Generalized rational Krylov decompositions with an application to rational approximation

Berljafa, Mario and Güttel, Stefan

2014

MIMS EPrint: **2014.59**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: http://eprints.maths.manchester.ac.uk/
And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

GENERALIZED RATIONAL KRYLOV DECOMPOSITIONS WITH AN APPLICATION TO RATIONAL APPROXIMATION

MARIO BERLJAFA* AND STEFAN GÜTTEL*

Abstract. Generalized rational Krylov decompositions are matrix relations which, under certain conditions, are associated with rational Krylov spaces. We study the algebraic properties of such decompositions and present an implicit Q theorem for rational Krylov spaces. Transformations on rational Krylov decompositions allow for changing the poles of a rational Krylov space without recomputation, and two algorithms are presented for this task. Using such transformations we develop a rational Krylov method for rational least squares fitting. Numerical experiments indicate that the proposed method converges fast and robustly. A MATLAB toolbox with implementations of the presented algorithms and experiments is provided.

Key words. rational Krylov decomposition, inverse eigenvalue problem, rational approximation

AMS subject classifications. 15A22, 65F15, 65F18, 30E10

1. Introduction. Numerical methods based on rational Krylov spaces have become an indispensable tool of scientific computing. Rational Krylov spaces were initially proposed by Ruhe in the 1980s for the purpose of solving large sparse eigenvalue problems [37, 39, 40]. Since then many more applications have been found in model order reduction [22, 17], large-scale matrix functions and matrix equations [13, 15, 1, 26, 27], and nonlinear eigenvalue problems [41, 30, 47, 28], to name a few.

In this paper we study various algebraic properties of rational Krylov spaces, using as starting point a *generalized rational Krylov decomposition*

$$AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}, \tag{1.1}$$

where $A \in \mathbb{C}^{N\times N}$ is a given matrix, and the matrices $V_{m+1} \in \mathbb{C}^{N\times(m+1)}$ and $\{\underline{K_m}, \underline{H_m}\} \subset \mathbb{C}^{(m+1)\times m}$ are of maximal rank. Throughout this paper the underlined matrices have one more row than they have columns.

The rational Arnoldi algorithm by Ruhe [39, 40] naturally generates decompositions of the form (1.1) in which case it is known (by construction) that the columns of V_{m+1} are an (orthonormal) basis of a rational Krylov space. Different choices of the so-called continuation vectors in the rational Arnoldi algorithm give rise to different decompositions, but all of them correspond to the same rational Krylov space. In this work we answer the converse question of when a decomposition (1.1) is associated with a rational Krylov space, and how transformations of such a decomposition affect the parameters of the rational Krylov space.

Our approach is inspired by the work of Stewart [43, 45] who studied transformations on a *(polynomial) Krylov decomposition*

$$AV_m = V_{m+1}\underline{H_m}, \tag{1.2}$$

which is a special case of (1.1) with $\underline{K_m} = \underline{I_m}$, the $m \times m$ identity matrix with an appended row of zeros. Indeed, all results in this paper apply to polynomial Krylov spaces as well.

* School of Mathematics, The University of Manchester, Alan Turing Building, Oxford Road, M13 9PL Manchester, United Kingdom, mario.berljafa@manchester.ac.uk, stefan.guettel@manchester.ac.uk

The outline of this work is as follows: in section 2 we study algebraic properties of rational Arnoldi decompositions (a special case of (1.1) where $(\underline{H_m}, \underline{K_m})$ is an unreduced upper-Hessenberg pencil) and relate these decompositions to the poles and the starting vector of a rational Krylov space. Section 3 provides a rational implicit Q theorem about the uniqueness of rational Arnoldi decompositions. We also show how the rational functions associated with the rational Krylov space can be evaluated at any point $z \in \mathbb{C}$ by computing a full QR factorization of $z\underline{K_m} - \underline{H_m}$. In section 4 we show that when the lower $m \times m$ part of the pencil $(\underline{H_m}, \underline{K_m})$ is regular then the range of V_{m+1} in (1.1) is a rational Krylov space. Via transformations on such decompositions we are able to move the poles of a rational Krylov space to arbitrary positions (even to the eigenvalues of A), and we give two algorithms for this task. Finally, in section 5 we incorporate one of these algorithms into an iterative method, called RKFIT. Given matrices $\{A, F\} \subset \mathbb{C}^{N\times N}$ and a vector $\boldsymbol{v} \in \mathbb{C}^N$, RKFIT attempts to find a rational function R_m of type (m, m) such that $R_m(A)\boldsymbol{v} \approx F\boldsymbol{v}$ in the Euclidian norm.

All algorithms and numerical experiments presented in this paper are contained in a MATLAB toolbox [2] available for download.[1]

Notation. Matrices are labeled with uppercase Latin letters and their elements with the corresponding lowercase letter, e.g., $A = \big[a_{ij}\big]$, $\underline{H_m} = \big[h_{ij}\big]$. Vectors are labeled with lowercase letters in bold, e.g., $\boldsymbol{b}$ or $\boldsymbol{v}_k$. Hence, we also use $V_{m+1} = \big[\boldsymbol{v}_1 \quad \dots \quad \boldsymbol{v}_{m+1}\big]$ to partition V_{m+1} in columns. Depending on the context, $\boldsymbol{h}_k$ may represent just the leading k rows of the kth column of $\underline{H_m}$, whilst $\boldsymbol{h_k}$ the leading $k+1$ rows. The kth canonical vector is denoted by $\boldsymbol{e}_k$. By $\overline{\mathcal{R}}(V)$ we denote the range of a matrix V. The linear space of polynomials of degree at most m is denoted by $\mathcal{P}_m$. Finally, $\overline{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$ and $\overline{\mathbb{C}}^* := \overline{\mathbb{C}} \setminus \{0\}$.

2. Rational Arnoldi decompositions. Given a matrix $A \in \mathbb{C}^{N\times N}$, a starting vector $\boldsymbol{v} \in \mathbb{C}^N$, and an integer $m < N$, the associated *polynomial Krylov space of order* $m+1$ is defined as $\mathcal{K}_{m+1} = \mathcal{K}_{m+1}(A, \boldsymbol{v}) = \operatorname{span}\{\boldsymbol{v}, A\boldsymbol{v}, \dots, A^m\boldsymbol{v}\}$. There exists an integer $M \le N$, called the *invariance index* for $(A, \boldsymbol{v})$, such that

$$\mathcal{K}_1 \subset \mathcal{K}_2 \subset \cdots \subset \mathcal{K}_{M-1} \subset \mathcal{K}_M = \mathcal{K}_{M+1}.$$

Throughout this work we assume that $0 < m < M$, in which case the space $\mathcal{K}_{m+1}$ is isomorphic to $\mathcal{P}_m$, i.e., any $\boldsymbol{w} \in \mathcal{K}_{m+1}$ corresponds to a polynomial $p \in \mathcal{P}_m$ satisfying $\boldsymbol{w} = p(A)\boldsymbol{v}$, and vice versa.

Given a nonzero polynomial $q_m \in \mathcal{P}_m$ with roots disjoint from the spectrum $\Lambda(A)$, we define the associated *rational Krylov space* as

$$\mathcal{Q}_{m+1} = \mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m) := q_m(A)^{-1}\mathcal{K}_{m+1}(A, \boldsymbol{v}). \tag{2.1}$$

Note that $q_m(A)$ is nonsingular since no root of q_m is an eigenvalue of A and therefore $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$ is well defined. Further, for the starting vector we have $\boldsymbol{v} \in \mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$. The assumption that $q_m(A)$ is nonsingular and the relation (2.1) also imply that the spaces $\mathcal{Q}_{m+1}$ and $\mathcal{K}_{m+1}(A, \boldsymbol{v})$ are of the same dimension for all m. Therefore $\mathcal{Q}_{m+1}$ is A-variant (i.e., $A\mathcal{Q}_{m+1} \not\subseteq \mathcal{Q}_{m+1}$) if and only if $m+1 < M$.

The roots of q_m are called *poles* of the rational Krylov space and denoted by $\xi_1, \xi_2, \dots, \xi_m$. If $\deg(q_m) < m$ then $m - \deg(q_m)$ of the poles are set to ∞. In this

[1] `http://guettel.com/rktoolbox` as of November 2014.

case we refer to ∞ as a formal (multiple) root of q_m. We now show that the poles of a rational Krylov space are uniquely determined by the starting vector and vice versa.

LEMMA 2.1. *Let $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$ be a given A-variant rational Krylov space, i.e., $m+1 < M$. Then the poles of $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$ are uniquely determined by the starting vector $\boldsymbol{v}$, or equivalently, the starting vector of $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$ is uniquely, up to scaling, determined by the roots of q_m.*

Proof. We first show that given an A-variant polynomial Krylov space $\mathcal{K}_{m+1}(A, \boldsymbol{q})$, all vectors $\boldsymbol{w} \in \mathcal{K}_{m+1}(A, \boldsymbol{q})$ that satisfy $\mathcal{K}_{m+1}(A, \boldsymbol{q}) = \mathcal{K}_{m+1}(A, \boldsymbol{w})$ are of the form $\boldsymbol{w} = \alpha \boldsymbol{q}$, $\alpha \neq 0$. Assume to the contrary that there exists a polynomial p_j with $1 \leq \deg(p_j) = j \leq m$ such that $\boldsymbol{w} = p_j(A)\boldsymbol{q}$. Then $A^{m+1-j}\boldsymbol{w} \in \mathcal{K}_{m+1}(A, \boldsymbol{w})$, but for the same vector we have $A^{m+1-j}\boldsymbol{w} = A^{m+1-j}p_j(A)\boldsymbol{q} \notin \mathcal{K}_{m+1}(A, \boldsymbol{q})$. This is a contradiction to $\mathcal{K}_{m+1}(A, \boldsymbol{q}) = \mathcal{K}_{m+1}(A, \boldsymbol{w})$.

To show that the poles are uniquely determined by the starting vector $\boldsymbol{v}$, assume that $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m) = \mathcal{Q}_{m+1}(A, \boldsymbol{v}, \widehat{q}_m)$. Using the definition of a rational Krylov space (2.1), this is equivalent to $\mathcal{K}_{m+1}(A, q_m(A)^{-1}\boldsymbol{v}) = \mathcal{K}_{m+1}(A, \widehat{q}_m(A)^{-1}\boldsymbol{v})$, which in turn is equivalent to $\mathcal{K}_{m+1}(A, \widehat{q}_m(A)\boldsymbol{v}) = \mathcal{K}_{m+1}(A, q_m(A)\boldsymbol{v})$. This space is A-variant, hence by the above argument we know that $q_m(A)\boldsymbol{v} = \alpha\widehat{q}_m(A)\boldsymbol{v}$, $\alpha \neq 0$. This vector is an element of $\mathcal{K}_{m+1}(A, \boldsymbol{v})$ which is isomorphic to $\mathcal{P}_m$. Therefore $q_m = \alpha\widehat{q}_m$ and hence q_m and $\widehat{q}_m$ have identical roots. Similarly one shows that if $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m) = \mathcal{Q}_{m+1}(A, \widehat{\boldsymbol{v}}, q_m)$, then $\boldsymbol{v} = \alpha\widehat{\boldsymbol{v}}$ with $\alpha \neq 0$. ∎

In the following we aim to establish a one-to-one correspondence between rational Krylov spaces and a particular type of matrix decompositions. As a consequence we are able to study the algebraic properties of rational Krylov spaces using these decompositions. Recall that a matrix $\underline{H_m} \in \mathbb{C}^{(m+1)\times m}$ is called *upper-Hessenberg* if all the elements below the first subdiagonal are zero, i.e., if $i > j+1$ implies $h_{ij} = 0$. Further, we say that $\underline{H_m}$ is *unreduced* if none of the elements on the first subdiagonal are zero, i.e., $h_{j+1,j} \neq 0$. For convenience, we now generalize this terminology from matrices to pencils $\left(\underline{H_m}, \underline{K_m}\right)$.

DEFINITION 2.2. *Let $\{\underline{K_m}, \underline{H_m}\} \subset \mathbb{C}^{(m+1)\times m}$ be upper-Hessenberg matrices. We say that $\left(\underline{H_m}, \underline{K_m}\right)$ is an* unreduced upper-Hessenberg pencil *if $|h_{j+1,j}| + |k_{j+1,j}| \neq 0$ for all $j = 1, \ldots, m$.*

We are now ready to introduce the notion of a rational Arnoldi decomposition, which is a generalization of decompositions generated by Ruhe's rational Arnoldi algorithm [39, 40]; see also [9, Definition 2.2] and [26, Definition 5.5].

DEFINITION 2.3. *Let $A \in \mathbb{C}^{N\times N}$ be a given matrix. A relation of the form*

$$AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m} \tag{2.2}$$

is called a rational Arnoldi decomposition (RAD) *if $V_{m+1} \in \mathbb{C}^{N\times(m+1)}$ is of full column rank, $\left(\underline{H_m}, \underline{K_m}\right)$ is an unreduced upper-Hessenberg pencil of size $(m+1)\times m$, and the quotients $h_{j+1,j}/k_{j+1,j}$, called* poles *of the decomposition, are outside the spectrum $\Lambda(A)$ for $j = 1, \ldots, m$.*

The columns of V_{m+1} are called the basis *of the decomposition and they span the* space *of the decomposition. If V_{m+1} is orthonormal, we say that* (2.2) *is an* orthonormal RAD.

It is noteworthy that both $\underline{H_m}$ and $\underline{K_m}$ in the RAD (2.2) are of full rank. To see this take any $\xi \in \mathbb{C}$ and subtract $\xi V_{m+1}\underline{K_m}$ from both sides of (2.2). This leads to

$$(A - \xi I)V_{m+1}\underline{K_m} = V_{m+1}(\underline{H_m} - \xi\underline{K_m}). \tag{2.3}$$

Since $\left(\underline{H_m}, \underline{K_m}\right)$ is unreduced there are at most m numbers ξ such that $\underline{H_m} - \xi \underline{K_m}$ is not unreduced. For any other ξ the right-hand side in (2.3) is of full rank and so must be the left-hand side. Therefore $\underline{K_m}$ is of full rank. If A is nonsingular, comparing the ranks of the left- and right-hand side in (2.2) we now see that $\underline{H_m}$ is of full rank as well. If A is singular, then zero is not an allowed pole and therefore $\underline{H_m}$ is unreduced and hence of full rank.

Furthermore, any RAD (2.2) can be transformed into an orthonormal RAD using the thin QR factorization $V_{m+1} = QR$. Setting $\widehat{V}_{m+1} = Q$, $\underline{\widehat{K}_m} = R\underline{K_m}$, and $\underline{\widehat{H}_m} = R\underline{H_m}$, we obtain the decomposition $A\widehat{V}_{m+1}\underline{\widehat{K}_m} = \widehat{V}_{m+1}\underline{\widehat{H}_m}$, satisfying $\mathcal{R}(\widehat{V}_{m+1}) = \mathcal{R}(V_{m+1})$, and $h_{j+1,j}/k_{j+1,j} = \widehat{h}_{j+1,j}/\widehat{k}_{j+1,j}$ for all $j = 1, \ldots, m$. We call these two RADs equivalent.

DEFINITION 2.4. *Two RADs with the same matrix $A \in \mathbb{C}^{N\times N}$ are* equivalent *if they span the same space and have the same poles.*

Note that we do not impose equal ordering of the poles for two RADs to be equivalent. From now on we assume all RADs to be orthonormal. In Theorem 2.5 we show that for every rational Krylov space $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$ there exists an RAD (2.2) spanning $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$ and conversely, if such a decomposition exists it spans a rational Krylov space. To proceed it is convenient to write the polynomial q_m in factored form, and to label separately all the leading factors

$$q_0(z) = 1, \quad \text{and} \quad q_j(z) = \prod_{i=1}^{j} \left(h_{i+1,i} - k_{i+1,i} z\right), \quad j = 1, \ldots, m, \tag{2.4}$$

with some scalars $\{h_{i+1,i}, k_{i+1,i}\}_{i=1}^{m} \subset \mathbb{C}$ such that $\xi_i = h_{i+1,i}/k_{i+1,i}$. Since (2.1) is independent of the scaling of q_m any choice of the scalars $h_{i+1,i}$ and $k_{i+1,i}$ is valid as long as their ratio is ξ_i. When we make use of (2.4) without specifying the order of appearance of the poles, we mean any order. The fact that $q_j \,|\, q_{j+1}$ gives rise to a sequence of nested rational Krylov spaces

$$\mathcal{Q}_1 \subset \mathcal{Q}_2 \subset \cdots \subset \mathcal{Q}_{m+1},$$

where $\mathcal{Q}_{j+1} = \mathcal{Q}_{j+1}(A, \boldsymbol{v}, q_j)$ for $j = 0, 1, \ldots, m$.

THEOREM 2.5. *Let $\mathcal{V}_{m+1}$ be a vector space of dimension $m+1$. Then $\mathcal{V}_{m+1}$ is a rational Krylov space with starting vector $\boldsymbol{v} \in \mathcal{V}_{m+1}$ and poles $\xi_1, \ldots, \xi_m$ if and only if there exists an RAD* (2.2) *with $\mathcal{R}(V_{m+1}) = \mathcal{V}_{m+1}$, $\boldsymbol{v}_1 = \boldsymbol{v}$, and poles $\xi_1, \ldots, \xi_m$.*

Proof. Let (2.2) hold and define the polynomials $\{q_j\}_{j=0}^{m}$ as in (2.4). Note that these are nonzero polynomials since the pencil $\left(\underline{H_m}, \underline{K_m}\right)$ is unreduced. We show by induction that

$$\mathcal{V}_{j+1} := \operatorname{span}\left\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{j+1}\right\} = q_j(A)^{-1}\mathcal{K}_{j+1}(A, \boldsymbol{v}), \tag{2.5}$$

for $j = 1, \ldots, m$, and with $\boldsymbol{v} = \boldsymbol{v}_1$. In particular, for $j = m$ we obtain $\mathcal{V}_{m+1} = q_m(A)^{-1}\mathcal{K}_{m+1}(A, \boldsymbol{v})$. Consider $j = 1$. Reading (2.2) column-wise, first column only, and rearranging the terms yields

$$q_1(A)\boldsymbol{v}_2 = (h_{21}I - k_{21}A)\,\boldsymbol{v}_2 = (k_{11}A - h_{11}I)\,\boldsymbol{v}_1. \tag{2.6}$$

Therefore, $\boldsymbol{v}_2 = q_1(A)^{-1}\,(k_{11}A - h_{11}I)\,\boldsymbol{v}_1 \in q_1(A)^{-1}\mathcal{K}_2(A, \boldsymbol{v})$ which together with the fact $\boldsymbol{v}_1 \in q_1(A)^{-1}\mathcal{K}_2(A, \boldsymbol{v})$ proves (2.5) for $j = 1$. Let us assume that (2.5) holds for

$j = 1, \ldots, n-1 < m$. We now consider the case $j = n$. Comparing the nth column on the left- and the right-hand side in (2.2) and rearranging the terms gives

$$\left(h_{n+1,n} I - k_{n+1,n} A\right) \boldsymbol{v}_{n+1} = \sum_{i=1}^{n} \left(k_{in} A - h_{in} I\right) \boldsymbol{v}_i, \tag{2.7}$$

$$\text{and hence} \qquad q_n(A)\boldsymbol{v}_{n+1} = \sum_{i=1}^{n} \left(k_{in} A - h_{in} I\right) q_{n-1}(A)\boldsymbol{v}_i. \tag{2.8}$$

By the induction hypothesis $\boldsymbol{v}_i \in q_{n-1}(A)^{-1}\mathcal{K}_n(A, \boldsymbol{v})$, therefore

$$\left(k_{in} A - h_{in} I\right) q_{n-1}(A)\boldsymbol{v}_i \in \mathcal{K}_{n+1}(A, \boldsymbol{v}), \quad i = 1, \ldots, n. \tag{2.9}$$

It follows from (2.8) and (2.9) that $\boldsymbol{v}_{n+1} \in q_n(A)^{-1}\mathcal{K}_{n+1}(A, \boldsymbol{v})$. The induction hypothesis asserts $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\} \subseteq q_n(A)^{-1}\mathcal{K}_{n+1}(A, \boldsymbol{v})$ which concludes this direction.

Alternatively, let $\mathcal{V}_{m+1} = q_m(A)^{-1}\mathcal{K}_{m+1}(A, \boldsymbol{v})$ be a rational Krylov space with a basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m+1}\}$ satisfying (2.5). Thus for $n = 1, \ldots, m$ there holds

$$\boldsymbol{v}_{n+1} \in q_n(A)^{-1}\mathcal{K}_{n+1}(A, \boldsymbol{v}) \Leftrightarrow \left(h_{n+1,n} I - k_{n+1,n} A\right) \boldsymbol{v}_{n+1} \in q_{n-1}(A)^{-1}\mathcal{K}_{n+1}(A, \boldsymbol{v}).$$

Since $\mathcal{K}_{n+1}(A, \boldsymbol{v}) = \mathcal{K}_n(A, \boldsymbol{v}) + A\mathcal{K}_n(A, \boldsymbol{v})$ we have $q_{n-1}(A)^{-1}\mathcal{K}_{n+1}(A, \boldsymbol{v}) = \mathcal{Q}_n + A\mathcal{Q}_n$. Consequently, there exist numbers $\{h_{in}, k_{in}\}_{i=1}^{n} \subset \mathbb{C}$ such that (2.7) holds. These relations can be merged into matrix form to get (2.2) with the pencil $\left(\underline{H_m}, \underline{K_m}\right)$ being unreduced as a consequence of q_m being a nonzero polynomial. ∎

Elaborating further on the proof of the previous theorem, we retrieve an explicit formula for the vectors $\boldsymbol{v}_j$, given in Theorem 2.6 below. This result appears to some extent in [37, 39], correct up to a normalization factor and given without proof. We stress that the result holds irrespectively of the RAD (2.2) being orthonormal or not.

THEOREM 2.6. *Let the RAD* (2.2) *be given. Then*

$$\boldsymbol{v}_{j+1} = p_j(A)q_j(A)^{-1}\boldsymbol{v}_1, \quad j = 0, 1, \ldots, m, \tag{2.10}$$

where $p_0(z) = 1$ *and* $p_j(z) = \det\left(zK_j - H_j\right)$*, for* $j = 1, \ldots, m$*. The polynomials* q_j *are given by* (2.4)*.*

Proof. The proof goes by induction on j. For $j = 0$ the relation (2.10) holds, and from (2.6) it follows for $j = 1$.

Assume (2.10) has been established for $j = 1, \ldots, n < m$ and insert it into (2.8), giving rise to

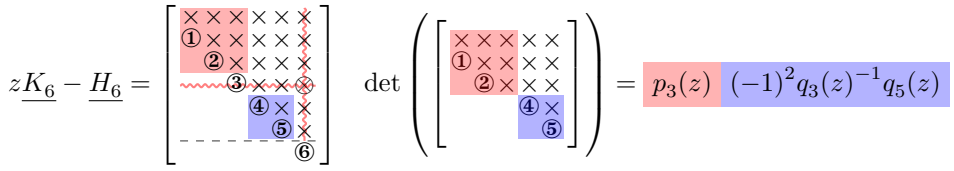
$$q_n(A)\boldsymbol{v}_{n+1} = \sum_{i=1}^{n} \left(k_{in} A - h_{in} I\right) q_{n-1}(A)p_{i-1}(A)q_{i-1}(A)^{-1}\boldsymbol{v}_1. \tag{2.11}$$

We obtain (2.10) for $j = n+1$ by noticing that the right-hand side of (2.11) represents the Laplace expansion of $\det\left(zK_n - H_n\right)$ along the last column. Indeed

$$q_n(A)\boldsymbol{v}_{n+1} = \sum_{i=1}^{n} (-1)^{i+n} \left(k_{in} A - h_{in} I\right) p_{i-1}(A)(-1)^{n-i} q_{i-1}(A)^{-1} q_{n-1}(A)\boldsymbol{v}_1.$$

See also Figure 2.1 for an illustration. ∎

Note that $p_j(z)$ is the determinant of the upper $j \times j$ submatrix of $z\underline{K_j} - \underline{H_j}$, whilst $(-1)^j q_j(z)$ is the determinant of its lower $j \times j$ submatrix.



(a) Upper-Hessenberg structure. (b) Contribution from a minor.

Fig. 2.1: Sketch illustrating the proof of Theorem 2.6. Part (a) shows the upper-Hessenberg structure of the shifted pencil $z\underline{K_j} - \underline{H_j}$, for $j = 6$. The elements marked with numbers, like $① = zk_{21} - h_{21}$, are those carrying the poles. The contribution of the element $\otimes = zk_{46} - h_{46}$ in the Laplace expansion of the determinant $\det(zK_6 - H_6)$ along the last column of the matrix is $(-1)^{4+6}(zk_{46} - h_{46})\det(M_\otimes)$. Here, $M_\otimes$ is the minor of $zK_6 - H_6$ resulting from the removal of the 4th row and last column, and is shown in part (b).

3. A rational implicit Q theorem. A special case of an orthonormal rational Arnoldi decomposition (2.2) is the polynomial Arnoldi decomposition (1.2). The corresponding polynomial q_m is constant and $\mathcal{R}(V_{m+1}) = \mathcal{K}_{m+1}(A, \boldsymbol{v}_1)$. The implicit Q theorem, see [44, Theorem 3.3], states that once the first column of V_{m+1} is fixed, so is, up to column scaling, the whole matrix V_{m+1}. Since $\underline{H_m} = V_{m+1}^* A V_m$ any scaling of V_{m+1} also affects $\underline{H_m}$. If V_{m+1} is rescaled to $\widehat{V}_{m+1} = V_{m+1} D_{m+1}$, with $|D_{m+1}| = I_{m+1}$, then $\underline{\widehat{H}_m} = D_{m+1}^* \underline{H_m} D_m$. There is no essential difference between V_{m+1} and $\underline{H_m}$ on one side and $\widehat{V}_{m+1}$ and $\underline{\widehat{H}_m}$ on the other. In this sense we say that V_{m+1} and $\underline{H_m}$ are essentially uniquely determined by the first column of V_{m+1}. With Theorem 3.2 below we now generalize this result to RADs.

Apart from the column scaling of V_{m+1}, in the rational case the decomposition (2.2) is also invariant (in the sense that it spans the same space, the poles remain unchanged, and the upper-Hessenberg structure is preserved) under right-multiplication by upper-triangular nonsingular matrices T_m. We make this precise.

DEFINITION 3.1. *Two orthonormal RADs, namely, $AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}$ and $A\widehat{V}_{m+1}\underline{\widehat{K}_m} = \widehat{V}_{m+1}\underline{\widehat{H}_m}$, are called* essentially equal *if there exist a unitary diagonal matrix D_{m+1} of size $m+1$ and an upper-triangular nonsingular matrix T_m of size m, such that $\widehat{V}_{m+1} = V_{m+1} D_{m+1}$, $\underline{\widehat{H}_m} = D_{m+1}^* \underline{H_m} T_m$, and $\underline{\widehat{K}_m} = D_{m+1}^* \underline{K_m} T_m$.*

Essentially equal orthonormal RADs form an equivalence class and we call any of its elements essentially unique.

Note that two orthonormal RADs may be equivalent but not essentially equal, as the poles may be ordered differently. We are now ready to generalize the implicit Q theorem to the rational case.

THEOREM 3.2. *Let $A \in \mathbb{C}^{N\times N}$ satisfy an orthonormal rational Arnoldi decomposition $AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}$ with poles $\xi_j = h_{j+1,j}/k_{j+1,j}$. For every $j = 1, \ldots, m$ the orthonormal matrix V_{j+1} and the pencil $(\underline{H_j}, \underline{K_j})$ are essentially uniquely determined by the first column of V_{m+1} and the poles $\xi_1, \ldots, \xi_j$.*

Proof. Let $A\widehat{V}_{m+1}\underline{\widehat{K}_m} = \widehat{V}_{m+1}\underline{\widehat{H}_m}$ be an orthonormal RAD with $\widehat{V}_{m+1}\boldsymbol{e}_1 = V_{m+1}\boldsymbol{e}_1$ and $\widehat{h}_{j+1,j}/\widehat{k}_{j+1,j} = h_{j+1,j}/k_{j+1,j}$ for all $j = 1, \ldots, m$. We show by induction that $A\widehat{V}_{m+1}\underline{\widehat{K}_m} = \widehat{V}_{m+1}\underline{\widehat{H}_m}$ is essentially equal to $AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}$.

We assume without loss of generality that $h_{j+1,j} \neq 0$ for all $j = 1, \ldots, m$.

Otherwise, if $h_{j+1,j} = 0$ for some j, then $0 = \xi_j \notin \Lambda(A)$ and we can consider $V_{m+1}\underline{K_m} = A^{-1}V_{m+1}\underline{H_m}$ at that step j, thus interchanging the roles of $\underline{H_m}$ and $\underline{K_m}$ and using A^{-1} instead of A. Since $\big(\underline{H_m}, \underline{K_m}\big)$ is unreduced, $k_{j+1,j} \neq 0$ if $h_{j+1,j} = 0$. The relation $AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}$ can be shifted for all $\xi \in \overline{\mathbb{C}}^* \setminus \Lambda(A)$ to provide

$$A^{(\xi)}V_{m+1}\underline{L_m^{(\xi)}} = V_{m+1}\underline{H_m}, \tag{3.1}$$

where $A^{(\xi)} := (I - A/\xi)^{-1}A$ and $\underline{L_m^{(\xi)}} := \big(\underline{K_m} - \underline{H_m}/\xi\big)$. We make frequent use of this relation, reading it column-wise. It is worth noticing that the jth column of $\underline{L_m^{(\xi_j)}}$ has all but eventually the leading j components equal to zero, and that $\underline{L_j^{(\xi)}}$ is of full rank for all j and ξ. Analogous results hold for $A\widehat{V}_{m+1}\underline{\widehat{K}_m} = \widehat{V}_{m+1}\underline{\widehat{H}_m}$. We are now ready to prove the statement.

Define $d_1 := 1$, so that $\widehat{\boldsymbol{v}}_1 = d_1\boldsymbol{v}_1$. The first column in (3.1) for $\xi = \xi_1$ yields

$$\ell_{11}^{(\xi_1)}A^{(\xi_1)}\boldsymbol{v}_1 = h_{11}\boldsymbol{v}_1 + h_{21}\boldsymbol{v}_2.$$

Since $\boldsymbol{v}_1^*\boldsymbol{v}_1 = 1$ and $\boldsymbol{v}_1^*\boldsymbol{v}_2 = 0$, we have

$$h_{11} = \ell_{11}^{(\xi_1)}\boldsymbol{v}_1^*A^{(\xi_1)}\boldsymbol{v}_1.$$

We then have

$$\begin{aligned} h_{21}\boldsymbol{v}_2 &= \ell_{11}^{(\xi_1)}A^{(\xi_1)}\boldsymbol{v}_1 - h_{11}\boldsymbol{v}_1, \\ \boldsymbol{v}_2 &= \ell_{11}^{(\xi_1)}\big[A^{(\xi_1)}\boldsymbol{v}_1 - \big(\boldsymbol{v}_1^*A^{(\xi_1)}\boldsymbol{v}_1\big)\boldsymbol{v}_1\big]/h_{21}. \end{aligned}$$

Since $\|\boldsymbol{v}_2\|_2 = 1$ and $h_{21} \neq 0$ by assumption, we have $\ell_{11}^{(\xi_1)} \neq 0$. Analogously

$$\widehat{h}_{11} = \widehat{\ell}_{11}^{(\xi_1)}\boldsymbol{v}_1^*A^{(\xi_1)}\boldsymbol{v}_1, \quad \widehat{\boldsymbol{v}}_2 = \widehat{\ell}_{11}^{(\xi_1)}\big[A^{(\xi_1)}\boldsymbol{v}_1 - \big(\boldsymbol{v}_1^*A^{(\xi_1)}\boldsymbol{v}_1\big)\boldsymbol{v}_1\big]/\widehat{h}_{21}, \text{ and } \widehat{\ell}_{11}^{(\xi_1)} \neq 0.$$

Obviously, $\boldsymbol{v}_2$ and $\widehat{\boldsymbol{v}}_2$ are collinear and since they are both of unit 2-norm, there exists a unimodular scalar $d_2 \in \mathbb{C}$ such that $\widehat{\boldsymbol{v}}_2 = d_2\boldsymbol{v}$. Defining $t_1 := \widehat{\ell}_1^{(\xi_1)}/\ell_1^{(\xi_1)}$, and $D_2 := \operatorname{diag}(d_1, d_2)$, and making use of $A^{(\xi_1)}\boldsymbol{v}_1 - \big(\boldsymbol{v}_1^*A^{(\xi_1)}\boldsymbol{v}_1\big)\boldsymbol{v}_1 = \widehat{h}_{21}\widehat{\boldsymbol{v}}_2/\widehat{\ell}_{11}^{(\xi_1)} = h_{21}\boldsymbol{v}_2/\ell_{11}^{(\xi_1)}$, we obtain $\underline{\widehat{H}_1} = D_2^*\underline{H_1}T_1$. From $\underline{K_1} = \underline{L_1^{(\xi_1)}} + \underline{H_1}/\xi_1$ and $\underline{\widehat{K}_1} = \underline{\widehat{L}_1^{(\xi_1)}} + \underline{\widehat{H}_1}/\xi_1$ we see that indeed $\underline{\widehat{K}_1} = D_2^*\underline{K_1}T_1$. This proves the statement for $j = 1$.

Suppose that, for $2 \leq j \leq m$, we have $\widehat{V}_j = V_jD_j$, $\underline{\widehat{H}_{j-1}} = D_j^*\underline{H_{j-1}}T_{j-1}$, and $\underline{\widehat{K}_{j-1}} = D_j^*\underline{K_{j-1}}T_{j-1}$, for a diagonal unitary matrix $D_j = \operatorname{diag}(d_1, \ldots, d_j)$ and upper-triangular nonsingular matrix T_{j-1}.

The jth column in (3.1) for $\xi = \xi_j$ gives

$$A^{(\xi_j)}V_j\boldsymbol{l}_j^{(\xi_j)} = V_{j+1}\underline{\boldsymbol{h}_j}.$$

Since $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j+1}$ are orthonormal we have

$$\boldsymbol{h}_j = V_j^*A^{(\xi_j)}V_j\boldsymbol{l}_j^{(\xi_j)}.$$

Rearranging the two equations above we deduce

$$\begin{aligned} h_{j+1,j}\boldsymbol{v}_{j+1} &= A^{(\xi_j)}V_j\boldsymbol{l}_j^{(\xi_j)} - V_j\boldsymbol{h}_j \\ &= A^{(\xi_j)}V_j\boldsymbol{l}_j^{(\xi_j)} - V_jV_j^*A^{(\xi_j)}V_j\boldsymbol{l}_j^{(\xi_j)} \\ &= \big(I - V_jV_j^*\big)A^{(\xi_j)}V_j\boldsymbol{l}_j^{(\xi_j)}. \end{aligned}$$

Expanding $\boldsymbol{l}_j^{(\xi_j)}$ as $\boldsymbol{l}_j^{(\xi_j)} =: \underline{L_{j-1}^{(\xi_j)}}\boldsymbol{z}_{j-1} + \boldsymbol{q}_j$, where $\boldsymbol{q}_j^* \underline{L_{j-1}^{(\xi_j)}} = \boldsymbol{0}^*$, gives

$$\begin{aligned} h_{j+1,j}\boldsymbol{v}_{j+1} &= \left(I - V_jV_j^*\right) A^{(\xi_j)}V_j \left(\underline{L_{j-1}^{(\xi_j)}}\boldsymbol{z}_{j-1} + \boldsymbol{q}_j\right) \\ &= \left(I - V_jV_j^*\right) A^{(\xi_j)}V_j\underline{L_{j-1}^{(\xi_j)}}\boldsymbol{z}_{j-1} + \left(I - V_jV_j^*\right) A^{(\xi_j)}V_j\boldsymbol{q}_j \\ &= \left(I - V_jV_j^*\right) A^{(\xi_j)}V_j\boldsymbol{q}_j. \end{aligned} \tag{3.2}$$

To obtain the last equality we have used $A^{(\xi_j)}V_j\underline{L_{j-1}^{(\xi_j)}} = V_j\underline{H_{j-1}}$, which are the first $j-1$ columns in (3.1) with $\xi = \xi_j$. Note that since $h_{j+1,j} \neq 0$ the vector $\boldsymbol{q}_j$ is also nonzero. We label analogously $\widehat{\boldsymbol{l}}_j^{(\xi_j)} =: \underline{\widehat{L}_{j-1}^{(\xi_j)}}\widehat{\boldsymbol{z}}_{j-1} + \widehat{\boldsymbol{q}}_j$, where $\widehat{\boldsymbol{q}}_j^* \underline{\widehat{L}_{j-1}^{(\xi_j)}} = \boldsymbol{0}^*$, and obtain

$$\begin{aligned} \widehat{h}_{j+1,j}\widehat{\boldsymbol{v}}_{j+1} &= \left(I - \widehat{V}_j\widehat{V}_j^*\right)A^{(\xi_j)}\widehat{V}_j\widehat{\boldsymbol{q}}_j, & \widehat{\boldsymbol{q}}_j^*\underline{\widehat{L}_{j-1}^{(\xi_j)}} &= \boldsymbol{0}^*, \\ \widehat{h}_{j+1,j}\widehat{\boldsymbol{v}}_{j+1} &= \left(I - V_jV_j^*\right)A^{(\xi_j)}V_jD_j\widehat{\boldsymbol{q}}_j, & \widehat{\boldsymbol{q}}_j^*D_j^*\underline{L_{j-1}^{(\xi_j)}} &= \boldsymbol{0}^*, \end{aligned}$$

where in the last equality above we have applied post-multiplication by T_{j-1}^{-1}. Since $\underline{L_{j-1}^{(\xi_j)}} \in \mathbb{C}^{j\times(j-1)}$ is of full column rank, $\boldsymbol{q}_j$ and $D_j\widehat{\boldsymbol{q}}_j$ are collinear, i.e., there exists a nonzero scalar $0 \neq \gamma \in \mathbb{C}$ such that $D_j\widehat{\boldsymbol{q}}_j = \gamma\boldsymbol{q}_j$. As a consequence $\boldsymbol{v}_{j+1}$ and $\widehat{\boldsymbol{v}}_{j+1}$ are collinear as well. Furthermore, as $\|\boldsymbol{v}_{j+1}\|_2 = \|\widehat{\boldsymbol{v}}_{j+1}\|_2 = 1$, there exists a unimodular scalar $d_{j+1} \in \mathbb{C}$ such that $\widehat{\boldsymbol{v}}_{j+1} = d_{j+1}\boldsymbol{v}_{j+1}$. We also observe $\widehat{h}_{j+1,j} = d_{j+1}^*\gamma h_{j+1,j}$.

It remains to find such $\boldsymbol{t}_j \in \mathbb{C}^j$ that $T_j = \begin{bmatrix}\underline{T_{j-1}} & \boldsymbol{t}_j\end{bmatrix}$ is nonsingular and that additionally $\underline{\widehat{H}_j} = D_{j+1}^*\underline{H_j}T_j$ and $\underline{\widehat{K}_j} = D_{j+1}^*\underline{K_j}T_j$. From $D_j\widehat{\boldsymbol{q}}_j = \gamma\boldsymbol{q}_j$ we infer

$$\begin{aligned} D_j\left(\widehat{\boldsymbol{l}}_j^{(\xi_j)} - \underline{\widehat{L}_{j-1}^{(\xi_j)}}\widehat{\boldsymbol{z}}_{j-1}\right) &= \gamma\left(\boldsymbol{l}_j^{(\xi_j)} - \underline{L_{j-1}^{(\xi_j)}}\boldsymbol{z}_{j-1}\right), \\ \widehat{\boldsymbol{l}}_j^{(\xi_j)} &= D_j^*\underline{L_{j-1}^{(\xi_j)}}\left(T_{j-1}\widehat{\boldsymbol{z}}_{j-1} - \gamma\boldsymbol{z}_{j-1}\right) + \gamma D_j^*\boldsymbol{l}_j^{(\xi_j)}, \\ \widehat{\boldsymbol{l}}_j^{(\xi_j)} &= D_j^*L_j^{(\xi_j)}\boldsymbol{t}_j, \end{aligned}$$

where $\boldsymbol{t}_j = \begin{bmatrix} T_{j-1}\widehat{\boldsymbol{z}}_{j-1} - \gamma\boldsymbol{z}_{j-1} \\ \gamma \end{bmatrix}$. Finally, using the equation above, the relation $\widehat{\boldsymbol{h}}_j = \widehat{V}_j^*A^{(\xi_j)}\widehat{V}_j\widehat{\boldsymbol{l}}_j^{(\xi_j)}$, and again $A^{(\xi_j)}V_j\underline{L_{j-1}^{(\xi_j)}} = V_j\underline{H_{j-1}}$, we derive $\widehat{\boldsymbol{h}}_j = D_j^*H_j\boldsymbol{t}_j$. With $\widehat{h}_{j+1,j} = d_{j+1}^*\gamma h_{j+1,j}$ we get $\underline{\widehat{H}_j} = D_{j+1}^*\underline{H_j}T_j$. We can consider $\underline{\widehat{K}_j}$ similarly. ∎

A further comment for the case $m = N-1$ is required. For the polynomial case, i.e., $\underline{K_{N-1}} = \underline{I_{N-1}}$, we have $AV_{N-1} = V_N\underline{H_{N-1}}$. The vector $\boldsymbol{h}_N = V_N^*AV_N\boldsymbol{e}_N$ is uniquely defined by the starting vector and A and $AV_N = V_NH_N$ holds. This last decomposition is usually stated as the (polynomial) implicit Q theorem and essential uniqueness of H_N is claimed. Let us consider a more general RAD, namely, $AV_N\underline{K_{N-1}} = V_N\underline{H_{N-1}}$. Defining $\boldsymbol{h}_N := V_N^*AV_N\boldsymbol{k}_N$ for an arbitrary $\boldsymbol{k}_N \in \mathbb{C}^N$ we see that $AV_NK_N = V_NH_N$. Therefore we cannot say that $\left(H_N, K_N\right)$ is essentially unique. In fact, essential uniqueness is related to both V_{m+1} and the pencil $\left(\underline{H_m}, \underline{K_m}\right)$ concurrently.

As already mentioned, a polynomial Krylov space $\mathcal{K}_{m+1}(A, \boldsymbol{v})$ with orthonormal basis V_{m+1} is related to a decomposition of the form

$$AV_m = V_{m+1}\underline{H_m} = V_mH_m + h_{m+1,m}\boldsymbol{v}_{m+1}\boldsymbol{e}_m^T, \tag{3.3}$$

where H_m is upper-Hessenberg. For a rational Krylov space we have an RAD (2.2) with an upper-Hessenberg pencil $\left(\underline{H_m}, \underline{K_m}\right)$ rather than a single upper-Hessenberg matrix $\underline{H_m}$. It has been shown, for example in [16, 46, 48, 33], that decompositions of the form (3.3) with H_m being *semiseparable plus diagonal*[2] are related to rational Krylov spaces in the same way as RADs are. Corresponding implicit Q theorems have been developed. The semiseparable structure can be used to develop (short) recurrences related to rational Krylov spaces, see for instance [29, 36].

We prefer to work with the pencil $\left(\underline{H_m}, \underline{K_m}\right)$ instead of the semiseparable plus diagonal representation since the former is widely used in practice. In fact, this pencil is a by-product of the rational Krylov method used to construct rational Krylov bases, and which is stated in Algorithm 3.1. We use the notation

$$A^{(\xi)} = \begin{cases} (I - A/\xi)^{-1} A, & \text{if } \xi \in \overline{\mathbb{C}}^* \setminus \Lambda(A), \\ A^{-1}, & \text{if } \xi = 0 \text{ and } A \text{ is nonsingular,} \end{cases}$$

and correspondingly

$$\underline{L_m^{(\xi)}} = \begin{cases} \underline{K_m} - \underline{H_m}/\xi, & \text{if } \xi \in \overline{\mathbb{C}}^*, \\ \underline{H_m}, & \text{otherwise.} \end{cases}$$

We merely remark that the application of $A^{(\xi)}$ in line 3 of Alg. 3.1 corresponds to the solution of a (large sparse) linear system (if $\xi \neq \infty$) or matrix-vector product (if $\xi = \infty$). Typically this is the computationally most expensive step per iteration.

The vector $\boldsymbol{q}_j$ in Alg. 3.1 is often called *continuation combination* as it specifies onto which linear combination of the previously computed basis vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j$ the operator $A^{(\xi)}$ is applied to. The choice made in line 7 is due to Ruhe [40] and it guarantees that the new vector $\boldsymbol{w}$ will be linearly independent of the previous vectors and hence expand the space, provided that we have not reached the invariance index. This can be argued as follows.

From Theorem 2.5 we know that for any rational Krylov space (of full dimension) there exists an orthonormal RAD spanning it. Equation (3.2) shows that the continuation combination $\boldsymbol{q}_j$, with the operator transformed to $A^{(\xi_j)}$, indeed expands the space. Moreover, a continuation combination must have a nonzero direction in $\boldsymbol{q}_j$ in order to expand the space. Otherwise, if the continuation combination is of the form $\underline{L_{j-1}^{(\xi_j)}} \boldsymbol{z}_{j-1}$, then (3.1) with m replaced by $j-1$ implies $\boldsymbol{w} := A^{(\xi_j)} V_j \underline{L_{j-1}^{(\xi_j)}} \boldsymbol{z}_{j-1} = V_j \underline{H_{j-1}} \boldsymbol{z}_{j-1} \in \mathcal{Q}_j(A, \boldsymbol{v}, q_{j-1})$ and the space is not expanded.

In (2.10) we have given explicit formulas for the orthonormal vectors $\boldsymbol{v}_{j+1}$, implicitly defined by the starting vector $\boldsymbol{v}_1$ and the poles $\xi_1, \ldots, \xi_j$. Note that the determinants appearing in the formulas do not change if the pencil $\left(\underline{H_j}, \underline{K_j}\right)$ is right-multiplied by an upper-triangular nonsingular matrix. We now give a formula for (a multiple of) the continuation vector $\boldsymbol{q}_{j+1} = \boldsymbol{q}_{j+1}^{(\xi_{j+1})}$ used in Alg. 3.1.

[2] A matrix S is called *(upper) semiseparable* if all submatrices consisting of elements in the lower-triangular part of S only are of rank at most 1. Examples of semiseparable matrices are inverse upper-Hessenberg matrices. The *diagonal* matrix D in $H_m = S + D$ carries the finite poles ξ_j whilst the infinite ones can be replaced by any finite number. The structure of S captures the difference between finite and infinite poles.

Algorithm 3.1 Rational Krylov method [37, 39, 40]. RK Toolbox: `rat_krylov`

Input: $A \in \mathbb{C}^{N\times N}$, $\boldsymbol{v} \in \mathbb{C}^N$ and poles $\{\xi_j\}_{j=1}^m \subset \overline{\mathbb{C}} \setminus \Lambda(A)$ with $m < M$.
Output: RAD $AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}$ spanning $\mathcal{Q}_{m+1}(A, \boldsymbol{v})$ with poles $\{\xi_j\}_{j=1}^m$.

1. Set $\boldsymbol{v}_1 := \boldsymbol{v}/\|\boldsymbol{v}\|_2$, and $\boldsymbol{q}_1 := [1]$.
2. **for** $j = 1, \ldots, m$ **do**
3. Compute $\boldsymbol{w} := A^{(\xi_j)} V_j \boldsymbol{q}_j$.
4. Project $\boldsymbol{h}_j := V_j^* \boldsymbol{w}$, and compute $h_{j+1,j} := \|\boldsymbol{w} - V_j \boldsymbol{h}_j\|_2$.
5. Compute $\boldsymbol{v}_{j+1} := \left(\boldsymbol{w} - V_j \boldsymbol{h}_j\right)/h_{j+1,j}$ orthogonal to $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j$.
6. **if** $\xi_j \neq 0$ **then** Set $\underline{\boldsymbol{k}_j} := \underline{\boldsymbol{q}_j} + \underline{\boldsymbol{h}_j}/\xi_j$.
else Set $\underline{\boldsymbol{k}_j} := \underline{\boldsymbol{h}_j}$, and replace $\underline{\boldsymbol{h}_j} := \underline{\boldsymbol{q}_j}$. **end if**
7. For $j \neq m$ set $\boldsymbol{q}_{j+1} := \boldsymbol{q}_{j+1}^{(\xi_{j+1})}$, where $\underline{L_j^{(\xi_{j+1})}} =: Q_{j+1}^{(\xi_{j+1})} \underline{R_j}$ is a QR factorization.
8. **end for**

THEOREM 3.3. *Let* (2.2) *be an RAD associated with an A-variant space $\mathcal{R}(V_{m+1})$, and let the polynomials $p_j, q_j \in \mathcal{P}_j$ be as in* (2.10). *Define the polynomials*

$$p_j^{[m]}(z) := q_m(z) q_j(z)^{-1} p_j(z) \in \mathcal{P}_m, \quad j = 0, 1, \ldots, m.$$

Equivalently, $p_j^{[m]}(z)$ is the determinant of the $m \times m$ minor of $z\underline{K_m} - \underline{H_m}$ resulting from the removal of the jth row. Then for any $\xi \in \mathbb{C}$ there holds

$$\boldsymbol{q}_{m+1}^{(\xi)} \neq \boldsymbol{0} \quad \textit{and} \quad \left(\boldsymbol{q}_{m+1}^{(\xi)}\right)^* \underline{L_m^{(\xi)}} = \boldsymbol{0}^*,$$

where $\boldsymbol{q}_{m+1}^{(\xi)} := \begin{bmatrix} p_0^{[m]}(\xi) & p_1^{[m]}(\xi) & \ldots & p_m^{[m]}(\xi) \end{bmatrix}^*$.

Proof. Let $\xi \in \mathbb{C}$ be an arbitrary scalar. Label the roots of $p_j(z)$ as $\vartheta_1^{[j]}, \ldots, \vartheta_j^{[j]}$, for $j = 1, \ldots, m$, and the roots of q_m as $\xi_1, \ldots, \xi_m$, eventual roots at infinity included. It follows from (2.10) that for all $j = 1, \ldots, m$ and $i = 1, \ldots, j$ we have $\xi_j \neq \vartheta_i^{[j]}$. Otherwise we would have $\boldsymbol{v}_{j+1} \in \mathcal{Q}_j(A, \boldsymbol{v}_1, q_{j-1})$ and V_{j+1} would not be of full column rank. We are ready to prove that $\boldsymbol{q}_{m+1}^{(\xi)} \neq \boldsymbol{0}$. Assume that $\boldsymbol{q}_{m+1}^{(\xi)} = \boldsymbol{0}$. In particular, the first component of $\boldsymbol{q}_{m+1}^{(\xi)}$ is zero, and thus so is $p_0^{[m]}(\xi) = 0$. Hence $\xi \in \{\xi_1, \ldots, \xi_m\}$. Since $p_1^{[m]}(\xi) = 0$ and $\xi_1 \neq \vartheta_1^{[1]}$ we may further restrict $\xi \in \{\xi_2, \ldots, \xi_m\}$. Looking at $p_j^{[m]}(\xi) = 0$ for the remaining $j = 2, \ldots, m$ we exclude one by one all the ξ_j and have $\xi \in \emptyset$. Hence there is no ξ such that $\boldsymbol{q}_{m+1}^{(\xi)} = \boldsymbol{0}$.

Let us now prove that $\left(\boldsymbol{q}_{m+1}^{(\xi)}\right)^* \underline{L_m^{(\xi)}} = \boldsymbol{0}^*$. Note that $\mathcal{R}(V_{m+1}) = \mathcal{Q}_{m+1}(A, \boldsymbol{v}_1, q_m)$ implies $\mathcal{R}(q_m(A)V_{m+1}) = \mathcal{K}_{m+1}(A, \boldsymbol{v}_1)$, and further, $\mathcal{K}_{m+1}(A, \boldsymbol{v}_1)$ is A-variant since $\mathcal{R}(V_{m+1})$ is. Left-multiplying the RAD (2.2) with $q_m(A)$ yields $Aq_m(A)V_{m+1}\underline{K_m} = q_m(A)V_{m+1}\underline{H_m}$. The columns of $\breve{V}_{m+1} := q_m(A)V_{m+1}$ satisfy $\breve{\boldsymbol{v}}_j = p_{j-1}^{[m]}(A)\boldsymbol{v}_1$ and span the A-variant space $\mathcal{K}_{m+1}(A, \boldsymbol{v}_1)$. The natural isomorphism between $\mathcal{K}_{m+1}(A, \boldsymbol{v}_1)$ and $\mathcal{P}_m$ allows us to write the decomposition in scalar form as

$$z \begin{bmatrix} p_0^{[m]}(z) & p_1^{[m]}(z) & \ldots & p_m^{[m]}(z) \end{bmatrix} \underline{K_m} = \begin{bmatrix} p_0^{[m]}(z) & p_1^{[m]}(z) & \ldots & p_m^{[m]}(z) \end{bmatrix} \underline{H_m},$$

for all $z \in \mathbb{C}$. If $\xi = 0$ the result follows by setting $z = 0$. Otherwise, using $z = \xi$, and subtracting $\begin{bmatrix} p_0^{[m]}(\xi) & p_1^{[m]}(\xi) & \ldots & p_m^{[m]}(\xi) \end{bmatrix} \underline{H_m}$, gives the result for $\xi \neq 0$. ∎

REMARK 3.4 (Evaluating rational functions). *The introduction of polynomials* $p_j^{[m]}$ *is necessary only for the case when* $q_m(\xi) = 0$*, since then* $q_m(\xi)^{-1}$ *is infinite. When* $q_m(\xi) \neq 0$*, we can replace equivalently the evaluation* $p_j^{[m]}(\xi)$ *with the evaluation of the rational functions* $r_j(\xi) := q_j(\xi)^{-1} p_j(\xi)$*, for all* $j = 0, \dots, m$*. As* $r_0(\xi) = 1$*, we see that* $\overline{r_j(\xi)} = q^{(\xi)}_{j+1,m+1}/q^{(\xi)}_{1,m+1}$*, the ratio of the* $(j+1)$*st and first element of the vector* $\boldsymbol{q}^{(\xi)}_{m+1}$*. The rational functions* r_j *are such that* $V_{m+1} = \begin{bmatrix} r_0(A)\boldsymbol{v}_1 & r_1(A)\boldsymbol{v}_1 & \dots & r_m(A)\boldsymbol{v}_1 \end{bmatrix}$*, and Theorem 3.3 can be used to evaluate the functions* $r_j(z)$ *at arbitrary points* $z \in \mathbb{C}$*.*

4. Moving the poles. Let us give a brief resume. For a fixed rational Krylov space $\mathcal{Q}_{m+1} = \mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$ the poles are uniquely defined by the starting vector $\boldsymbol{v}$ and, up to scaling of $\boldsymbol{v}$, the reverse is true. Further, by Theorem 2.5, there exists an orthonormal RAD (2.2) spanning $\mathcal{Q}_{m+1}$ with starting vector $\boldsymbol{v}$ and poles q_m. Upon fixing the order of appearance of the poles, Theorem 3.2 guarantees the RAD to be essentially unique.

Observe that $\mathcal{Q}_{m+1}$ can be interpreted as a rational Krylov space with starting vector being almost any vector from $\mathcal{Q}_{m+1}$. Indeed, let a nonzero polynomial $\breve{q}_m \in \mathcal{P}_m$ have roots disjoint from $\Lambda(A)$, then

$$\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m) = \mathcal{Q}_{m+1}(A, \breve{q}_m(A) q_m(A)^{-1} \boldsymbol{v}, \breve{q}_m). \tag{4.1}$$

We are now interested in transforming an RAD (2.2) for $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$ into one for $\mathcal{Q}_{m+1}(A, \breve{q}_m(A) q_m(A)^{-1} \boldsymbol{v}, \breve{q}_m)$. For the moment this is only of theoretical importance, however, in subsection 4.3 we show a connection with implicit filtering and provide references to the literature. Further, an application of the ideas developed here to rational approximation is given in section 5.

To get an RAD for $\mathcal{Q}_{m+1}(A, \breve{q}_m(A) q_m(A)^{-1} \boldsymbol{v}, \breve{q}_m)$ one can focus on either the "new starting vector" or the "new poles". The result is essentially the same and both the starting vector and poles change. We first look at the case when the new starting vector is given as $\breve{\boldsymbol{v}} = V_{m+1}\boldsymbol{c}$, for a nonzero $\boldsymbol{c} \in \mathbb{C}^{m+1}$, and later we focus on the case when the new poles $\breve{q}_m$ are prescribed.

4.1. Moving the poles implicitly. Let $\breve{\boldsymbol{v}} = V_{m+1}\boldsymbol{c} \in \mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$ be a nonzero vector and take any nonsingular matrix P of size $m+1$ such that $P\boldsymbol{e}_1 = \boldsymbol{c}$. Then

$$A\breve{V}_{m+1}\underline{\breve{K}_m} = \breve{V}_{m+1}\underline{\breve{H}_m}, \tag{4.2}$$

where $\breve{V}_{m+1} = V_{m+1}P$, $\underline{\breve{H}_m} = P^{-1}\underline{H_m}$, and $\underline{\breve{K}_m} = P^{-1}\underline{K_m}$. This construction guarantees the first column $\breve{\boldsymbol{v}}_1$ of $\breve{V}_{m+1}$ to be $\breve{\boldsymbol{v}}$, however, the pencil $\left(\underline{\breve{H}_m}, \underline{\breve{K}_m}\right)$ may loose the upper-Hessenberg structure. In the following we aim at recovering this structure in (4.2) without affecting $\breve{\boldsymbol{v}}_1$. For that purpose we generalize the notion of RADs by first giving a technical definition. For a matrix $\underline{X_m} \in \mathbb{C}^{(m+1)\times m}$ the notation $\underline{X_{-m}}$ is used to denote its lower $m \times m$ submatrix.

DEFINITION 4.1. *Let* $\{\underline{\breve{K}_m}, \underline{\breve{H}_m}\} \subset \mathbb{C}^{(m+1)\times m}$ *be matrices. We say that the pencil* $\left(\underline{\breve{H}_m}, \underline{\breve{K}_m}\right)$ *is* regular *if the lower* $m \times m$ *subpencil* $\left(\underline{\breve{H}_{-m}}, \underline{\breve{K}_{-m}}\right)$ *is regular, i.e.,* $\breve{q}_m(z) = \det\left(z\underline{\breve{K}_{-m}} - \underline{\breve{H}_{-m}}\right)$ *is not identically equal to zero.*

Note that an upper-Hessenberg pencil of size $(m+1) \times m$ is unreduced if and only if it is regular. We are now ready to introduce decompositions of the form (4.2).

Algorithm 4.1 Implicit pole placement. RK Toolbox: `move_poles_impl`

Input: Generalized RAD $AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}$ and unit 2-norm $\boldsymbol{e}_1 \neq \boldsymbol{c} \in \mathbb{C}^{m+1}$.
Output: Generalized RAD (4.2) spanning $\mathcal{R}(V_{m+1})$ with $\breve{\boldsymbol{v}}_1 = V_{m+1}\boldsymbol{c}$.

1. Define $P := I_{m+1} - 2\boldsymbol{u}\boldsymbol{u}^*$, where $\boldsymbol{u} := (\boldsymbol{c} - \boldsymbol{e}_1)/\|\boldsymbol{c} - \boldsymbol{e}_1\|_2$.
2. Find unitary $Q = \mathrm{blkdiag}(1, Q_m)$ and Z, of order $m+1$ and m respectively, such that $Q^*P\underline{H_m}Z$ and $Q^*P\underline{K_m}Z$ are both upper-Hessenberg.
3. Define $\breve{V}_{m+1} := V_{m+1}PQ$, $\underline{\breve{H}_m} := Q^*P\underline{H_m}Z$ and $\underline{\breve{K}_m} := Q^*P\underline{K_m}Z$.

DEFINITION 4.2. *A relation of the form* (4.2) *where* $\breve{V}_{m+1}$ *is of full column rank and* $\big(\underline{\breve{H}_m}, \underline{\breve{K}_m}\big)$ *is regular is called a* generalized rational Krylov decomposition. *The generalized eigenvalues of* $\big(\underline{\breve{H}_{-m}}, \underline{\breve{K}_{-m}}\big)$ *are called* poles *of the decomposition. If the poles of* (4.2) *are outside the spectrum* $\Lambda(A)$*, then* (4.2) *is called a* rational Krylov decomposition (RKD).

The notion of *(orthonormal) basis, space* and *equivalent decompositions* are the same as for RADs. We call a generalized RKD with an upper-Hessenberg pencil a *generalized RAD*. The two definitions above let us speculate that the unique poles associated with $\breve{\boldsymbol{v}}$ are the eigenvalues of $\big(\underline{\breve{H}_{-m}}, \underline{\breve{K}_{-m}}\big)$. The justification follows from Theorem 2.5 (or Theorem 3.2) and the following result.

THEOREM 4.3. *Any generalized RKD is equivalent to a generalized RAD with the same starting vector.*

Proof. Let (4.2) be a generalized RKD. We need to bring both $\underline{\breve{H}_m}$ and $\underline{\breve{K}_m}$ into upper-Hessenberg form. To achieve this it suffices to bring $\big(\underline{\breve{H}_{-m}}, \underline{\breve{K}_{-m}}\big)$ into generalized Schur form. The existence of unitary $Q_m, Z_m \in \mathbb{C}^{m\times m}$ such that $Q_m^*\underline{\breve{H}_{-m}}Z_m$ and $Q_m^*\underline{\breve{K}_{-m}}Z_m$ are both upper-triangular follows from [19, Theorem 7.7.1]. Multiplying $A\breve{V}_{m+1}\underline{\breve{H}_m} = \breve{V}_{m+1}\underline{\breve{K}_m}$ from the right with Z_m and "inserting" $I_{m+1} = Q_{m+1}Q_{m+1}^*$, where $Q_{m+1} = \mathrm{blkdiag}(1, Q_m)$, we obtain the generalized RAD

$$A\underbrace{\big(\breve{V}_{m+1}Q_{m+1}\big)}_{V_{m+1}}\underbrace{Q_{m+1}^*\underline{\breve{K}_m}Z_m}_{\underline{K_m}} = \underbrace{\big(\breve{V}_{m+1}Q_{m+1}\big)}_{V_{m+1}}\underbrace{Q_{m+1}^*\underline{\breve{H}_m}Z_m}_{\underline{H_m}}.$$

Note that $\mathcal{R}(\breve{V}_{m+1}) = \mathcal{R}(V_{m+1})$ with the poles of $\big(\underline{\breve{H}_m}, \underline{\breve{K}_m}\big)$ and $\big(\underline{H_m}, \underline{K_m}\big)$ being identical. The first vector $\breve{\boldsymbol{v}}_1 = \boldsymbol{v}_1$ is unaffected. ∎

This discussion is summarized in Algorithm 4.1, used to replace the starting vector $\boldsymbol{v}$ with $\breve{\boldsymbol{v}} = V_{m+1}\boldsymbol{c}$. Note that there is no guarantee that by transforming an RAD the resulting decomposition is also an RAD, i.e., some poles may be moved to eigenvalues of A. We prove later (cf. Theorem 4.4) that if $\breve{\boldsymbol{v}} = V_{m+1}\boldsymbol{c} = p_m(A)q_m(A)^{-1}\boldsymbol{v}$ then the poles of the decomposition are always the roots of p_m, even if they coincide with eigenvalues of A.

4.2. Moving the poles explicitly. If the vector $\breve{\boldsymbol{v}}$ is not given as a linear combination $\breve{\boldsymbol{v}} = V_{m+1}\boldsymbol{c}$ of the basis vectors V_{m+1} but rather by specifying the new poles $\breve{q}_m$ one can compute $\boldsymbol{c} = V_{m+1}^*\breve{\boldsymbol{v}}$, where $\breve{\boldsymbol{v}} = \breve{q}_m(A)q_m(A)^{-1}\boldsymbol{v}$, and still use Alg. 4.1 to recover the new decomposition. The vector $\breve{\boldsymbol{v}} = \breve{q}_m(A)q_m(A)^{-1}\boldsymbol{v}$ can be computed cheaply as a rational Arnoldi approximation $\breve{\boldsymbol{v}} = V_{m+1}\breve{q}_m(A_{m+1})q_m(A_{m+1})^{-1}V_{m+1}^*\boldsymbol{v}$, where $A_{m+1} := V_{m+1}^*AV_{m+1}$, see for instance [27]. In the following we present an approach that works directly with the pencil $\big(\underline{H_m}, \underline{K_m}\big)$, changing the poles iteratively

one by one, and thence requires no information about the reduced matrix A_{m+1}.

Moving the first pole. The poles are the ratios of the subdiagonal elements of $(\underline{H_m}, \underline{K_m})$. Applying a Givens rotation G acting on planes $(1,2)$ from the left of the pencil does not destroy the upper-Hessenberg structure and, as we show, can move the first pole anywhere. We now derive the formulas for $s = e^{i\phi}\sin\vartheta$ and $c = \cos\vartheta$ satisfying $c^2 + |s|^2 = 1$ and such that the Givens rotation

$$G = \operatorname{blkdiag}\left(\begin{bmatrix} c & -s \\ \overline{s} & c \end{bmatrix}, I_{m-1}\right)$$

replaces the pole ξ_1 with $\breve{\xi}_1$ when applied to the pencil from the left. Define $\underline{\breve{H}_m} = G\underline{H_m}$ and $\underline{\breve{K}_m} = G\underline{K_m}$. This gives

$$\begin{aligned} \breve{h}_{11} &= ch_{11} - sh_{21}, \quad \breve{k}_{11} = ck_{11} - sk_{21}, \\ \breve{h}_{21} &= \overline{s}h_{11} + ch_{21}, \quad \breve{k}_{21} = \overline{s}k_{11} + ck_{21}. \end{aligned} \tag{4.3}$$

Additionally, G is chosen so that $\breve{\xi}_1 = \breve{h}_{21}/\breve{k}_{21}$. Using the notation $t = \overline{s}/c$, we derive

$$t = \begin{cases} -k_{21}/k_{11}, & \breve{\xi}_1 = \infty, \\ \big(\breve{\xi}_1 k_{21} - h_{21}\big)/\big(h_{11} - \breve{\xi}_1 k_{11}\big), & \breve{\xi}_1 \neq \infty. \end{cases}$$

Using standard trigonometric relations we arrive at

$$s = \frac{\overline{t}}{\sqrt{1+|t|^2}}, \qquad c = \frac{1}{\sqrt{1+|t|^2}}$$

if $t \neq \infty$, and otherwise, $s = 1$ and $c = 0$.

Formula (4.1) asserts (with the roots of $\breve{q}_m$ being $\breve{\xi}_1, \xi_2, \ldots, \xi_m$) that this process replaces the starting vector $\boldsymbol{v}_1$ with a multiple of $\big(A - \breve{\xi}_1 I\big)\big(A - \xi_1 I\big)^{-1}\boldsymbol{v}_1$, where for notational convenience only we assume both ξ_1 and $\breve{\xi}_1$ to be finite. Let us verify that. Define $\breve{V}_{n+1} = V_{n+1}G^*$. In particular,

$$\breve{\boldsymbol{v}}_1 = c\boldsymbol{v}_1 - \overline{s}\boldsymbol{v}_2. \tag{4.4}$$

Recall that (2.6) reads $\big(h_{21}I - k_{21}A\big)\boldsymbol{v}_2 = \big(k_{11}A - h_{11}I\big)\boldsymbol{v}_1$. Hence, using the relation (2.6) within (4.4) together with (4.3) provides

$$\big(h_{21}I - k_{21}A\big)\breve{\boldsymbol{v}}_1 = \big[c\big(h_{21}I - k_{21}A\big) - \overline{s}\big(k_{11}A - h_{11}I\big)\big]\,\boldsymbol{v}_1 = \big(\breve{h}_{21}I - \breve{k}_{21}A\big)\boldsymbol{v}_1. \tag{4.5}$$

Note that (2.6) holds even if $h_{21}/k_{21} = \xi_1 \in \Lambda(A)$ as long as the generalized RAD (2.2) exists. As we impose no constraints on $\breve{\xi}_1$, we conclude that (4.5) holds even if $\breve{\xi}_1 \in \Lambda(A)$ and/or $\xi_1 \in \Lambda(A)$. If however $\xi_1 \notin \Lambda(A)$ we can further write

$$\breve{\boldsymbol{v}}_1 = \big(\breve{h}_{21}I - \breve{k}_{21}A\big)\big(h_{21}I - k_{21}A\big)^{-1}\boldsymbol{v}_1.$$

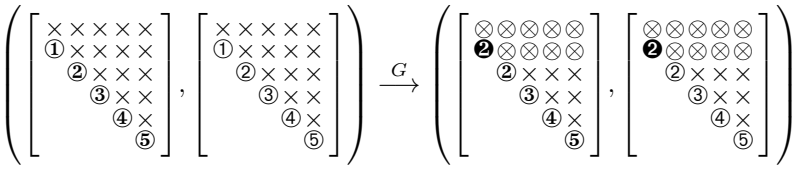
Moving all poles. Changing the other ratios with Givens rotations results in the loss of the upper-Hessenberg structure. However, the poles are the eigenvalues of the pencil $(\underline{H_{-m}}, \underline{K_{-m}})$ which is (already) in generalized Schur form. After changing the first pole, using the Givens rotation approach just described, the poles can be reordered (see for instance [31, 32]) with the aim of bringing an unchanged pole to

Algorithm 4.2 Explicit pole placement. RK Toolbox: `move_poles_expl`

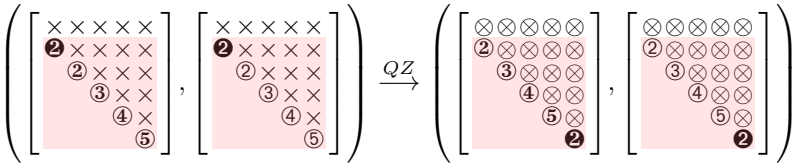
Input: Generalized RAD $AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}$ and $\breve{\boldsymbol{\xi}} = \{\breve{\xi}_j\}_{j=1}^{k} \subset \overline{\mathbb{C}}$, $1 \le k \le m$.

Output: Generalized RAD (4.2) spanning $\mathcal{R}(V_{m+1})$ and having poles $\breve{\boldsymbol{\xi}} \cup \{\xi_j\}_{j=k+1}^{m}$.

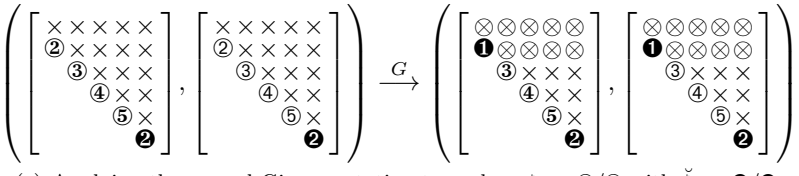
1. Set $\breve{V}_{m+1} := V_{m+1}$, $\underline{\breve{H}_m} := \underline{H_m}$, and $\underline{\breve{K}_m} := \underline{K_m}$.
2. Label $\xi_j := h_{j+1,j}/k_{j+1,j}$ for $j = 1, \ldots, k$.
3. **for** $j = 1, \ldots, k$ **do**
4. Find Givens rotation G to replace the pole ξ_j with $\breve{\xi}_\ell$ where $\ell = k - j + 1$.
5. Update $\breve{V}_{m+1} := \breve{V}_{m+1}G^*$, $\underline{\breve{H}_m} := G\underline{\breve{H}_m}$, and $\underline{\breve{K}_m} := G\underline{\breve{K}_m}$.
6. Find unitary $Q_{m+1} = \mathrm{blkdiag}(1, Q_\ell, I_{m-\ell})$ and $Z_m = \mathrm{blkdiag}(Z_\ell, I_{m-\ell})$, to circularly shift the ℓ poles from position $(2,1)$ to position $(\ell+1, \ell)$ for one place forward so that $\breve{\xi}_\ell$ gets pushed back to position $(\ell+1, \ell)$.
7. Update $\breve{V}_{m+1} := \breve{V}_{m+1}Q_{m+1}$, $\underline{\breve{H}_m} := Q_{m+1}^*\underline{\breve{H}_m}Z_m$, and $\underline{\breve{K}_m} := Q_{m+1}^*\underline{\breve{K}_m}Z_m$.
8. **end for**



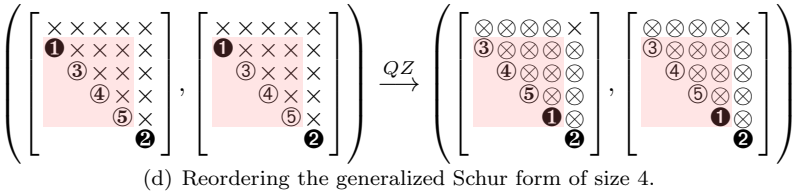
(a) Applying the first Givens rotation to replace $\xi_1 = ①/①$ with $\breve{\xi}_2 = ❷/❷$.



(b) Reordering the generalized Schur form of the lower 5×5 part.



(c) Applying the second Givens rotation to replace $\xi_2 = ②/②$ with $\breve{\xi}_1 = ❶/❶$.



(d) Reordering the generalized Schur form of size 4.

Fig. 4.1: Looking at the 6×5 upper-Hessenberg pencil while Algorithm 4.2 is applied on the corresponding RAD with $k = 2$. The original poles are the ratios ①/①,..., ⑤/⑤. The first two poles are replaced with ❷/❷ and ❶/❶. The transition from $\times$ to $\otimes$ symbolizes that the element potentially changes.

the front of the decomposition so that it can be changed using a Givens rotation. This process is formalized in Algorithm 4.2 and an illustration is presented in Figure 4.1.

Let us now consider Alg. 4.2 when $k = m$. For notational convenience only, we assume the poles to be finite. As we have shown in (4.5), after applying the first Givens rotation the starting vector $\boldsymbol{v}_1$ gets replaced with $\boldsymbol{v}_1^{[1]}$ satisfying

$$\big(A - \xi_1 I\big)\boldsymbol{v}_1^{[1]} = \gamma_1\big(A - \breve{\xi}_m I\big)\boldsymbol{v}_1, \tag{4.6}$$

where $0 \neq \gamma_1 \in \mathbb{C}$ is a scaling factor. By reordering the poles we do not affect the "new starting vector" $\boldsymbol{v}_1^{[1]}$ and bring ξ_2 to the leading positions, i.e., second row, first column, where the next Givens rotation acts. Thus, for $j = 2$ the Givens rotation replaces $\boldsymbol{v}_1^{[1]}$ with $\boldsymbol{v}_1^{[2]}$ satisfying $\big(A-\xi_2 I\big)\boldsymbol{v}_1^{[2]} = \gamma_2\big(A-\breve{\xi}_{m-1}I\big)\boldsymbol{v}_1^{[1]}$, for some $0 \neq \gamma_2 \in \mathbb{C}$. Using (4.6) we obtain

$$\big(A - \xi_1 I\big)\big(A - \xi_2 I\big)\boldsymbol{v}_1^{[2]} = \gamma_1\gamma_2\big(A - \breve{\xi}_{m-1}I\big)\big(A - \breve{\xi}_m I\big)\boldsymbol{v}_1.$$

Reasoning inductively we deduce

$$q_m(A)\breve{\boldsymbol{v}}_1 = \gamma\breve{q}_m(A)\boldsymbol{v}_1, \tag{4.7}$$

where $0 \neq \gamma \in \mathbb{C}$ is a scalar, $\breve{\boldsymbol{v}}_1 = \boldsymbol{v}_1^{[m]}$, q_m is given by (2.4), and $\breve{q}_m$ is defined in an analogous manner. The above discussion is the gist of the following result.

THEOREM 4.4. *Let $\mathcal{Q}_{m+1} = \mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$ be A-variant. If the generalized RKD $A\breve{V}_{m+1}\underline{\breve{K}_m} = \breve{V}_{m+1}\underline{\breve{H}_m}$ with poles $\breve{q}_m$ spans $\mathcal{Q}_{m+1}$ then $\breve{\boldsymbol{v}}_1 = \gamma\breve{q}_m(A)q_m(A)^{-1}\boldsymbol{v}$ with a scalar $0 \neq \gamma \in \mathbb{C}$. Alternatively, if $\breve{\boldsymbol{v}}_1 = \breve{q}_m(A)q_m(A)^{-1}\boldsymbol{v}$ then there exists a generalized RKD $A\breve{V}_{m+1}\underline{\breve{K}_m} = \breve{V}_{m+1}\underline{\breve{H}_m}$ with poles $\breve{q}_m$ spanning $\mathcal{Q}_{m+1}$.*

Proof. If $A\breve{V}_{m+1}\underline{\breve{K}_m} = \breve{V}_{m+1}\underline{\breve{H}_m}$ spans $\mathcal{Q}_{m+1}$ we can transform it into an equivalent generalized RAD (cf. Theorem 4.3) and then, using Alg. 4.2, into $AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}$, having poles q_m and still spanning $\mathcal{Q}_{m+1}$. According to Lemma 2.1, $\boldsymbol{v}_1$ is collinear with $\boldsymbol{v}$. Therefore, it follows from (4.7) that $\breve{\boldsymbol{v}}_1 = \gamma\breve{q}_m(A)q_m(A)^{-1}\boldsymbol{v}$ for some scalar $0 \neq \gamma \in \mathbb{C}$. The other direction follows from Theorem 2.5 and (4.7) after using Alg. 4.2. ∎

Theorem 4.4 shows that Alg. 4.1 and Alg. 4.2 are equivalent, provided that equivalent input data are given. It also shows, together with Theorem 2.5 and Theorem 4.3, that an $(m+1)$-dimensional space $\mathcal{V}_{m+1}$ is a rational Krylov space if and only if there exist a generalized RKD spanning $\mathcal{V}_{m+1}$.

REMARK 4.5 (Recovering the polynomial Krylov space). *With $\breve{q}_m(z) = 1$, there holds $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m) = \mathcal{Q}_{m+1}(A, q_m(A)^{-1}\boldsymbol{v}, \breve{q}_m) = \mathcal{K}_{m+1}(A, q_m(A)^{-1}\boldsymbol{v})$, and we can recover a polynomial Arnoldi decomposition for $\mathcal{K}_{m+1}(A, q_m(A)^{-1}\boldsymbol{v})$ from an RAD for $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$ using Alg. 4.2 with all poles $\breve{\xi}_j$ set to infinity. In this particular case, a simpler method is to bring the pencil $\big(\underline{H_m}, \underline{K_m}\big)$ from upper-Hessenberg–upper-Hessenberg to upper-Hessenberg–upper-triangular form using just Givens rotations; see, e.g., [42, p. 495]. Bringing the pencil to upper-triangular–upper-Hessenberg form would move all the poles to zero.*

4.3. Implicit filters in the rational Krylov method. Implicit filtering aims at compressing the space $\mathcal{Q}_{m+1}(A, \boldsymbol{v}_1, q_m)$ into $\mathcal{Q}_{m+1-k}(A, p_k(A)q_k(A)^{-1}\boldsymbol{v}_1, \breve{q}_{m-k})$, where $1 \leq k \leq m$, $q_m = q_k \cdot \breve{q}_{m-k}$, and $p_k \in \mathcal{P}_k$ is a polynomial with roots (infinity allowed) in the region we want to filter out. In applications this technique is usually

used to deal with large memory requirements or orthogonalization costs for V_{m+1}, or to purge unwanted or spurious eigenvalues (see, e.g., [5, 8, 9] and the references given therein). Implicit filtering for RADs was first introduced in [9] and further studied in [8]. Alg. 4.2 can easily be used for implicit filtering. In fact, applying Alg. 4.2 with the k poles $\breve{\xi}_j$ being the roots of p_k implicitly applies the filter $p_k(A)q_k(A)^{-1}$ to the RAD. The k "new" poles correspond to the rightmost k columns in $\breve{V}_{m+1}$, $\underline{\breve{K}_m}$ and $\underline{\breve{H}_m}$, cf. Figure 4.1. Hence, truncating the decomposition to the leading $m+1-k$ columns completes the process. The derivation and algorithms in [8, 9] are different, and it would perhaps be interesting to compare them. This is, however, not done here. Pertinent ideas for polynomial Krylov methods have recently appeared in [5] where the authors relate implicit filtering in the Krylov–Schur algorithm [43, 45] with partial eigenvalue assignment.

As an alternative to Alg. 3.1 for a Hermitian matrix A, it was proposed in [34] to use the spectral transformation Lanczos method with change of (the repeated) pole. The approach for changing poles taken here is different and more general.

5. An application to rational least squares approximation. Given matrices $\{A, F\} \subset \mathbb{C}^{N\times N}$ and a unit 2-norm vector $\boldsymbol{v} \in \mathbb{C}^N$, we consider in this section the following rational least squares problem: find a rational function R_m of type (m, m), with $m < M$, such that

$$\|F\boldsymbol{v} - R_m(A)\boldsymbol{v}\|_2 \to \min. \tag{5.1}$$

This is a nonlinear approximation problem as the denominator of R_m is unknown. Hence an iterative algorithm is required.

Let $q_m \in \mathcal{P}_m$ with no roots in $\Lambda(A)$ be a given polynomial and consider the linear space of rational functions of type (m, m) with denominator q_m, denoted by $\mathcal{P}_m/q_m$. Each element $R_m \in \mathcal{P}_m/q_m$ is in a one-to-one correspondence with an element $R_m(A)\boldsymbol{v}$ of $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$. Instead of (5.1) we now consider a linear approximation problem: find a unit 2-norm vector $\breve{\boldsymbol{v}} \in \mathcal{Q}_{m+1} = \mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$ as

$$\breve{\boldsymbol{v}} = \operatorname*{argmin}_{\substack{\boldsymbol{y}\in\mathcal{Q}_{m+1}\\ \|\boldsymbol{y}\|_2=1}} \min_{R_m\in\mathcal{P}_m/q_m} \|F\boldsymbol{y} - R_m(A)\boldsymbol{v}\|_2. \tag{5.2}$$

This means that $F\breve{\boldsymbol{v}}$ is best approximated by an element of $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$. Problem (5.2) is easy to solve. Let $V_{m+1} \in \mathbb{C}^{N\times(m+1)}$ be an orthonormal basis of $\mathcal{Q}_{m+1}$ and write $\boldsymbol{y} = V_{m+1}\boldsymbol{c}$ with $\boldsymbol{c} \in \mathbb{C}^{m+1}$ and $\|\boldsymbol{c}\|_2 = 1$. Then the inner minimum in (5.2) is a linear least squares problem whose solution $R_m(A)\boldsymbol{v}$ is given by orthogonal projection of $F\boldsymbol{y}$ onto $\mathcal{Q}_{m+1}$, i.e., $R_m(A)\boldsymbol{v} = V_{m+1}V_{m+1}^*F\boldsymbol{y}$ minimizes $\|F\boldsymbol{y} - R_m(A)\boldsymbol{v}\|_2$ for a fixed vector $\boldsymbol{y}$. Hence (5.2) reduces to

$$\breve{\boldsymbol{v}} = \operatorname*{argmin}_{\substack{\boldsymbol{y}=V_{m+1}\boldsymbol{c}\\ \|\boldsymbol{c}\|_2=1}} \|(I - V_{m+1}V_{m+1}^*)F\boldsymbol{y}\|_2.$$

A minimizing coefficient vector $\boldsymbol{c} = V_{m+1}^*\boldsymbol{y}$ can be obtained as a right singular vector of $(I - V_{m+1}V_{m+1}^*)FV_{m+1}$ corresponding to a smallest singular value $\sigma_{\min}$. We now exploit that by Theorem 4.4 we can associate with $\breve{\boldsymbol{v}}$ a "new" rational Krylov space $\mathcal{Q}_{m+1}(A, \breve{\boldsymbol{v}}, \breve{q}_m) = \mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$, where the roots of $\breve{q}_m$, the "new" poles, can be computed from $\boldsymbol{c}$ using Alg. 4.1. The vector–pole pair $(\breve{\boldsymbol{v}}, \breve{q}_m)$ is optimal in the sense that $\|F\breve{\boldsymbol{v}} - \breve{R}_m(A)\breve{\boldsymbol{v}}\|_2$ is minimal (and equal to $\sigma_{\min}$) among all $\breve{R}_m \in \mathcal{P}_m/\breve{q}_m$, and

Algorithm 5.1 Rational Krylov fitting (RKFIT). RK Toolbox: `rkfit`

Input: $\{A, F\} \subset \mathbb{C}^{N\times N}$, $\boldsymbol{v} \in \mathbb{C}^N$ and poles $\{\xi_j\}_{j=1}^m \subset \overline{\mathbb{C}} \setminus \Lambda(A)$ with $m < M$.
Output: Poles $\xi_1, \xi_2, \ldots, \xi_m$ and coefficient vector $\boldsymbol{c} \in \mathbb{C}^{m+1}$.

1. **repeat**
2. Compute RAD $AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}$ with $\boldsymbol{v}_1 = \boldsymbol{v}/\|\boldsymbol{v}\|_2$ and poles $\{\xi_j\}_{j=1}^m$.
3. Compute a right singular vector $\boldsymbol{c} \in \mathbb{C}^{m+1}$ of $(FV_{m+1} - V_{m+1}V_{m+1}^*FV_{m+1})$ corresponding to a smallest singular value $\sigma_{\min}$.
4. Form $A\breve{V}_{m+1}\underline{\breve{H}_m} = \breve{V}_{m+1}\underline{\breve{K}_m}$ spanning $\mathcal{R}(V_{m+1})$ with $\breve{\boldsymbol{v}}_1 = V_{m+1}\boldsymbol{c}$, cf. Alg. 4.1.
5. Obtain new poles $\xi_1, \xi_2, \ldots, \xi_m$ as the poles of $(\underline{\breve{H}_m}, \underline{\breve{K}_m})$.
6. **until** $\sigma_{\min}$ is small enough or a maximal iteration count is exceeded.
7. Compute RAD $AV_{m+1}\underline{K_m} = V_{m+1}\underline{H_m}$ with $\boldsymbol{v}_1 = \boldsymbol{v}/\|\boldsymbol{v}\|_2$ and poles $\{\xi_j\}_{j=1}^m$.
8. Compute $\boldsymbol{c} = V_{m+1}^*F\boldsymbol{v}$.

there is no better vector–pole pair associated with $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$. Replacing $\breve{\boldsymbol{v}}$ back to $\boldsymbol{v}$, we hope that the new rational Krylov space $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, \breve{q}_m)$ contains a better approximation to $F\boldsymbol{v}$ than $\mathcal{Q}_{m+1}(A, \boldsymbol{v}, q_m)$. In this case we have found an improved denominator $\breve{q}_m$ for the rational function R_m in (5.1).

The procedure described is iterated, computing a new rational Krylov space at each iteration and changing the poles by modifying the starting vector. The complete procedure is given in Algorithm 5.1 under the name RKFIT, which stands for *Rational Krylov Fitting*. A MATLAB implementation of RKFIT is available in [2].

Discussion. In this and the following subsections we briefly discuss Alg. 5.1 in a list of comments and some numerical experiments. A more detailed analysis will be given in a separate publication [3].

1. If $A = \operatorname{diag}(\lambda_j)$ and $F = \operatorname{diag}(\varphi_j)$ are diagonal matrices, and $\boldsymbol{v} = [v_1, \ldots, v_N]^T$, then (5.1) corresponds to a rational weighted least squares problem

$$\|F\boldsymbol{v} - R_m(A)\boldsymbol{v}\|_2^2 = \sum_{j=1}^{N} |v_j|^2 \cdot |\varphi_j - R_m(\lambda_j)|^2 \to \min.$$

Nonlinear rational optimization problems of this type are nonconvex and hence no numerical solution method can come with a guarantee to find a global minimum. In fact, the existence of a minimum is not even guaranteed, but this has not prevented the development of solution methods for these practically important problems. One popular approach is known as *vector fitting* [25, 23], which similarly to Alg. 5.1 is based on the iterative relocation of poles of rational functions. In contrast to Alg. 5.1, vector fitting uses a partial fraction representation of the rational functions [25].

2. The use of partial fractions as basis functions may result in poorly conditioned linear algebra problems to be solved, and *orthonormal vector fitting* [11] tries to overcome this problem by using instead an expansion of R_m into orthonormal rational functions; orthonormal with respect to a measure supported on the imaginary axis. The orthonormal rational functions in [11] are computed by a Gram–Schmidt procedure applied to partial fractions, which merely transforms an ill-conditioned basis into an orthonormal basis and still incurs numerical problems of ill-conditioning, see [24]. If the sampling points λ_j can be chosen freely, then one way to improve stability is to choose them based on quadrature rules associated with orthogonal polynomials (see, e.g., the discussion of quadrature-based vector fitting in [12]).

3. In the case that A is a normal matrix, RKFIT can be interpreted as an orthonormal vector fitting algorithm where two rational functions r_m and $\breve{r}_m$ with common denominator q_m are orthonormal with respect to a discrete inner product defined as

$$\langle r_m, \breve{r}_m \rangle := (\breve{r}_m(A)\boldsymbol{v})^*(r_m(A)\boldsymbol{v})/\|\boldsymbol{v}\|_2^2.$$

(See [6, 10] for the theory of orthogonal rational functions and their relation to rational Krylov spaces.) RKFIT is different from orthonormal vector fitting in that it uses a discrete inner product defined by a measure not necessarily supported on the imaginary axis. The orthonormal rational functions are computed by the rational Krylov method without the need for explicit quadrature. In [7] it is advocated to use orthogonal rational basis functions with fixed poles for least squares fitting. This leads to a linear least squares problem but does not resolve the problem of pole relocation.

4. If A has Jordan blocks of size 2 or greater then also derivatives of $R_m(z)$ at (some of) the eigenvalues of A are fitted. Consider, for example,

$$A = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}, \text{ and } R_m(A) = \begin{bmatrix} R_m(\lambda) & R_m'(\lambda) & \cdots & \frac{R_m^{(N-1)}(\lambda)}{(N-1)!} \\ & R_m(\lambda) & \ddots & \vdots \\ & & \ddots & R_m'(\lambda) \\ & & & R_m(\lambda) \end{bmatrix}.$$

Then each component of $R_m(A)\boldsymbol{v}$ is a weighted sum of derivatives $R_m^{(j)}(\lambda)$ and one can use $\boldsymbol{v}$ to choose the weights as required. This generalizes naturally to matrices A with more than one Jordan block.

5. If $F = f(A)$, each iteration of Alg. 5.1 requires the computation of $f(A)V_{m+1}$. If one is interested in scalar rational approximation problems where A is a diagonal matrix (see point 1), or a Jordan block matrix (see point 4), then $f(A)$ is easy to compute. Otherwise rational Krylov techniques can be used to approximate $f(A)V_{m+1}$ directly (see, e.g., the review [27]).

6. If A, F, and $\boldsymbol{v}$ are real-valued and the initial poles $\{\xi_j\}_{j=1}^m$ appear in complex conjugate pairs, it is natural to enforce real arithmetic in all operations of Alg. 5.1. This can be achieved by using the real form of the rational Krylov method [38] instead of Alg. 3.1, and a generalized real Schur form (see, e.g., [44, § 3.1]) in Step 2 of Alg. 4.1. Our RKFIT implementation [2] provides a `'real'` option for this purpose.

7. The output vector $\boldsymbol{c} \in \mathbb{C}^{m+1}$ returned by Alg. 5.1 collects the coefficients of the approximant $R_m(A)\boldsymbol{v}$ in the rational Krylov basis V_{m+1}, i.e., $R_m(A)\boldsymbol{v} = V_{m+1}\boldsymbol{c}$. Using Theorem 3.3 and Remark 3.4 we find that $R_m(z)$ can be evaluated for any point $z \in \mathbb{C}$ by computing a full QR factorization of $z\underline{K_m} - \underline{H_m}$ and forming an inner product of $\boldsymbol{c}$ with the last column $\boldsymbol{q}_{m+1}^{(z)}$ of the Q factor scaled by its first entry, i.e.,

$$R_m(z) = \frac{(\boldsymbol{q}_{m+1}^{(z)})^*\boldsymbol{c}}{(\boldsymbol{q}_{m+1}^{(z)})^*\boldsymbol{e}_1}.$$

In the following we discuss three experiments with the aim of providing further insight and showing the applicability of RKFIT. Accompanying MATLAB scripts to reproduce these experiments are available as part of [2]. All computations were performed with MATLAB version R2013a on an Intel Core i5-3570 processor running Scientific Linux, Release 6.4 (Carbon).

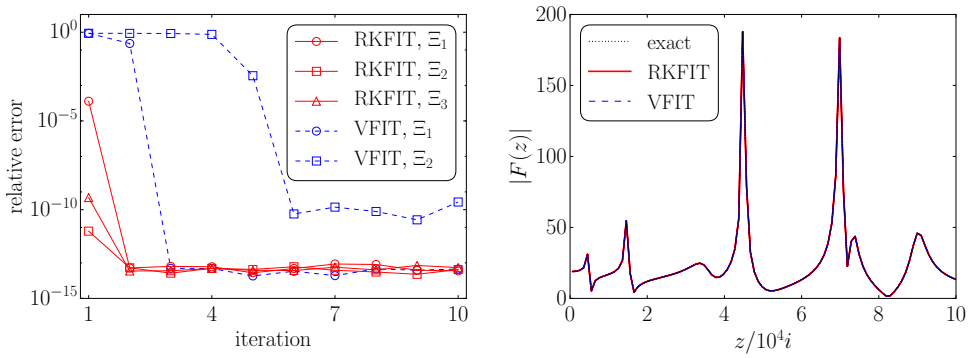


Fig. 5.1: Least-squares approximation of a rational function $F(z)$ of type $(19, 18)$ using RKFIT and the vector fitting code VFIT. Left: Relative error $\|F(A)\boldsymbol{v}-R_m(A)\boldsymbol{v}\|_2/\|F(A)\boldsymbol{v}\|_2$ after each iteration of RKFIT (solid red) and VFIT (dashed blue). The two convergence curves for each method correspond to different choices for the initial poles Ξ_1 (circles), Ξ_2 (squares), and Ξ_3 (triangles), respectively. Right: Plot of $|F(z)|$ over an interval on the imaginary axis overlaid with the approximants $|R_m(z)|$ obtained after 10 iterations of RKFIT and VFIT with initial poles Ξ_1 (the curves are visually indistinguishable).

5.1. Experiment 1: Fitting an artificial frequency response. We first consider a diagonal matrix $A \in \mathbb{C}^{N\times N}$ with $N = 200$ linearly spaced eigenvalues in the interval $[10^{-5}i, 10^5 i]$. The matrix $F = F(A)$ is a rational matrix function of type $(19, 18)$ given in partial fraction form in [25, subsection 4.1], and $\boldsymbol{v} = [1, 1, \ldots, 1]^T$. We compare RKFIT to the vector fitting code VFIT [25, 23] which is available online.[3]

We consider three different sets of starting poles, namely

- Ξ_1: 9 log-spaced poles in $[10^3 i, 10^5 i]$ and their complex conjugates;
- Ξ_2: 12 log-spaced poles in $[10^6 i, 10^9 i]$ and their complex conjugates;
- Ξ_3: 18 infinite poles (applicable to RKFIT only);

and run 10 iterations of RKFIT and VFIT, respectively.

The numerical results are shown in Figure 5.1. On the left we see the relative error $\|F(A)\boldsymbol{v} - R_m(A)\boldsymbol{v}\|_2/\|F(A)\boldsymbol{v}\|_2$ after each iteration. We observe that RKFIT converges within the first 2 iterations for all three sets of initial poles Ξ_1, Ξ_2, and Ξ_3. VFIT requires 3 iterations starting with Ξ_1 and it fails to converge within 10 iterations when being initialized with the poles Ξ_2. In the later case MATLAB warnings about solves of close-to-singular linear systems seem to indicate that the partial fraction basis used in VFIT is ill-conditioned. RKFIT, on the other hand, always uses discrete orthonormal rational bases and performs robustly with respect to changes in the initial poles. The choice of infinite initial poles Ξ_3 is interesting in that it requires no a-priori knowledge of the pole location (choosing all poles to be infinite is not possible in the available VFIT code). On the right of Figure 5.1 we show a plot $|F(z)|$ over an interval on the imaginary axis together with the RKFIT and VFIT approximants $|R_m(z)|$. This plot essentially coincides with [25, Figure 1] (it does not exactly coincide as apparently the figure in that paper has been produced with a smaller number of sampling points, causing some "spikes" to be missed or reduced).

[3] `http://www.sintef.no/Projectweb/VECTFIT/Downloads/VFUT3/` as of November 2014.

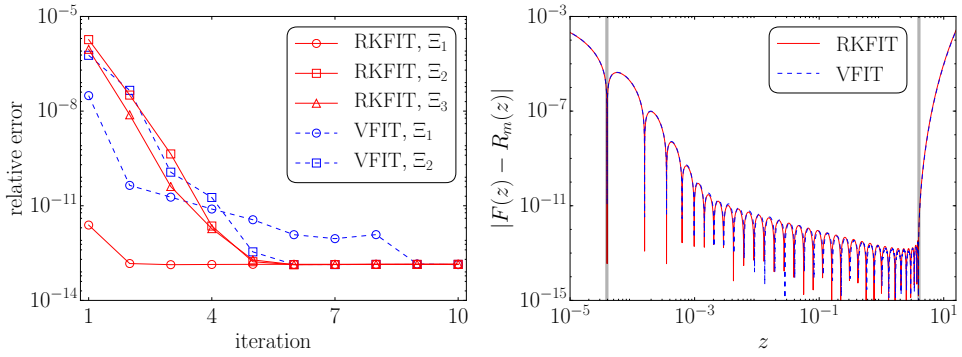


Fig. 5.2: Least-squares approximation of the function $F(z) = z^{1/2}$ using RKFIT and the vector fitting code VFIT. Left: This plot shows the relative approximation error $\|F(A)\boldsymbol{v} - R_m(A)\boldsymbol{v}\|_2/\|F(A)\boldsymbol{v}\|_2$ after each iteration of RKFIT (solid red) and VFIT (dashed blue). The convergence curves for each method correspond to different choices for the initial poles Ξ_1 (circles), Ξ_2 (squares), and Ξ_3 (triangles), respectively. Right: This is the plot of $|F(z) - R_m(z)|$ over an interval on the positive real axis obtained after 10 iterations of RKFIT and VFIT with initial poles Ξ_1. The vertical lines indicate the spectral interval of A.

5.2. Experiment 2: Square root of a symmetric matrix. We consider the approximation of $F\boldsymbol{v}$ with the matrix square root $F = A^{1/2}$, $A = \text{tridiag}(-1, 2, -1) \in \mathbb{R}^{100\times 100}$, and $\boldsymbol{v} = [1, 0, \ldots, 0]^T$. Again, we test different sets of initial poles, namely

- Ξ_1: 16 log-spaced poles in $[-10^8, -10^{-8}]$;
- Ξ_2: 16 linearly spaced poles in $[0, 4]$;
- Ξ_3: 16 infinite poles (applicable to RKFIT only).

Note that the initial poles Ξ_1 are placed on the branch cut of $z^{1/2}$, which is a reasonable initial guess for the poles of R_m. Some of the poles Ξ_2 are located very close to the eigenvalues of A whose spectral interval is approximately $[0, 4]$. The convergence of the relative error per iteration of RKFIT and VFIT is shown on the left of Figure 5.2. In order to use VFIT for this problem we have diagonalized A and provided the code with weights corresponding to the components of $\boldsymbol{v}$ in the eigenvector basis of A. All tests converge within at most 9 iterations, with the fastest convergence achieved by RKFIT with initial guess Ξ_1. On the right of Figure 5.2 we show the error $|z^{1/2} - R_m(z)|$ over an interval containing the spectrum of A.

5.3. Experiment 3: Exponential of a nonnormal matrix. We consider the approximation of $F\boldsymbol{v}$ with the matrix exponential $F = \exp(A)$ of a Grcar matrix A of size $N = 100$ generated in MATLAB via `A = -5*gallery('grcar',N,3)`. The eigenvalues and 10^{-6}-pseudospectrum of A are shown on the right of Figure 5.3. The vector is $\boldsymbol{v} = [1, 1, \ldots, 1]^T$ and we consider different sets of initial poles for RKFIT,

- Ξ_1: 16 poles equal to 0;
- Ξ_2: 16 poles equal to -10;
- Ξ_3: 16 infinite poles.

Note that A is not diagonalizable and therefore VFIT cannot be applied as in the previous two experiments. On the left of Figure 5.3 we observe excellent convergence of RKFIT within 2 iterations starting with the initial poles Ξ_1 and Ξ_3.

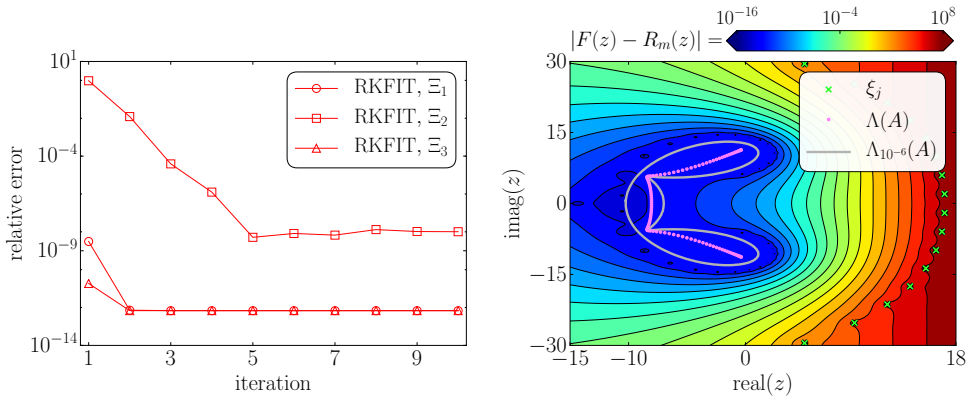


Fig. 5.3: Least-squares approximation of the function $F(z) = \exp(z)$ using RKFIT. Left: This plot shows the relative approximation error $\|F(A)\boldsymbol{v} - R_m(A)\boldsymbol{v}\|_2/\|F(A)\boldsymbol{v}\|_2$ after each iteration of RKFIT (solid red) for different choices of initial poles Ξ_1, Ξ_2, and Ξ_3, respectively. Right: A plot of $|F(z) - R_m(z)|$ over a region in the complex plane together with the poles of R_m (green crosses), where R_m is the rational least squares approximant obtained after 10 iterations of RKFIT with initial poles Ξ_1. The eigenvalues of the Grcar matrix and its 10^{-6}-pseudospectrum are also shown.

With the initial poles Ξ_2 the error stagnates on a higher level, possibly trapped nearby a non-global minimum. As is the case with any nonlinear iteration, RKFIT is not guaranteed to converge to a global minimum (if it even exists). We currently do not have a good explanation why the initial guess Ξ_2 is bad, but we have verified that $\xi = -10$ lies in the 10^{-6}-pseudospectrum of A and hence the initial rational Krylov space may have too large components in just a few eigendirections of A.

6. Summary and future work. We introduced the notion of generalized rational Krylov decompositions and studied their connections with rational Krylov spaces. We generalized the implicit Q theorem to the rational case and have provided some insight for the continuation combination proposed by Ruhe [40] for building rational Krylov spaces. Algorithms for transforming generalized RKDs and thereby changing the poles and starting vector of the associated spaces were presented. These algorithms, in particular Alg. 4.2, can also be employed for implicit restarting in polynomial and rational Krylov decompositions and even eigenvalue assignment, cf. [5]. A comparison with existing algorithms for the same purpose might be interesting.

We introduced the RKFIT algorithm for rational least squares approximation. A more detailed analysis of the convergence properties will be subject of future work. We will extend the MATLAB code in [2] to return the computed rational approximant in partial fraction form, although this conversion itself may be ill-conditioned in particular when the poles of the approximant are close to each other and/or far away from the eigenvalues of A. A further extension of RKFIT will handle "lucky breakdowns" in the case when the rational Krylov space becomes (nearly) A-invariant. It should be possible to robustify RKFIT as it was done for linearized least squares and Padé approximation in [21, 20], where close-to-zero singular values lead to a reduction of the approximation degree. In an upcoming work [3] we demonstrate how RKFIT can be used to numerically solve some pole optimization problems associated with

matrix function approximation, similar to those solved analytically in [14, 18, 35]. Following the idea developed in [4], our approach is based on a surrogate diagonal matrix D having similar spectral properties as A but being much cheaper to invert in the pre-computation of the poles.

REFERENCES

[1] B. Beckermann and L. Reichel, *Error estimation and evaluation of matrix functions via the Faber transform*, SIAM J. Numer. Anal., 47 (2009), pp. 3849–3883.

[2] M. Berljafa and S. Güttel, *A Rational Krylov Toolbox for MATLAB*, MIMS EPrint 2014.56, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, 2014. The toolbox is available for download at `http://guettel.com/rktoolbox/`.

[3] M. Berljafa and S. Güttel, *The RKFIT algorithm for nonlinear rational approximation*, tech. report, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, 2015. In preparation.

[4] R.-U. Börner, O. G. Ernst, and S. Güttel, *Three-dimensional transient electromagnetic modeling using rational Krylov methods*, MIMS EPrint 2014.36, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, 2014.

[5] Z. Bujanović and Z. Drmač, *A new framework for implicit restarting of the Krylov–Schur algorithm*, Numer. Linear Algebra Appl., 22 (2015), pp. 220–232.

[6] A. Bultheel, P. Gonzalez-Vera, E. Hendriksen, and O. Njastad, *Orthogonal Rational Functions*, Cambridge University Press, Cambridge, UK, 1999.

[7] A. Bultheel, M. Van Barel, and P. Van gucht, *Orthogonal basis functions in discrete least-squares rational approximation*, J. Comput. Appl. Math., 164-165 (2004), pp. 175–194.

[8] G. De Samblanx and A. Bultheel, *Using implicitly filtered RKS for generalised eigenvalue problems*, J. Comput. Appl. Math., 107 (1999), pp. 195–218.

[9] 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.

[10] K. Deckers and A. Bultheel, *Rational Krylov sequences and orthogonal rational functions*, Tech. Report TW499, Department of Computer Science, KU Leuven, Leuven, Belgium, 2007.

[11] D. Deschrijver, B. Haegeman, and T. Dhaene, *Orthonormal vector fitting: A robust macromodeling tool for rational approximation of frequency domain responses*, IEEE Transactions on Advanced Packaging, 30 (2007), pp. 216–225.

[12] Z. Drmač, S. Gugercin, and C. Beattie, *Quadrature-based vector fitting for discretized $\mathcal{H}_2$ approximation*, SIAM J. Sci. Comput., 37 (2015), pp. A625–A652.

[13] V. Druskin and L. Knizhnerman, *Extended Krylov subspaces: approximation of the matrix square root and related functions*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 755–771.

[14] V. Druskin, L. Knizhnerman, and M. Zaslavsky, *Solution of large scale evolutionary problems using rational Krylov subspaces with optimized shifts*, SIAM J. Sci. Comput., 31 (2009), pp. 3760–3780.

[15] J. Eshof and M. Hochbruck, *Preconditioning Lanczos approximations to the matrix exponential*, SIAM J. Sci. Comput., 27 (2006), pp. 1438–1457.

[16] D. Fasino, *Rational Krylov matrices and QR steps on Hermitian diagonal-plus-semiseparable matrices*, Numer. Linear Algebra Appl., 12 (2005), pp. 743–754.

[17] K. Gallivan, E. Grimme, and P. Van Dooren, *A rational Lanczos algorithm for model reduction*, Numer. Algorithms, 12 (1996), pp. 33–63.

[18] T. Göckler and V. Grimm, *Uniform approximation of φ-functions in exponential integrators by a rational Krylov subspace method with simple poles*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1467–1489.

[19] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, USA, 4th ed., 2013.

[20] P. Gonnet, S. Güttel, and L. N. Trefethen, *Robust Padé approximation via SVD*, SIAM Rev., 55 (2013), pp. 101–117.

[21] P. Gonnet, R. Pachón, and L. N. Trefethen, *Robust rational interpolation and least-squares*, Electron. Trans. Numer. Anal., 38 (2011), pp. 146–167.

[22] S. Gugercin, A. Antoulas, and C. Beattie, *A rational Krylov iteration for optimal H2 model reduction*, in Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems, Kyoto, Japan, 2006, pp. 1665–1667.

[23] B. Gustavsen, *Improving the pole relocating properties of vector fitting*, IEEE Trans. Power

Del., 21 (2006), pp. 1587–1592.
[24] B. Gustavsen, *Comments on "A comparative study of vector fitting and orthonormal vector fitting techniques for EMC applications"*, in Proceedings of the 18th International Zurich Symposium on Electromagnetic Compatibility, Zurich, Switzerland, 2007, pp. 131–134.
[25] B. Gustavsen and A. Semlyen, *Rational approximation of frequency domain responses by vector fitting*, IEEE Trans. Power Del., 14 (1999), pp. 1052–1061.
[26] S. Güttel, *Rational Krylov Methods for Operator Functions*, PhD thesis, Institut für Numerische Mathematik und Optimierung der Technischen Universität Bergakademie Freiberg, Freiberg, Germany, 2010. Available online at `http://nbn-resolving.de/urn:nbn:de:bsz:105-qucosa-27645`.
[27] S. Güttel, *Rational Krylov approximation of matrix functions: Numerical methods and optimal pole selection*, GAMM Mitteilungen, 36 (2013), pp. 8–31.
[28] S. Güttel, R. V. Beeumen, K. Meerbergen, and W. Michiels, *NLEIGS: A class of fully rational Krylov methods for nonlinear eigenvalue problems*, to appear in SIAM J. Sci. Comput., (2014).
[29] C. Jagels and L. Reichel, *Recursion relations for the extended krylov subspace method*, Linear Algebra Appl., 434 (2011), pp. 1716–1732.
[30] E. Jarlebring and H. Voss, *Rational Krylov for nonlinear eigenproblems, an iterative projection method*, Appl. Math., 50 (2005), pp. 543–554.
[31] 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.
[32] D. Kressner, *Block algorithms for reordering standard and generalized Schur forms*, ACM Trans. Math. Software, 32 (2006), pp. 521–532.
[33] T. Mach, M. Pranić, 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.
[34] K. Meerbergen, *Changing poles in the rational Lanczos method for the Hermitian eigenvalue problem*, Numer. Linear Algebra Appl., 8 (2001), pp. 33–52.
[35] I. Moret and P. Novati, *RD-rational approximations of the matrix exponential*, BIT Numerical Mathematics, 44 (2004), pp. 595–615.
[36] M. Pranić and L. Reichel, *Rational Gauss quadrature*, SIAM J. Numer. Anal., 52 (2014), pp. 832–851.
[37] A. Ruhe, *Rational Krylov sequence methods for eigenvalue computation*, Linear Algebra Appl., 58 (1984), pp. 391–405.
[38] A. Ruhe, *The rational Krylov algorithm for nonsymmetric eigenvalue problems. III: Complex shifts for real matrices*, BIT, 34 (1994), pp. 165–176.
[39] A. Ruhe, *Rational Krylov algorithms for nonsymmetric eigenvalue problems. II. Matrix pairs*, Linear Algebra Appl., 198 (1994), pp. 283–295.
[40] A. Ruhe, *Rational Krylov: A practical algorithm for large sparse nonsymmetric matrix pencils*, SIAM J. Sci. Comput., 19 (1998), pp. 1535–1551.
[41] A. Ruhe, *The rational Krylov algorithm for nonlinear matrix eigenvalue problems*, J. Math. Sci. (N. Y.), 114 (2003), pp. 1854–1856.
[42] A. Ruhe and D. Skoogh, *Rational Krylov algorithms for eigenvalue computation and model reduction*, in Applied Parallel Computing Large Scale Scientific and Industrial Problems, B. Kågström, J. Dongarra, E. Elmroth, and J. Waśniewski, eds., vol. 1541 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 1998, pp. 491–502.
[43] G. W. Stewart, *A Krylov–Schur algorithm for large eigenproblems*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 601–614.
[44] G. W. Stewart, *Matrix Algorithms, Volume II: Eigensystems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2001.
[45] G. W. Stewart, *Addendum to "A Krylov–Schur algorithm for large eigenproblems"*, SIAM J. Matrix Anal. Appl., 24 (2002), pp. 599–601.
[46] M. Van Barel, D. Fasino, L. Gemignani, and N. Mastronardi, *Orthogonal rational functions and structured matrices*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 810–829.
[47] R. Van Beeumen, K. Meerbergen, and W. Michiels, *A rational Krylov method based on Hermite interpolation for nonlinear eigenvalue problems*, SIAM J. Sci. Comput., 35 (2013), pp. A327–A350.
[48] R. Vandebril and D. S. Watkins, *A generalization of the multishift QR-algorithm*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 759–779.