for k from 1 to $n-1$,

$$L_k^{rat}(A, B, \boldsymbol{e}_1, (\xi_2, \ldots, \xi_k), (\varrho_2, \ldots, \varrho_k)),$$

are upper triangular matrices with nonvanishing diagonal elements.

5.3. Proof of the implicit Q theorem. We are ready to prove Theorem 5.1.

Proof. Choose a tuple of $n-1$ shifts P different from the poles Ξ. Corollary 5.8 states that $K_n^{\mathrm{rat}}(\hat{A}, \hat{B}, \boldsymbol{e}_1, \Xi, \mathrm{P})$ and $K_n^{\mathrm{rat}}(\check{A}, \check{B}, \boldsymbol{e}_1, \Xi, \mathrm{P})$ are $n\times n$ nonsingular upper triangular matrices. The elementary rational matrix $M(\varrho, \xi)$ is transformed via $\hat{Q}$ and $\check{Q}$ to $\hat{M}(\varrho, \xi) = \hat{Q}^*\, M(\varrho, \xi)\, \hat{Q}$ and $\check{M}(\varrho, \xi) = \check{Q}^*\, M(\varrho, \xi)\, \check{Q}$.

It follows that

$$\begin{aligned}
&\hat{Q}\, K_n^{\mathrm{rat}}(\hat{A}, \hat{B}, \boldsymbol{e}_1, \Xi, \mathrm{P}) \\
&\quad = \hat{Q} \left[\boldsymbol{e}_1\ \hat{M}(\varrho_1, \xi_1)\, \boldsymbol{e}_1\ \ldots\ \left(\prod_{i=1}^{n-1} \hat{M}(\varrho_i, \xi_i) \right)\ \boldsymbol{e}_1 \right] \\
&\quad = \hat{Q} \left[\boldsymbol{e}_1\ \hat{Q}^* M(\varrho_1, \xi_1) \hat{Q}\, \boldsymbol{e}_1\ \ldots\ \hat{Q}^* \left(\prod_{i=1}^{n-1} M(\varrho_i, \xi_i) \right) \hat{Q}\, \boldsymbol{e}_1 \right] \\
&\quad = \left[\hat{\boldsymbol{q}}_1\ M(\varrho_1, \xi_1)\, \hat{\boldsymbol{q}}_1\ \ldots\ \left(\prod_{i=1}^{n-1} M(\varrho_i, \xi_i) \right)\ \hat{\boldsymbol{q}}_1 \right] \\
&\quad = \sigma \left[\check{\boldsymbol{q}}_1\ M(\varrho_1, \xi_1)\, \check{\boldsymbol{q}}_1\ \ldots\ \left(\prod_{i=1}^{n-1} M(\varrho_i, \xi_i) \right)\ \check{\boldsymbol{q}}_1 \right] \\
&\quad = \sigma \check{Q}\, K_n^{\mathrm{rat}}(\check{A}, \check{B}, \boldsymbol{e}_1, \Xi, \mathrm{P}).
\end{aligned}$$

Since the upper triangular matrices K_n^{rat} are nonsingular, the uniqueness of the QR factorization implies the existence of a unitary diagonal matrix D_1 such that $\hat{Q} = \check{Q}D_1$.

It remains to prove that a similar relation holds for the matrices $\hat{Z}$ and $\check{Z}$. Let us first prove that $\hat{Z}$ and $\check{Z}$ also share a first column up to unimodular scaling. From the relations $(\beta_1\hat{A} - \alpha_1\hat{B}) = \hat{Q}^*(\beta_1 A-\alpha_1 B)\hat{Z}$ and $(\beta_1\check{A} - \alpha_1\check{B}) = \check{Q}^*(\beta_1 A-\alpha_1 B)\check{Z}$

with $\xi_1 = \alpha_1/\beta_1$, it follows that

$$\begin{aligned} \hat{\boldsymbol{z}}_1 &= \hat{Z}\boldsymbol{e}_1 = (\beta_1 A - \alpha_1 B)^{-1}\hat{Q}(\beta_1 \hat{A} - \alpha_1 \hat{B})\boldsymbol{e}_1, \\ \check{\boldsymbol{z}}_1 &= \check{Z}\boldsymbol{e}_1 = (\beta_1 A - \alpha_1 B)^{-1}\check{Q}(\beta_1 \check{A} - \alpha_1 \check{B})\boldsymbol{e}_1. \end{aligned} \tag{24}$$

Since both $(\beta_1\hat{A} - \alpha_1\hat{B})\boldsymbol{e}_1$ and $(\beta_1\check{A} - \alpha_1\check{B})\boldsymbol{e}_1$ reduce to a scalar multiple of $\boldsymbol{e}_1$ and $\hat{Q}\boldsymbol{e}_1 = \sigma\check{Q}\boldsymbol{e}_1$ we get $\check{\boldsymbol{z}}_1 = \tilde{\sigma}\hat{\boldsymbol{z}}_1$. Using Corollary 5.8 and similar reasoning as before, it is shown that the following two QR factorizations are equal:

$$\hat{Z}\, L_{n-1}^{\mathrm{rat}}(\hat{A}, \hat{B}, \boldsymbol{e}_1, \Xi_s, \mathrm{P}_s) = \tilde{\sigma}\check{Z} L_{n-1}^{\mathrm{rat}}(\check{A}, \check{B}, \boldsymbol{e}_1, \Xi_s, \mathrm{P}_s)$$

with $\Xi_s = (\xi_2, \ldots, \xi_{n-1})$ and $\mathrm{P}_s = (\varrho_2, \ldots, \varrho_{n-1})$. In this case the L_{n-1} matrices are of size $n\times n-1$. Uniqueness of the QR factorization implies essential uniqueness of the first $n-1$ columns of $\hat{Z}$ and $\check{Z}$. Nonetheless also the last column of $\hat{Z}$ and $\check{Z}$ are essentially the same as they are orthogonal to the first $n-1$ columns. We conclude that $\hat{Z} = \check{Z}D_2$, with D_2 a unitary diagonal matrix. ∎

When the Hessenberg pair is not proper, uniqueness can only be guaranteed up to the pole that causes the problem. This is similar to the Hessenberg case. In practice this is in fact good news as a breakdown signals a deflation.

6. Implicit rational subspace iteration. It is well-known that Francis's QR algorithm [8, 9] effects nested subspace iteration with a change of coordinate system driven by polynomial Krylov subspaces [28, Theorem 6.3], [34, p. 396]. This result is generalized in this section for the rational QZ method.

Starting with a proper Hessenberg pair (A, B) with $\Xi = (\xi_1, \ldots, \xi_{n-1})$, a single iteration of the rational QZ method with shift ϱ and new pole $\hat{\xi}_{n-1}$ results in a new proper Hessenberg pair,

$$(\hat{A}, \hat{B}) = Q^*\,(A, B)\,Z,$$

with $\hat{\Xi} = (\xi_2, \ldots, \xi_{n-1}, \hat{\xi}_{n-1})$. This equivalence transformation simultaneously performs two similarity transformations on the matrices,

$$\hat{M}(\varrho, \xi) = Q^*\, M(\varrho, \xi)\, Q \quad \text{and} \quad \hat{N}(\varrho, \xi) = Z^*\, N(\varrho, \xi)\, Z \tag{25}$$

for all ϱ and ξ.

The following theorem formalizes the convergence behavior of the RQZ method.

THEOREM 6.1. *Consider a single RQZ step $(\hat{A}, \hat{B}) = Q^*\,(A, B)\,Z$ with shift ϱ, pole tuple $\Xi = (\xi_1, \ldots, \xi_{n-1})$ prior to the RQZ step, and $\hat{\Xi} = (\xi_2, \ldots, \xi_{n-1}, \hat{\xi}_{n-1})$ afterward. Assume all poles are different from the eigenvalues, and the shift ϱ is different from all eigenvalues and poles. For $k = 1, \ldots, n-1$, this effects subspace iteration driven by $M(\varrho, \xi_k)$ and $N(\varrho, \xi_{k+1})$, and we get*

$$\mathcal{R}(Q_{:,1:k}) = M(\varrho, \xi_k)\,\mathcal{E}_k \qquad \text{and} \qquad \mathcal{R}(Z_{:,1:k}) = N(\varrho, \xi_{k+1})\,\mathcal{E}_k \tag{26}$$

with $\xi_n \coloneqq \hat{\xi}_{n-1}$. The change of coordinate system maps both $\mathcal{R}(Q_{:,1:k})$ and $\mathcal{R}(Z_{:,1:k})$ back to $\mathcal{E}_k$.

Proof. We make use of the properties of Lemmas 5.3 and 5.6, Theorem 5.7, (25), and $\boldsymbol{q}_1 = \gamma M(\varrho, \xi_1)\, \boldsymbol{e}_1$ (equation (3)). We get

$$\begin{aligned}
\mathcal{R}(Q_{:,1:k}) &= Q\,\mathcal{E}_k = Q\,\mathcal{K}_k^{\text{rat}}(\hat{A}, \hat{B}, \boldsymbol{e}_1, \hat{\Xi}) \\
&= Q \prod_{i=2}^{k} \hat{M}(\varrho, \xi_i) \cdot \mathcal{K}_k(\hat{M}(\check{\varrho}, \varrho), \boldsymbol{e}_1) \\
&= \prod_{i=2}^{k} M(\varrho, \xi_i) \cdot \mathcal{K}_k(M(\check{\varrho}, \varrho), Q\boldsymbol{e}_1) \\
&= \prod_{i=2}^{k} M(\varrho, \xi_i) \cdot \mathcal{K}_k(M(\check{\varrho}, \varrho), M(\varrho, \xi_1)\boldsymbol{e}_1) \\
&= M(\varrho, \xi_k) \prod_{i=1}^{k-1} M(\varrho, \xi_i) \cdot \mathcal{K}_k(M(\check{\varrho}, \varrho), \boldsymbol{e}_1) \\
&= M(\varrho, \xi_k)\,\mathcal{E}_k.
\end{aligned}$$

The first equality is clear, and the second equality uses Theorem 5.7. The third equality applies part II of Lemma 5.6. The fourth equality relies on (25) to change from $\hat{M}$ to M. The fifth equality uses the expression for $\boldsymbol{q}_1$, the sixth uses the commutativity property, and the last equality again applies Lemma 5.6 and Theorem 5.7.

The second result follows a similar reasoning. The only difference is the relation between $\boldsymbol{z}_1$ and $\boldsymbol{e}_1$. Starting from the same argument as in (24) we get, for some constants $\gamma, \check{\gamma}$, and $\tilde{\gamma}$,

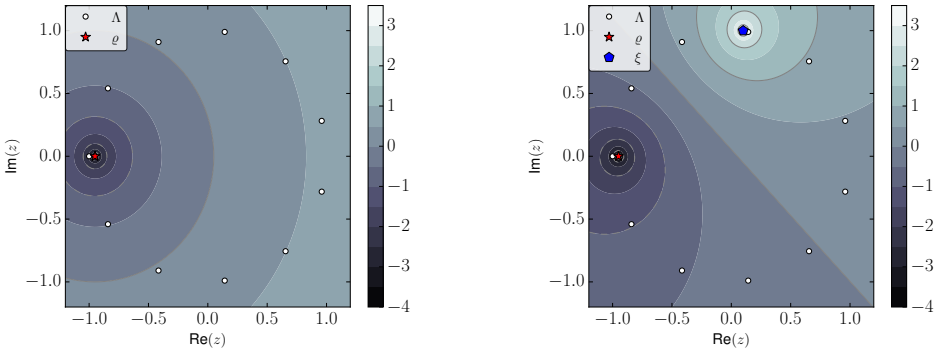
$$\boldsymbol{z}_1 = \gamma(\beta_2 A - \alpha_2 B)^{-1}\boldsymbol{q}_1 = \check{\gamma}\,(\beta_2 A - \alpha_2 B)^{-1} M(\varrho, \xi_1)\,\boldsymbol{e}_1 = \tilde{\gamma}\, N(\varrho, \xi_2)\,\boldsymbol{e}_1. \qquad \square$$

A single shifted RQZ step will execute a QR step with shift ϱ on the entire space simultaneously with RQ steps having shifts ξ_i on selected subspaces. The shift ϱ is rapidly moving from top to bottom and thus affects all subspaces. The poles on the other hand are slowly moving upward, one row during each step, and as such do not act on all subspaces in a single RQZ step. The shifts will rapidly initiate convergence at the bottom, and the poles slowly push converged eigenvalues to the top. This is an explanation of why, in the classical QZ algorithm, the zero eigenvalues in B appear at the top: they are pushed there by the poles at infinity. Moreover, it is also clear from the analysis that picking a shift equal to a pole will lead to cancellation in some of the factors thereby slowing down convergence.

Note that in the formulation of Theorem 6.1 the shift and poles are assumed to be different from the eigenvalues of the matrix pair. This is imposed to ensure that the required inverses exist. However, in practical implementations, these parameters will typically converge toward an eigenvalue. This is in fact a desirable situation as it will lead to deflations.

In the QZ algorithm [16], all poles are at ∞ and the two driving functions reduce to $M(\varrho, \infty)$ and $N(\varrho, \infty)$ which is equivalent to $AB^{-1} - \varrho I$ and $B^{-1}A - \varrho I$. In the RQZ method, the poles can be chosen freely and as such they can be utilized to influence the convergence of the method as was illustrated in the numerical experiments of subsections 3.2 and 4.3. Note that, as the poles only shift one row up during every RQZ step, it takes $n-1$ iterations before a pole has moved from the bottom to the top and has influenced all vectors in the subspace iteration.

To further clarify the result of Theorem 6.1 consider the simplified case where all the poles of the Hessenberg pair are equal to the same value ξ different from the



(a) Filter with $\xi = \infty$

(b) Filter with $\xi = 0.1 + 1i$

FIG. 13. *Logarithm of the absolute value of the rational filter, $|q(\lambda_i)^s|$, after $s = 2$ iterations with $\varrho = -0.95$, and ξ either at ∞ or at $0.1 + 1i$. The eigenvalues λ_i are shown with circles, the shift ϱ is indicated with a star, and the pole with a pentagon. Darker regions agree with convergence at the end of the pencil and lighter regions with convergence at the top of the pencil.*

eigenvalues of (A, B). Assume that the RQZ algorithm is applied s times on this proper Hessenberg pair with the same shift ϱ. At the end of each RQZ step the last pole is again restored to ξ. Then the subspace iterations, as considered from the initial pair, are given by

$$Q : \mathcal{E}_k \to M(\varrho, \xi)^s \mathcal{E}_k, \quad \text{and} \quad Z : \mathcal{E}_k \to N(\varrho, \xi)^s \mathcal{E}_k.$$

Denote $q(z) = (z - \varrho)/(z - \xi)$ and let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of the pair (A, B), so that $q(\lambda_i)^s$ is the rational filter that is implicitly applied during these s iterations to λ_i. Assume the eigenvalues are ordered such that

$$|q(\lambda_1)^s| \leq |q(\lambda_2)^s| \leq \cdots \leq |q(\lambda_{n-1})^s| \leq |q(\lambda_n)^s|,$$

and then the convergence factor of an eigenvalue at the end of the Hessenberg pencil is given by $|q(\lambda_1)^s|/|q(\lambda_2)^s|$, while the convergence factor at the top of the Hessenberg pencil is given by $|q(\lambda_{n-1})^s|/|q(\lambda_n)^s|$. As such, a good choice of both poles and shifts can accelerate convergence and lead to deflations.

As an example consider a problem of size 11 with eigenvalues located on the unit circle in the complex plane. Figure 13 shows the absolute value of the rational filter after s=2 iterations for two different choices for the rational function q. Figure 13(a) shows the filter, $q_\infty(z)^2$, with shift $\varrho = -0.95$ and all the poles at ∞. This situation corresponds to the QZ method applied twice with the same shift to a Hessenberg, triangular pair. The shift ϱ is located close to the eigenvalue $\lambda_1 = -1$ such that $|q_\infty(\lambda_1)^2| = 2.5 \cdot 10^{-3}$ is the minimal value of the filter over all eigenvalues. The convergence factor of λ_1 at the end of the pencil is approximately $8.22 \cdot 10^{-3}$. At the top of the pencil there is no convergence in this case as $|q(\lambda_{n-1})^2|/|q(\lambda_n)^2| = 1$. Figure 13(b) shows the same experiment but this time the poles are located at $\xi = 0.1+1i$ which is in the vicinity of another eigenvalue. This situation corresponds to the RQZ method applied twice with the same shift to a Hessenberg pair with $\Xi = (\xi, \ldots, \xi)$. The rational filter, $q_\xi(z)^2$, leads to a convergence factor of λ_1 at the end of the pencil of approximately $1.21 \cdot 10^{-2}$. The convergence of λ_1 at the end of the

pencil is slower with q_ξ^2 compared to q_∞^2. However, q_ξ^2 will also lead to convergence at the top of the pencil as the convergence factor is $|q(\lambda_{n-1})^2|/|q(\lambda_n)^2| \approx 7.46 \cdot 10^{-3}$. We observe that using q_ξ leads to convergence of another eigenvalue, whereas q_∞ does not.

It is clear that both the shifts and the poles can accelerate the convergence but they do influence each other.

When the shifts are changed in every iteration and the poles of the Hessenberg pair are not the same then the filter q becomes dependent on the index k and will be a product of terms with different shifts,

$$q_k(\lambda) = \prod_{i=1}^{s} (\lambda - \varrho_i)/(\lambda - \xi_k^{(i)}) \tag{27}$$

with $\xi_k^{(i)}$ the pole at iteration i in position k (or $k+1$) for Q (or Z) as shown in Theorem 6.1.

Provided a good choice of shifts and poles is made during repeated application of the RQZ algorithm the pair (A, B) will converge to a pair of upper triangular matrices.

7. Filtering rational Krylov subspaces. In the last part, we will apply the concept of the RQZ method within the rational Krylov method [18, 19, 20, 21] to filter and restart the rational Krylov method. The rational Krylov method is an iterative method applicable for, for example, computing a select subset of eigenvalues of large-scale eigenvalue problems. Our formulation is in terms of a large-scale complex-valued matrix pair (A, B) of dimension $N \times N$.

Starting from a regular pair (A, B), a nonzero vector $\boldsymbol{v} \in \mathbb{C}^N$, and a tuple of poles $\Xi = (\xi_1, \ldots, \xi_k)$ different from the spectrum, the rational Krylov method iteratively constructs an orthonormal basis $V_{k+1} \in \mathbb{C}^{N \times k+1}$ of the rational Krylov subspace $\mathcal{K}_{k+1}^{\text{rat}}(A, B, \boldsymbol{v}, \Xi)$.

It also constructs a $(k+1) \times k$ recurrence matrix pair $(\underline{H}_k, \underline{G}_k)$ in Hessenberg form. This rational Krylov Hessenberg pair contains the poles that are used in the rational Krylov method as its subdiagonal elements: $\xi_i = h_{i+1,i}/g_{i+1,i}$ for i from 1 to k. This is similar to the square Hessenberg pairs that are used in the first part of this paper. As long as the rational Krylov method does not break down, the pair $(\underline{H}_k, \underline{G}_k)$ can be considered as proper according to two out of three conditions of Definition 2.1. The third condition concerning the linear independence of the last row of the pair does not hold for the rectangular rational Krylov Hessenberg pencil. If the other two conditions are satisfied, we nonetheless say that $(\underline{H}_k, \underline{G}_k)$ is a proper rational Krylov Hessenberg pair.

The rational Krylov recurrence relation,

$$A\, V_{k+1}\, \underline{G}_k = B\, V_{k+1}\, \underline{H}_k, \tag{28}$$

holds throughout the rational Krylov method. We refer the interested reader to [21] for further algorithmic details on the iterative scheme.

The rational Krylov method generalizes Arnoldi's method [1], which generates an orthonormal basis for a standard Krylov subspace $\mathcal{K}_{k+1}(A, \boldsymbol{v})$. These Krylov methods have a growing orthogonalization cost and growing memory requirements with increasing subspace dimension. To overcome this one could apply implicit filtering and restart.

Sorensen [23] applied Francis's QR algorithm to filter a standard Krylov subspace and implicitly restart the Arnoldi iteration. The implicit QR algorithm can be applied

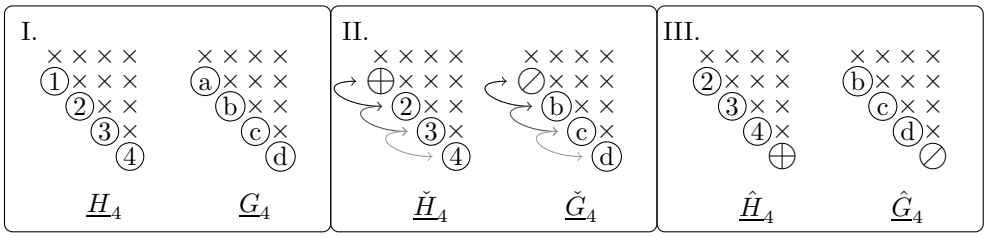


FIG. 14. *RQZ-like procedure to change the first pole in an rational Krylov Hessenberg pair to a new shift (pane* II*) and move it to the last position in the rational Krylov Hessenberg pair (pane* II–III*).*

to restart Arnoldi's method because they are both based on the Hessenberg matrix structure. As a generalization, the RQZ algorithm can be applied to filter a rational Krylov subspace and restart the rational Krylov iteration because both methods use the structure of Hessenberg pairs. Berljafa and Güttel [2, section 4.3] already proposed this technique of changing and swapping the poles in the rational Krylov method as a way to implicitly filter a rational Krylov subspace. The first algorithm to apply an implicit filter in the rational Krylov method is due to De Samblanx, Meerbergen, and Bultheel [5]. However, their method relied on an explicit QZ algorithm which is quite costly and prone to numerical inaccuracies.

To filter a rational Krylov subspace we can thus use the concept of the RQZ method. The procedure is summarized in Figure 14. The initial situation of the rational Krylov Hessenberg is shown in pane I on the left. In pane II, the first pole ξ_1 is changed to a shift ϱ by computing a unitary transformation Q such that

$$\boldsymbol{q}_1 = \check{\gamma}(\underline{H}_k - \varrho\underline{G}_k)(\underline{H}_k - \xi_1\,\underline{G}_k)^\dagger \boldsymbol{e}_1 = \hat{\gamma}(\underline{H}_k - \varrho\underline{G}_k)\boldsymbol{e}_1. \tag{29}$$

The principle is the same as described in subsection 2.2; the only difference is that the inverse is replaced with the Moore-Penrose *pseudoinverse* $(\underline{H}_k - \xi_1\,\underline{G}_k)^\dagger$.

It is well-known [10] that $\boldsymbol{x}_{LS} = (\underline{H}_k - \xi_1\,\underline{G}_k)^\dagger \boldsymbol{b}$ is the least squares solution of minimal norm $\|\boldsymbol{x}\|_2$. As $\|\gamma\boldsymbol{e}_1 - (\underline{H}_k - \xi_1\underline{G}_k)\boldsymbol{e}_1\|_2 = 0$ when $\gamma = h_{11} - \xi_1 g_{11}$, we conclude that

$$(\underline{H}_k - \xi_1\,\underline{G}_k)^\dagger \boldsymbol{e}_1 = \gamma\boldsymbol{e}_1. \tag{30}$$

Pane II of Figure 14 further shows how the shift is swapped to the last position on the subdiagonal of $(\underline{H}_k, \underline{G}_k)$. The end result is displayed in pane III. The process summarized in Figure 14 effectively updates,

$$(\hat{\underline{H}}_k, \hat{\underline{G}}_k) = Q^*(\underline{H}_k, \underline{G}_k)Z,$$

in such a way that the pole tuple is changed to $\hat{\Xi} = (\xi_2, \ldots, \xi_k, \varrho)$. To maintain the rational Krylov recurrence (28), the orthonormal basis is updated as $\hat{V}_{k+1} = V_{k+1}Q$.

This does not change the span of V_{k+1}, i.e., $\mathcal{R}(\hat{V}_{k+1}) = \mathcal{R}(V_{k+1})$, but the vectors are *rearranged*. The new start vector is given by

$$\hat{\boldsymbol{v}} = \hat{V}_{k+1}\,\boldsymbol{e}_1 = V_{k+1}\,\boldsymbol{q}_1 = \gamma V_{k+1}(\underline{H}_k - \varrho\underline{G}_k)\boldsymbol{e}_1. \tag{31}$$

The rational Krylov recurrence (28) implies

$$(A - \varrho B)\,V_{k+1}\,(\underline{H}_k - \xi_1\underline{G}_k) = (A - \xi_1 B)\,V_{k+1}\,(\underline{H}_k - \varrho\underline{G}_k). \tag{32}$$

TABLE 2

Summary of the settings and results of the restarted rational Krylov iteration. The columns list the maximal subspace dimension m, the restart length p, the number of wanted Ritz values ℓ, the tolerance `tol`*, and the required number of restarts to reach convergence.*

Problem	m	p	ℓ	`tol`	# restarts
Cavity flow	40	20	8	10^{-7}	8
Obstacle flow	60	25	7	10^{-7}	11

Rearranging terms in (32) and combining this with (30), (31) we see that the new starting vector is given by

$$\hat{\boldsymbol{v}} = \gamma(A-\xi_1 B)^{-1}(A-\varrho B)\,\boldsymbol{v}. \tag{33}$$

From the uniqueness of a rational Krylov recurrence (28) [2, Theorem 3.2], it follows that $\mathcal{R}(\hat{V}_{k+1}) = \mathcal{K}^{\text{rat}}_{k+1}(A, B, \hat{\boldsymbol{v}}, \hat{\Xi})$.

The filter operation is finalized by removing the last pole ϱ from the subspace by reducing the order of the rational Krylov recurrence by one. This means that the trailing column and row of $(\underline{\hat{H}}_k, \underline{\hat{G}}_k)$ are removed, as well as the last vector of $\hat{V}_{k+1}$.

With the results of sections 5 and 6 in mind it is clear how this RQZ-like procedure applies a filter in the rational Krylov iteration.

As a numerical experiment, we revisit the two fluid flow problems of subsection 4.3. Instead of computing all eigenvalues we are now only interested in determining if the problems are stable. To this end, it is sufficient to determine the rightmost eigenvalues of both problems and check if they are situated in the left half-plane. As an input to the restarted rational Krylov method we have the matrix pair (A, B), a start vector $\boldsymbol{v}$, a tuple of poles Ξ, a maximal subspace dimension m, a restart length p, a number of desired Ritz values ℓ, and a tolerance `tol` up to which the ℓ Ritz values need to be converged. The residual is determined as $\|(\underline{H}_k - \theta\underline{G}_k)\,\boldsymbol{y}\|_2$ with $(\theta, \boldsymbol{y})$ the Ritz value and Ritz vector computed from (H_k, G_k), the upper $k\times k$ block of the recurrence pencil. The iteration starts with computing an initial rational Krylov subspace of dimension m. If the ℓ rightmost Ritz values have converged up to a maximal residual `tol`, the iteration is halted. Otherwise the p leftmost Ritz values are selected as shifts, the subspace is reduced to dimension $m-p$ by using the RQZ method to filter the subspace, and the subspace is again expanded to full dimension m. This procedure is repeated until the ℓ rightmost Ritz values have converged.

The settings and results are summarized in Table 2. In both cases, we selected poles along the imaginary axis, $\Xi = (-20i, -18i, \ldots, 18i, 20i)$, as we expect the rightmost eigenvalue to be situated close to it.

Figure 15 shows the rightmost part of the spectrum and the converged Ritz values. As can be seen, the method successfully converged to the correct eigenvalues within a reasonable number of restarts.

8. Conclusion. In this paper we proposed a rational QZ algorithm for the numerical solution of the dense (unsymmetric) generalized eigenvalue problem. The new algorithm operates on matrix pairs in Hessenberg, Hessenberg form rather than the Hessenberg, triangular form used in the classical QZ method. Hessenberg pairs link to rational Krylov and the associated poles are encoded in the subdiagonal elements of both Hessenberg matrices. A direct reduction method of a regular matrix pair to Hessenberg, Hessenberg form was proposed. Moreover, we have demonstrated that during the reduction a good choice of poles can lead to premature deflations. The iterative rational QZ algorithm differs from the classical algorithm in the sense that also

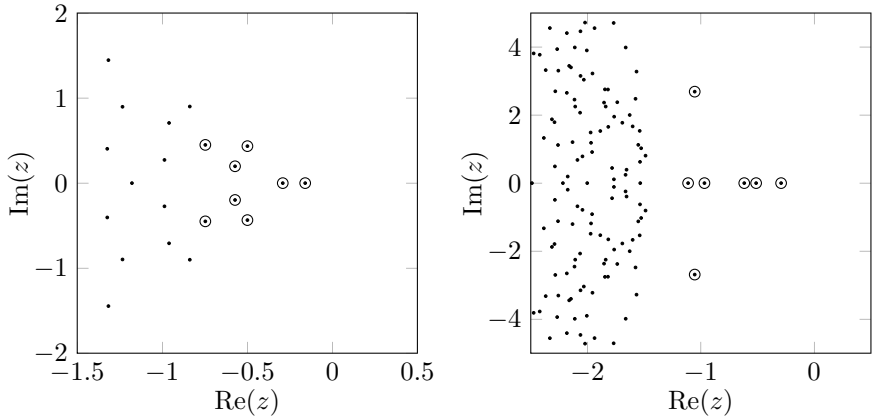


FIG. 15. *Rightmost part of the spectrum of the cavity flow (left) and obstacle flow (right) problems. The eigenvalues (•) and Ritz values (○) are shown.*

poles can be introduced in each QZ step. Numerical experiments confirm that a good choice of poles allows the RQZ method to outperform the QZ algorithm by reducing the number of iterations per eigenvalue. The implicit chasing technique is justified by an implicit Q theorem, which is proved in a novel manner operating directly on the matrix pair and exploiting the connections with rational Krylov. Our theoretical analysis revealed that an RQZ iteration implicitly performs nested subspace iteration driven by a pair of rational functions. Finally, we have applied the RQZ method as a filter in rational Krylov.

Acknowledgments. The authors thank Jared Aurentz and Thomas Mach for the fruitful discussions and suggestions related to this project, and the referees for their valuable feedback.

REFERENCES

[1] W. E. Arnoldi, *The principle of minimized iteration in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.
[2] M. Berljafa and S. Güttel, *Generalized rational Krylov decompositions with an application to rational approximation*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 894–916.
[3] R. F. Boisvert, R. Pozo, K. Remington, R. F. Barrett, and J. J. Dongarra, *Matrix market: A web resource for test matrix collections*, in Proceedings of the IFIP TC2/WG2.5 Working Conference on Quality of Numerical Software: Assessment and Enhancement, Chapman & Hall, London, UK, 1997, pp. 125–137.
[4] G. De Samblanx and A. Bultheel, *Using implicitly filtered RKS for generalised eigenvalue problems*, J. Comput. Appl. Math., 107 (1999), pp. 195–218.
[5] G. De Samblanx, K. Meerbergen, and A. Bultheel, *The implicit application of a rational filter in the RKS method*, BIT, 37 (1997), pp. 925–947.
[6] H. Elman, A. Ramage, and D. Silvester, *Algorithm* 866: *IFISS, a Matlab toolbox for modelling incompressible flow*, ACM Trans. Math. Software, 33 (2007), pp. 2–14.
[7] H. Elman, A. Ramage, and D. Silvester, *IFISS: A computational laboratory for investigating incompressible flow problems*, SIAM Rev., 56 (2014), pp. 261–273.
[8] J. G. F. Francis, *The QR transformation a unitary analogue to the LR transformation—Part* 1, Comput. J., 4 (1961), pp. 265–271.
[9] J. G. F. Francis, *The QR transformation—Part* 2, Comput. J., 4 (1962), pp. 332–345.
[10] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.
[11] B. Kågström and D. Kressner, *Multishift variants of the QZ algorithm with aggressive early deflation*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 199–227.

[12] B. Kågström and P. Poromaa, *Computing eigenspaces with specified eigenvalues of a regular matrix pair (A, B) and condition estimation: Theory, algorithms and software*, Numer. Algorithms, 12 (1996), pp. 369–407.
[13] L. Kaufman, *Some thoughts on the QZ algorithm for solving the generalized eigenvalue problem*, ACM Trans. Math. Software, 3 (1977), pp. 65–75.
[14] D. Kressner, *Numerical Methods for General and Structured Eigenvalue Problems*, Lect. Notes Comput. Sci. Eng. 46, Springer, New York, 2005.
[15] T. Mach, M. S. Pranic, and R. Vandebril, *Computing approximate (block) rational Krylov subspaces without explicit inversion with extensions to symmetric matrices*, Electron. Trans. Numer. Anal., 43 (2014), pp. 100–124.
[16] C. B. Moler and G. W. Stewart, *An algorithm for generalized matrix eigenvalue problems*, SIAM J. Numer. Anal., 10 (1973), pp. 1–52.
[17] E. Polizzi, *Density-matrix-based algorithm for solving eigenvalue problems*, Phys. Rev. B, 79 (2009), 115112.
[18] A. Ruhe, *Rational Krylov sequence methods for eigenvalue computation*, Linear Algebra Appl., 58 (1984), pp. 391–405.
[19] A. Ruhe, *Rational Krylov algorithms for nonsymmetric eigenvalue problems*, in Recent Advances in Iterative Methods, IMA Vol. Math. Appl. 60, Springer, New York, 1994, pp. 149–164.
[20] A. Ruhe, *Rational Krylov algorithms for nonsymmetric eigenvalue problems,* II: *Matrix pairs*, Linear Algebra Appl., 197–198 (1994), pp. 283–296.
[21] A. Ruhe, *Rational Krylov: A practical algorithm for large sparse nonsymmetric matrix pencils*, SIAM J. Sci. Comput., 19 (1998), pp. 1535–1551.
[22] T. Sakurai and H. Sugiura, *A projection method for generalized eigenvalue problems using numerical integration*, J. Comput. Appl. Math., 159 (2003), pp. 119–128.
[23] D. C. Sorensen, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 357–385.
[24] M. Van Barel and P. Kravanja, *Nonlinear eigenvalue problems and contour integrals*, J. Comput. Appl. Math., 292 (2016), pp. 526–540.
[25] R. Van Beeumen, K. Meerbergen, and W. Michiels, *Subspace iteration for generalized eigenvalue problems via contour integration and rational Krylov methods*, submitted, 2017.
[26] P. Van Dooren, *A generalized eigenvalue approach for solving Riccati equations*, SIAM J. Sci. and Stat. Comput., 2 (1981), pp. 121–135.
[27] R. Vandebril, M. Van Barel, and N. Mastronardi, *Matrix Computations and Semiseparable Matrices, Volume* II: *Eigenvalue and Singular Value Methods*, Johns Hopkins University Press, Baltimore, MD, 2008.
[28] R. Vandebril and D. S. Watkins, *A generalization of the multishift QR algorithm*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 759–779.
[29] R. Vandebril and D. S. Watkins, *An extension of the QZ algorithm beyond the Hessenberg-upper triangular pencil*, Electron. Trans. Numer. Anal., 40 (2013), pp. 17–35.
[30] R. C. Ward, *The combination shift QZ algorithm*, SIAM J. Numer. Anal., 12 (1975), pp. 835–853.
[31] D. S. Watkins, *Bulge exchanges in algorithms of QR-type*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 1074–1096.
[32] D. S. Watkins, *Performance of the QZ Algorithm in the Presence of Infinite Eigenvalues*, SIAM J. Matrix Anal. Appl., 22 (2000), pp. 364–375.
[33] D. S. Watkins, *The Matrix Eigenvalue Problem: GR and Krylov Subspace Methods*, SIAM, Philadelphia, 2007.
[34] D. S. Watkins, Amer. Math. Monthly, (2011), pp. 387–403.
[35] D. S. Watkins and L. Elsner, *Theory of decomposition and bulge-chasing algorithms for the generalized eigenvalue problem*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 943–967.