SIAM J. MATRIX ANAL. APPL.
Vol. 38, No. 2, pp. 297–321


# KRYLOV–SCHUR-TYPE RESTARTS FOR THE TWO-SIDED ARNOLDI METHOD*

IAN N. ZWAAN† AND MICHIEL E. HOCHSTENBACH†

**Abstract.** We consider the two-sided Arnoldi method and propose a two-sided Krylov–Schur-type restarting method. We discuss the restart for standard Rayleigh–Ritz extraction as well as harmonic Rayleigh–Ritz extraction. Additionally, we provide error bounds for Ritz values and Ritz vectors in the context of oblique projections and present generalizations of, e.g., the Bauer–Fike theorem and Saad's theorem. Applications of the two-sided Krylov–Schur method include the simultaneous computation of left and right eigenvectors and the computation of eigenvalue condition numbers. We demonstrate how the method can be used to find the least sensitive eigenvalues of a nonnormal matrix and how to approximate pseudospectra by using left and right shift-invariant subspaces. The results demonstrate that significant improvements in quality can be obtained over approximations with the (one-sided) Krylov–Schur method.

**Key words.** two-sided Krylov–Schur, Krylov–Schur, two-sided Arnoldi, dual Arnoldi, implicit restart, harmonic two-sided extraction, eigenvalue condition number, pseudospectra, least sensitive eigenvalues

**AMS subject classifications.** 65F15, 65F50, 15A18, 15A23

**DOI.** 10.1137/16M1078987

## 1. Introduction.

The two-sided Lanczos algorithm (cf., e.g., [16, sect. 6.4]) is an important alternative to the Arnoldi method (cf., e.g., [16, sect. 6.2]) for nonnormal matrices. The former uses short three-term recurrences at the expense of double the number of matrix-vector multiplications. But if one wants the eigenvectors, then either the bases must be stored, or they must be computed in a second run. This means that either the storage needed for two-sided Lanczos becomes roughly twice that of Arnoldi, or the number of matrix-vector multiplications doubles again. Moreover, in practice re-biorthogonalization is often necessary because of the loss of biorthogonality in finite precision arithmetic. The accuracy and stability of the computed bases may be improved by using the two-sided Arnoldi method, proposed by Ruhe [14], to replace biorthonormal by orthonormal bases. In this paper, we propose an efficient restarting technique for two-sided Arnoldi, inspired by the Krylov–Schur algorithm [20, 21]. We also investigate perturbation and convergence properties using error bounds for Ritz values and Ritz vectors in the context of oblique projections.

There already are generalizations of the Krylov–Schur method, for example, for Hamiltonian matrices and the product eigenproblem by Kressner [10, 11], as well as a block method for symmetric matrices by Zhou and Saad [27], a version for unitary eigenproblems by David and Watkins [5], and a method for the truncated SVD by Stoll [23]. Jaimoukha and Kasenally [7] present a restarted two-sided Krylov method for model order reduction; however, their method uses projections to remove unstable

*Received by the editors June 7, 2016; accepted for publication (in revised form) by M. Hochbruck December 19, 2016; published electronically April 13, 2017.
http://www.siam.org/journals/simax/38-2/M107898.html
**Funding:** This work was part of the research program "Innovative methods for large matrix problems," which is (partly) financed by the Netherlands Organization for Scientific Research (NWO).
†Department of Mathematics and Computer Science, TU Eindhoven, 5600 MB Eindhoven, The Netherlands (i.n.zwaan@tue.nl, http://ianzwaan.com; M.E.Hochstenbach@tue.nl, http://www.win.tue.nl/~hochsten).

elements without changing the initial vectors. Instead, we are interested in arbitrary exterior eigenvalues of general nonnormal matrices and allow our method to implicitly modify the initial vectors.

Applications that may benefit from two-sided Krylov–Schur include those where the condition number of eigenvalues is important and those where both the left and right eigenvectors are desired. In particular, we use two-sided Krylov–Schur to find eigenvalues with the lowest condition numbers and to approximate pseudospectra. The former may be useful to compute the least sensitive eigenvalues of parameterized matrices, or the most reliable eigenvalues of matrices containing uncertain data. The latter application can provide insight into the (worst-case) behavior of eigenvalues under perturbations. Our contribution is a new type of approximation using two shift-invariant subspaces.

The rest of this paper is organized as follows. First we review Stewart's Krylov–Schur method in section 2. Then we introduce a new two-sided Krylov–Schur method in section 3 and its harmonic counterpart in section 4. Section 5 explores the relation between two-sided Arnoldi and two-sided Lanczos. The focus of section 6 is on perturbation and convergence theory, and that of section 7 on distance properties. Finally, sections 8 and 9 contain the numerical experiments and conclusions.

Throughout this paper $\|\cdot\| = \|\cdot\|_2$ denotes the Euclidean norm, unless stated otherwise, $\|\cdot\|_F$ is the Frobenius norm, and $\sigma_{\min}(M)$ is the smallest singular value of a general matrix $M$.

**2. One-sided Krylov–Schur.** The Krylov–Schur method by Stewart [20, 21] combines the Arnoldi method with a restarting mechanism based on the Schur decomposition. Let $A$ be an $n \times n$ matrix and consider the Krylov subspace

$$\mathcal{V}_\ell = \mathcal{K}_\ell(A, \boldsymbol{v}) = \operatorname{span}\{\boldsymbol{v}, A\boldsymbol{v}, A^2\boldsymbol{v}, \ldots, A^{\ell-1}\boldsymbol{v}\}, \tag{1}$$

where $\ell \ll n$. It is well known that the Arnoldi method creates a basis $V_\ell$ for $\mathcal{V}_\ell$ satisfying the decomposition

$$AV_\ell = V_\ell H_\ell + \boldsymbol{v}_{\ell+1}\boldsymbol{h}_\ell^* = V_{\ell+1}\underline{H}_\ell, \tag{2}$$

where $V_{k+1} = [V_\ell \ \ \boldsymbol{v}_{\ell+1}]$ has orthonormal columns and $\underline{H}_\ell = [H_\ell;\ \boldsymbol{h}_\ell^*]$ is upper-Hessenberg. When $\underline{H}_\ell$ is an arbitrary full-rank $(\ell+1)\times\ell$, it is nevertheless possible to transform the decomposition into the described upper-Hessenberg form [20, Thm. 2.2]. To perform a restart, compute the Schur decomposition

$$H_\ell = QSQ^*,$$

where $Q$ is unitary and $S$ is upper triangular, and define $\widehat{V}_\ell = V_\ell Q$ and $\widehat{\boldsymbol{h}}_\ell = Q^*\boldsymbol{h}_\ell$; then

$$A\widehat{V}_\ell = \widehat{V}_\ell S + \boldsymbol{v}_{\ell+1}\widehat{\boldsymbol{h}}_\ell^*.$$

Partition the above decomposition as

$$A\begin{bmatrix}\widehat{V}_1 & \widehat{V}_2\end{bmatrix} = \begin{bmatrix}\widehat{V}_1 & \widehat{V}_2\end{bmatrix}\begin{bmatrix}S_{11} & S_{12}\\ 0 & S_{22}\end{bmatrix} + \boldsymbol{v}_{\ell+1}\begin{bmatrix}\widehat{\boldsymbol{h}}_1^* & \widehat{\boldsymbol{h}}_2^*\end{bmatrix},$$

where it may be assumed without loss of generality that the desired eigenvalues of $H_\ell$ are along the diagonal of $S_{11}$. Lastly, truncate to obtain

$$A\widehat{V}_1 = \widehat{V}_1 S_{11} + \boldsymbol{v}_{\ell+1}\widehat{\boldsymbol{h}}_1^*.$$

**Algorithm 1.** One-sided Krylov–Schur [20].

**Input:** $A \in \mathbb{C}^{n\times n}$, starting vector $\boldsymbol{v}_1$, minimum and maximum dimensions $m$ and $\ell$, tolerance tol.
**Output:** $V_{m+1}$ and $H_m$ such that $\|AV_m - V_m H_m\| \le \text{tol}$.
1. **for** *number of restarts* **do**
2. Expand the Krylov decomposition to $AV_\ell = V_\ell H_\ell + \boldsymbol{v}_{\ell+1}\boldsymbol{h}_\ell^*$.
3. Compute $H_\ell = QSQ^*$, and partition $Q = [Q_1\, Q_2]$ and $S = \left[\begin{smallmatrix} S_{11} & S_{12} \\ & S_{22} \end{smallmatrix}\right]$.
4. Set $V_m = V_\ell Q_1$, $H_m = S_{11}$, and $\boldsymbol{h}_m = Q_1^*\boldsymbol{h}_\ell$.
5. **if** $\|\boldsymbol{h}_m\| \le \text{tol}$ **then break**
6. **end**

We summarize the one-sided Krylov–Schur method in Algorithm 1.

The Krylov–Schur method extracts approximations to eigenvalues and eigenvectors using the standard Galerkin condition

$$AV_\ell \boldsymbol{c} - \theta V_\ell \boldsymbol{c} \perp \mathcal{V}_\ell.$$

However, it is also possible to extract eigenvalues by choosing a different test subspace $\mathcal{U}_\ell$ and imposing the modified Galerkin condition

$$AV_\ell \boldsymbol{c} - \theta V_\ell \boldsymbol{c} \perp \mathcal{U}_\ell.$$

In this case a Krylov–Schur-type restart is more elaborate [21], but allows, for instance, restarts with harmonic Ritz value extraction. The following two sections show how the one-sided Krylov–Schur restart can be modified to restart either two-sided Arnoldi or harmonic two-sided Arnoldi.

**3. Two-sided Krylov–Schur.** In this section we derive the two-sided Krylov–Schur method. Assume $A$ is a nonnormal $n \times n$ matrix, and consider the right Krylov subspace in (1) together with the left Krylov subspace

$$\mathcal{W}_\ell = \mathcal{K}_\ell(A^*, \boldsymbol{w}) = \operatorname{span}\{\boldsymbol{w}, A^*\boldsymbol{w}, (A^*)^2\boldsymbol{w}, \ldots, (A^*)^{\ell-1}\boldsymbol{w}\}.$$

The two-sided Arnoldi method proposed by Ruhe [14], and later as a block method by Cullum and Zhang [4], independently generates orthonormal bases for the right search space $\mathcal{V}_\ell$ and the left search space $\mathcal{W}_\ell$. This can be done by applying the (one-sided) Arnoldi method twice. Let the generated bases be denoted by $V_\ell$ and $W_\ell$, respectively; then the following relations are satisfied:

$$\begin{aligned} AV_\ell &= V_\ell H_\ell + \boldsymbol{v}_{\ell+1}\boldsymbol{h}_\ell^* = V_{\ell+1}\underline{H}_\ell, \\ A^*W_\ell &= W_\ell K_\ell + \boldsymbol{w}_{\ell+1}\boldsymbol{k}_\ell^* = W_{\ell+1}\underline{K}_\ell, \end{aligned} \tag{3}$$

where both $V_{\ell+1} = [V_\ell\ \boldsymbol{v}_{\ell+1}]$ and $W_{\ell+1} = [W_\ell\ \boldsymbol{w}_{\ell+1}]$ consist of orthonormal columns. The next step is to extract approximate eigenvectors and eigenvalues using the two-sided Rayleigh–Ritz method. For this purpose, the matrices $H_\ell$ and $K_\ell$ are modified to be Rayleigh quotients of $A$ and $A^*$, respectively. The Rayleigh quotient of a matrix $M$ is defined here as $Y^*MX$ for a full column rank matrix $X$ with left inverse $Y^*$ (cf., e.g., [19, p. 252]). Assuming $W_\ell^*V_\ell$ is nonsingular, let

$$\begin{aligned} \tilde{H}_\ell &= H_\ell + (W_\ell^*V_\ell)^{-1}W_\ell^*\boldsymbol{v}_{\ell+1}\boldsymbol{h}_\ell^*, \\ \tilde{K}_\ell &= K_\ell + (V_\ell^*W_\ell)^{-1}V_\ell^*\boldsymbol{w}_{\ell+1}\boldsymbol{k}_\ell^* \end{aligned}$$

and

$$\tilde{\boldsymbol{v}}_{\ell+1} = (I - V_\ell (W_\ell^* V_\ell)^{-1} W_\ell^*) \boldsymbol{v}_{\ell+1},$$
$$\tilde{\boldsymbol{w}}_{\ell+1} = (I - W_\ell (V_\ell^* W_\ell)^{-1} V_\ell^*) \boldsymbol{w}_{\ell+1}.$$

Then it is possible to rewrite (3) as

$$\begin{aligned} AV_\ell &= V_\ell \tilde{H}_\ell + \tilde{\boldsymbol{v}}_{\ell+1} \boldsymbol{h}_\ell^*, \\ A^* W_\ell &= W_\ell \tilde{K}_\ell + \tilde{\boldsymbol{w}}_{\ell+1} \boldsymbol{k}_\ell^*. \end{aligned} \tag{4}$$

Since $\tilde{\boldsymbol{v}}_{\ell+1}$ is orthogonal to $W_\ell$ and $\tilde{\boldsymbol{w}}_{\ell+1}$ is orthogonal to $V_\ell$, it follows that

$$\begin{aligned} \tilde{H}_\ell &= (W_\ell^* V_\ell)^{-1} W_\ell^* A V_\ell, \\ \tilde{K}_\ell &= (V_\ell^* W_\ell)^{-1} V_\ell^* A^* W_\ell. \end{aligned} \tag{5}$$

Because $(W_\ell^* V_\ell)^{-1} W_\ell^*$ and $(V_\ell^* W_\ell)^{-1} V_\ell^*$ are left inverses of $V_\ell$ and $W_\ell$, respectively, we recognize $\tilde{H}_\ell$ and $\tilde{K}_\ell$ as Rayleigh quotients of $A$ and $A^*$. Furthermore, the eigenvalues of $\tilde{H}_\ell$ and of $\tilde{K}_\ell^*$ satisfy the following proposition due to Cullum and Zhang [4].

PROPOSITION 1. *Using the previous definitions, $\tilde{H}_\ell$ and $\tilde{K}_\ell^*$ are similar if $W_\ell^* V_\ell$ is nonsingular.*

*Proof.* Since $(V_\ell^* W_\ell)^{-*} = (W_\ell^* V_\ell)^{-1}$, it is easy to deduce from (5) that

$$(W_\ell^* V_\ell) \tilde{H}_\ell = W_\ell^* A V_\ell = \tilde{K}_\ell^* (W_\ell^* V_\ell). \qquad \square$$

If $W_\ell^* V_\ell$ is singular, then one can perform additional steps of the Krylov process or remove vectors until $W_j^* V_j$ is nonsingular for some $j$.

We are now ready to derive a new two-sided restarting approach inspired by (one-sided) harmonic Krylov–Schur restarts [21]. Consider the Schur decompositions

$$\tilde{H}_\ell = QSQ^* \qquad \text{and} \qquad \tilde{K}_\ell = ZTZ^*,$$

where the eigenvalues of $\tilde{H}_\ell$ and $\tilde{K}_\ell$ are ordered along the diagonals of $S$ and $T$, respectively, and are such that $s_{jj} = t_{jj}^*$. If such a pairing cannot be found due to roundoff errors, then an alternative is to sort $s_{jj}$ and $t_{jj}^*$ independently based on some desirable quantity such as their distance to a target, the size of the real part, etc. Substituting the above Krylov–Schur decompositions in (4) yields

$$\begin{aligned} AV_\ell &= V_\ell QSQ^* + \tilde{\boldsymbol{v}}_{\ell+1} \boldsymbol{h}_\ell^*, \\ A^* W_\ell &= W_\ell ZTZ^* + \tilde{\boldsymbol{w}}_{\ell+1} \boldsymbol{k}_\ell^*. \end{aligned}$$

Let $\widehat{V}_\ell = V_\ell Q$, $\widehat{W}_\ell = W_\ell Z$, $\tilde{\boldsymbol{h}}_\ell = Q^* \boldsymbol{h}_\ell$, and $\tilde{\boldsymbol{k}}_\ell = Z^* \boldsymbol{k}_\ell$, so that

$$\begin{aligned} A\widehat{V}_\ell &= \widehat{V}_\ell S + \tilde{\boldsymbol{v}}_{\ell+1} \tilde{\boldsymbol{h}}_\ell^*, \\ A^* \widehat{W}_\ell &= \widehat{W}_\ell T + \tilde{\boldsymbol{w}}_{\ell+1} \tilde{\boldsymbol{k}}_\ell^*, \end{aligned} \tag{6}$$

and in partitioned form

$$\begin{aligned} A \begin{bmatrix} \widehat{V}_1 & \widehat{V}_2 \end{bmatrix} &= \begin{bmatrix} \widehat{V}_1 & \widehat{V}_2 \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} + \tilde{\boldsymbol{v}}_{\ell+1} \begin{bmatrix} \tilde{\boldsymbol{h}}_1^* & \tilde{\boldsymbol{h}}_2^* \end{bmatrix}, \\ A^* \begin{bmatrix} \widehat{W}_1 & \widehat{W}_2 \end{bmatrix} &= \begin{bmatrix} \widehat{W}_1 & \widehat{W}_2 \end{bmatrix} \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} + \tilde{\boldsymbol{w}}_{\ell+1} \begin{bmatrix} \tilde{\boldsymbol{k}}_1^* & \tilde{\boldsymbol{k}}_2^* \end{bmatrix}. \end{aligned} \tag{7}$$

We can now truncate the partitioned decompositions to

$$\begin{aligned} A\widehat{V}_1 &= \widehat{V}_1 S_{11} + \tilde{\boldsymbol{v}}_{\ell+1}\tilde{\boldsymbol{h}}_1^*, \\ A^*\widehat{W}_1 &= \widehat{W}_1 T_{11} + \tilde{\boldsymbol{w}}_{\ell+1}\tilde{\boldsymbol{k}}_1^*. \end{aligned} \tag{8}$$

The vector $\tilde{\boldsymbol{v}}_{\ell+1}$ is in general not orthogonal to $\widehat{V}_1$, and $\tilde{\boldsymbol{w}}_{\ell+1}$ is not orthogonal to $\widehat{W}_1$. This problem can be remedied by computing

$$\begin{aligned} A\widehat{V}_1 &= \widehat{V}_1 \widehat{H} + \widehat{\boldsymbol{v}}_{\ell+1}\widehat{\boldsymbol{h}}_1^*, \\ A^*\widehat{W}_1 &= \widehat{W}_1 \widehat{K} + \widehat{\boldsymbol{w}}_{\ell+1}\widehat{\boldsymbol{k}}_1^*, \end{aligned} \tag{9}$$

where $[\widehat{V}_1\ \widehat{\boldsymbol{v}}_{\ell+1}]$ and $[\widehat{W}_1\ \widehat{\boldsymbol{w}}_{\ell+1}]$ have orthonormal columns, and

$$\begin{aligned} \widehat{H} &= S_{11} + (\widehat{V}_1^*\tilde{\boldsymbol{v}}_{\ell+1})\tilde{\boldsymbol{h}}_1^*, \\ \widehat{K} &= T_{11} + (\widehat{W}_1^*\tilde{\boldsymbol{w}}_{\ell+1})\tilde{\boldsymbol{k}}_1^*, \\ \widehat{\boldsymbol{v}}_{\ell+1} &= \|(I - \widehat{V}_1\widehat{V}_1^*)\tilde{\boldsymbol{v}}_{\ell+1}\|^{-1}(I - \widehat{V}_1\widehat{V}_1^*)\tilde{\boldsymbol{v}}_{\ell+1}, \\ \widehat{\boldsymbol{w}}_{\ell+1} &= \|(I - \widehat{W}_1\widehat{W}_1^*)\tilde{\boldsymbol{w}}_{\ell+1}\|^{-1}(I - \widehat{W}_1\widehat{W}_1^*)\tilde{\boldsymbol{w}}_{\ell+1}, \\ \widehat{\boldsymbol{h}}_1 &= \|(I - \widehat{V}_1\widehat{V}_1^*)\tilde{\boldsymbol{v}}_{\ell+1}\|\,\tilde{\boldsymbol{h}}_1, \\ \widehat{\boldsymbol{k}}_1 &= \|(I - \widehat{W}_1\widehat{W}_1^*)\tilde{\boldsymbol{w}}_{\ell+1}\|\,\tilde{\boldsymbol{k}}_1. \end{aligned}$$

From here, the search spaces spanned by $\widehat{V}_1$ and $\widehat{W}_1$ can be expanded independently using the (one-sided) Arnoldi method. Below in Algorithm 2 we summarize the two-sided Krylov–Schur method for the computation of approximate right and left invariant subspaces.

Ordinarily, the oblique projections in steps 3 and 4 of Algorithm 2 must be repeated at least once in practice [22, sect. 7], which can be seen as the oblique analogue of reorthogonalization. Step 12 requires extra attention as well, since properly measuring the convergence in two-sided Krylov–Schur is more complex than in one-sided Krylov–Schur. Luckily, we can rely on the work of Kahan, Parlett, and Jiang [9], who investigate the convergence of two-sided Lanczos and derive a set of termination criteria. We describe some of their results below.

For two unit vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ with $\boldsymbol{w}^*\boldsymbol{v} \neq 0$, define the two-sided Rayleigh quotient

$$\rho = \rho(\boldsymbol{v}, \boldsymbol{w}^*) = \frac{\boldsymbol{w}^* A \boldsymbol{v}}{\boldsymbol{w}^*\boldsymbol{v}}$$

and the right and left residuals

$$\boldsymbol{r} = (A - \rho I)\boldsymbol{v} \qquad \text{and} \qquad \boldsymbol{s} = (A - \rho I)^*\boldsymbol{w};$$

then the partial derivatives of $\rho$ are

$$\partial_{\boldsymbol{v}}\rho(\boldsymbol{v}, \boldsymbol{w}^*) = \frac{\boldsymbol{s}^*}{\boldsymbol{w}^*\boldsymbol{v}} \qquad \text{and} \qquad \partial_{\boldsymbol{w}^*}\rho(\boldsymbol{v}, \boldsymbol{w}^*) = \frac{\boldsymbol{r}}{\boldsymbol{w}^*\boldsymbol{v}}.$$

Consequently, $\rho$ should not be used as an approximate eigenvalue unless the value of $\max\{\|\boldsymbol{s}\|, \|\boldsymbol{r}\|\}/|\boldsymbol{w}^*\boldsymbol{v}|$ is sufficiently small relative to $\rho$. An additional result shows that for an eigenvalue $\lambda$ near $\rho$, the bound

$$|\lambda - \rho| \leq \kappa(\lambda)\|E\| + \mathcal{O}(\|E\|^2)$$

**Algorithm 2.** Two-sided Krylov–Schur.

**Input:** Nonnormal $A \in \mathbb{C}^{n\times n}$, starting vectors $\boldsymbol{v}_1$ and $\boldsymbol{w}_1$, minimum and maximum dimensions $m$ and $\ell$.

**Output:** $V_{m+1}$, $W_{m+1}$, $\underline{H}_m$, and $\underline{K}_m$ such that $AV_m = V_{m+1}\underline{H}_m \approx V_m H_m$ and $A^*W_m = W_{m+1}\underline{K}_m \approx W_m K_m$.

1. **for** *number of restarts* **do**
2. Expand the Krylov decompositions to

$$AV_\ell = V_\ell H_\ell + \boldsymbol{v}_{\ell+1}\boldsymbol{h}_\ell^*, \qquad A^*W_\ell = W_\ell K_\ell + \boldsymbol{w}_{\ell+1}\boldsymbol{k}_\ell^*,$$

   using the Arnoldi process, and update $M_\ell = W_\ell^* V_\ell$.
3. Compute $H_\ell = H_\ell + M_\ell^{-1} W_\ell^* \boldsymbol{v}_{\ell+1}\boldsymbol{h}_\ell^*$ and $\boldsymbol{v}_{\ell+1} = \boldsymbol{v}_{\ell+1} - V_\ell M_\ell^{-1} W_\ell^* \boldsymbol{v}_{\ell+1}$.
4. Compute $K_\ell = K_\ell + M_\ell^{-*} V_\ell^* \boldsymbol{w}_{\ell+1}\boldsymbol{k}_\ell^*$ and $\boldsymbol{w}_{\ell+1} = \boldsymbol{w}_{\ell+1} - W_\ell M_\ell^{-*} V_\ell^* \boldsymbol{w}_{\ell+1}$.
5. Compute the Schur decompositions $H_\ell = QSQ^*$ and $K_\ell = ZTZ^*$.
6. Partition $Q$, $S$, $Z$, and $T$ as in (7).
7. Set $V_m = V_\ell Q_1$, $H_m = S_{11}$, $\boldsymbol{h}_m = Q_1^* \boldsymbol{b}_\ell$.
8. Set $W_m = W_\ell Z_1$, $K_m = T_{11}$, $\boldsymbol{k}_m = Z_1^* \boldsymbol{c}_\ell$.
9. Set $M_m = Z_1^* M_\ell Q_1$.
10. Set $H_m = H_m + (V_m^* \boldsymbol{v}_{\ell+1})\boldsymbol{h}_m^*$, $\boldsymbol{v}_{m+1} = (I - V_m V_m^*)\boldsymbol{v}_{\ell+1}$,
    $\boldsymbol{h}_m = \|\boldsymbol{v}_{m+1}\|\boldsymbol{h}_m$, $\boldsymbol{v}_{m+1} = \boldsymbol{v}_{m+1}/\|\boldsymbol{v}_{m+1}\|$.
11. Set $K_m = K_m + (W_m^* \boldsymbol{w}_{\ell+1})\boldsymbol{k}_m^*$, $\boldsymbol{w}_{m+1} = (I - W_m W_m^*)\boldsymbol{w}_{\ell+1}$,
    $\boldsymbol{k}_m = \|\boldsymbol{w}_{m+1}\|\boldsymbol{k}_m$, $\boldsymbol{w}_{m+1} = \boldsymbol{w}_{m+1}/\|\boldsymbol{w}_{m+1}\|$.
12. **if** *converged (cf.* (10)*)* **then break**
13. **end**

holds [9, sect. 5], where $\|E\| \le \max\{\|\boldsymbol{r}\|, \|\boldsymbol{s}\|\}$ and $\kappa(\lambda)$ is the condition number of $\lambda$. While $\kappa(\lambda)$ is unknown in practice, it can be approximated with $|\boldsymbol{w}^*\boldsymbol{v}|^{-1}$; see, e.g., Theorem 3 and [9, sect. 8].

In the context of two-sided Krylov–Schur we compute

$$\tilde{H}_\ell C = C\Theta \qquad (\text{with } \Theta = \operatorname{diag}(\theta_1, \ldots, \theta_\ell) \text{ and } \|\boldsymbol{c}_j\| = 1),$$
$$\tilde{K}_\ell D = D\Gamma \qquad (\text{with } \Gamma = \operatorname{diag}(\gamma_1, \ldots, \gamma_\ell) \text{ and } \|\boldsymbol{d}_j\| = 1),$$

where $\Theta = \Gamma^*$ in exact arithmetic, and let the right and left Ritz vectors be $\boldsymbol{v}_j = V_\ell \boldsymbol{c}_j$ and $\boldsymbol{w}_j = W_\ell \boldsymbol{d}_j$. Then the Rayleigh quotients $\rho_j = \rho(\boldsymbol{v}_j, \boldsymbol{w}_j)$ can be shown to equal the Ritz values $\theta_j = \bar{\gamma}_j$, so that the residuals satisfy

$$\boldsymbol{r}_j = \|(A - \rho_j I)\boldsymbol{v}_j\| = \|(A - \theta_j I)\boldsymbol{v}_j\| = \|\tilde{\boldsymbol{v}}_{\ell+1}\| \, |\boldsymbol{h}_\ell^* \boldsymbol{c}_j|,$$
$$\boldsymbol{s}_j = \|(A - \rho_j I)^* \boldsymbol{w}_j\| = \|(A - \bar{\gamma}_j I)^* \boldsymbol{w}_j\| = \|\tilde{\boldsymbol{w}}_{\ell+1}\| \, |\boldsymbol{k}_\ell^* \boldsymbol{d}_j|.$$

Using the sensitivities $\kappa_j = |\boldsymbol{w}_j^* \boldsymbol{v}_j|^{-1}$, we terminate, for example, if the relative error

$$\frac{\kappa_j}{|\rho_j|} \max\{\|\boldsymbol{r}_j\|, \|\boldsymbol{s}_j\|\} \tag{10}$$

is sufficiently small for the desired value(s) of $\rho_j$. In our tests using finite precision arithmetic, it was advantageous to use the right eigenvectors of both $\tilde{H}_\ell$ and $\tilde{K}_\ell$ instead of using the left and right eigenvectors of only one of the two. In some cases it may also be numerically preferable to use the Rayleigh quotients $\rho_j$ in place of the Ritz values $\theta_j$ and $\gamma_j$ [17, sect. 4].

In this section we have discussed a two-sided version of the Krylov–Schur algorithm in addition to a suitable stopping criterion. In the subsequent section we focus on a two-sided Krylov–Schur restart with harmonic eigenvalue extraction.

**4. Harmonic two-sided Krylov–Schur.** The eigenvalue extraction from the previous section corresponds to imposing the Galerkin conditions

$$\begin{aligned} AV_\ell \boldsymbol{c} - \theta V_\ell \boldsymbol{c} &\perp \mathcal{W}_\ell, \\ A^* W_\ell \boldsymbol{d} - \eta W_\ell \boldsymbol{d} &\perp \mathcal{V}_\ell. \end{aligned} \tag{11}$$

Suppose one is interested in interior eigenvalues near a target $\tau$ not equal to an eigenvalue. These eigenvalues are exterior eigenvalues of the shifted and inverted matrix $(A-\tau I)^{-1}$; hence, it makes sense for the extraction to impose the Petrov–Galerkin conditions

$$\begin{aligned} (A-\tau I)^{-1}\boldsymbol{v} - (\theta-\tau)^{-1}\boldsymbol{v} &\perp \mathcal{U}_1, \\ (A-\tau I)^{-*}\boldsymbol{w} - (\eta-\tau)^{-*}\boldsymbol{w} &\perp \mathcal{U}_2 \end{aligned}$$

for certain test spaces $\mathcal{U}_1$ and $\mathcal{U}_2$; see also [6, sect. 3.2]. It is straightforward to show that the choice $\boldsymbol{v} = V_\ell \boldsymbol{c}$, $\boldsymbol{w} = W_\ell \boldsymbol{d}$, $\mathcal{U}_1 = (A-\tau I)^* \mathcal{W}_\ell$, and $\mathcal{U}_2 = (A-\tau I)\mathcal{V}_\ell$ is equivalent to (11). For harmonic two-sided Rayleigh–Ritz one can take the test spaces

$$\begin{aligned} \mathcal{U}_1 &= (A-\tau I)^*(A-\tau I)^* \mathcal{W}_\ell, \\ \mathcal{U}_2 &= (A-\tau I)(A-\tau I)\mathcal{V}_\ell \end{aligned}$$

to obtain the equivalent conditions

$$\begin{aligned} (A-\theta I)\boldsymbol{v} &\perp (A-\tau I)^* \mathcal{W}_\ell, \\ (A-\eta I)^* \boldsymbol{w} &\perp (A-\tau I)\mathcal{V}_\ell \end{aligned}$$

after some manipulation. The former conditions lead to the generalized eigenvalue problems

$$\begin{aligned} W_\ell^*(A-\tau I)AV_\ell \boldsymbol{c} &= \theta W_\ell^*(A-\tau I)V_\ell \boldsymbol{c}, \\ V_\ell^*(A-\tau I)^* A^* W_\ell \boldsymbol{d} &= \eta V_\ell^*(A-\tau I)^* W_\ell \boldsymbol{d}. \end{aligned}$$

Since these are two conjugated generalized eigenvalue problems, it follows that they are satisfied by $\ell$ quadruples $(\theta, \eta, \boldsymbol{c}, \boldsymbol{d})$ with $\eta = \bar{\theta}$. If $W_\ell^*(A-\tau I)V_\ell$ is nonsingular, we receive the equivalent eigenvalue problems

$$\begin{aligned} (W_\ell^*(A-\tau I)V_\ell)^{-1} W_\ell^*(A-\tau I)AV_\ell \boldsymbol{c} &= \theta \boldsymbol{c}, \\ (V_\ell^*(A-\tau I)^* W_\ell)^{-1} V_\ell^*(A-\tau I)^* A^* W_\ell \boldsymbol{d} &= \bar{\theta}\boldsymbol{d}. \end{aligned}$$

Substituting the Arnoldi decompositions from (3) produces

$$\tilde{H}_\ell \boldsymbol{c} = \theta \boldsymbol{c} \qquad \text{and} \qquad \tilde{K}_\ell \boldsymbol{d} = \bar{\theta}\boldsymbol{d},$$

where $\tilde{H}_\ell$ and $\tilde{K}_\ell$ are rank-1 updates of $H_\ell$ and $K_\ell$, defined by

$$\begin{aligned} \tilde{H}_\ell &= H_\ell + ((\underline{K}_\ell - \bar{\tau}\underline{I})^* W_{\ell+1}^* V_\ell)^{-1} (\underline{K}_\ell - \bar{\tau}\underline{I})^* W_{\ell+1}^* \boldsymbol{v}_{\ell+1} \boldsymbol{h}_\ell^*, \\ \tilde{K}_\ell &= K_\ell + ((\underline{H}_\ell - \tau\underline{I})^* V_{\ell+1}^* W_\ell)^{-1} (\underline{H}_\ell - \tau\underline{I})^* V_{\ell+1}^* \boldsymbol{w}_{\ell+1} \boldsymbol{k}_\ell^*, \end{aligned}$$

and $\underline{I}$ is the identity matrix with an additional zero bottom row. Next, define

$$\tilde{\boldsymbol{v}}_{\ell+1} = (I - V_\ell((\underline{K}_\ell - \bar{\tau}\underline{I})^* W_{\ell+1}^* V_\ell)^{-1}(\underline{K}_\ell - \bar{\tau}\underline{I})^* W_{\ell+1}^*)\boldsymbol{v}_{\ell+1},$$
$$\tilde{\boldsymbol{w}}_{\ell+1} = (I - W_\ell((\underline{H}_\ell - \tau\underline{I})^* V_{\ell+1}^* W_\ell)^{-1}(\underline{H}_\ell - \tau\underline{I})^* V_{\ell+1}^*)\boldsymbol{w}_{\ell+1},$$

so that

$$AV_\ell = V_\ell \tilde{H}_\ell + \tilde{\boldsymbol{v}}_{\ell+1}\boldsymbol{h}_\ell^*,$$
$$A^* W_\ell = W_\ell \tilde{K}_\ell + \tilde{\boldsymbol{w}}_{\ell+1}\boldsymbol{k}_\ell^*,$$

and $\widetilde{H}_\ell$ and $\widetilde{K}_\ell$ are the Rayleigh quotients

$$\begin{aligned}
\tilde{H}_\ell &= (W_\ell^*(A - \tau I)V_\ell)^{-1} W_\ell^*(A - \tau I)AV_\ell \\
&= ((\underline{K}_\ell - \bar{\tau}\underline{I})^* W_{\ell+1}^* V_\ell)^{-1}(\underline{K}_\ell - \bar{\tau}\underline{I})^* W_{\ell+1}^* AV_\ell, \\
\tilde{K}_\ell &= (V_\ell^*(A - \tau I)^* W_\ell)^{-1} V_\ell^*(A - \tau I)^* A^* W_\ell \\
&= ((\underline{H}_\ell - \tau\underline{I})^* V_{\ell+1}^* W_\ell)^{-1}(\underline{H}_\ell - \tau\underline{I})^* V_{\ell+1}^* A^* W_\ell.
\end{aligned} \tag{12}$$

As in Proposition 1, the eigenvalues of the $\tilde{H}_\ell$ and $\tilde{K}_\ell^*$ from this section coincide.

PROPOSITION 2. *If $W_\ell^*(A - \tau I)V_\ell$ is nonsingular, then $\tilde{H}_\ell$ and $\tilde{K}_\ell^*$ in* (12) *are similar.*

*Proof.* The proof is comparable to the proof of Proposition 1, but with $W_\ell$ replaced by $(A - \tau I)^* W_\ell$. From (12) and $A(A - \tau I) = (A - \tau I)A$, it follows that

$$(W_\ell^*(A - \tau I)V_\ell)\tilde{H}_\ell = W_\ell^*(A - \tau I)AV_\ell = \tilde{K}_\ell^*(W_\ell^*(A - \tau I)^* V_\ell). \qquad \square$$

At this point we can compute Schur decompositions of $\tilde{H}_\ell$ and $\tilde{K}_\ell$ and continue analogously to the previous section. Algorithm 3 summarizes the harmonic two-sided Krylov–Schur method for the determination of approximate right and left invariant subspaces.

Notice that in step 3 of the algorithm we attempt to improve the accuracy by using a QR factorization of $\underline{K}_\ell - \bar{\tau}\underline{I}$, so that we essentially work with the orthonormal basis $W_{\ell+1}Q$ instead of $W_{\ell+1}(\underline{K} - \bar{\tau}\underline{I})$. The approach of step 4 is comparable, and $M_{\ell,\ell+1}$ and $M_{\ell+1,\ell}$ denote the $\ell \times (\ell+1)$ and $(\ell+1) \times \ell$ leading principal submatrices of $M_{\ell+1}$, respectively. In step 14 the same stopping conditions from section 3 can be used; however, in this case, using the Rayleigh quotients $\rho_j$ in place of the Ritz values $\theta_j$ and $\gamma_j$ is recommended (cf. [17, sect. 4]).

We have now seen the regular and harmonic two-sided Krylov–Schur algorithms. In the following section we discuss the relation between these two algorithms and the two-sided Lanczos algorithm.

**5. Relation with two-sided Lanczos.** As discussed in the introduction, the two-sided Lanczos method and two-sided Arnoldi are closely related. Specifically, if (3) is in upper-Hessenberg form with $\boldsymbol{h}_\ell = \|\boldsymbol{h}_\ell\|\boldsymbol{e}_\ell$ and $\boldsymbol{k}_\ell = \|\boldsymbol{k}_\ell\|\boldsymbol{e}_\ell$, and $W_\ell^* V_\ell$ is nonsingular, then it can be verified that $\tilde{H}$ and $\tilde{K}$ in (4) are also upper-Hessenberg. Now let $W_\ell^* V_\ell = LU$ be a decomposition into lower and upper triangular factors, and define the biorthonormal bases $\widehat{V}_\ell = VU^{-1}$ and $\widehat{W}_\ell = WL^{-*}$. Furthermore, let $T = U\tilde{H}_\ell U^{-1}$; then from the proof of Proposition 1 it follows that

$$T = U\tilde{H}_\ell U^{-1} = L^{-1}LU\tilde{H}_\ell U^{-1} = L^{-1}\tilde{K}_\ell^* LUU^{-1} = (L^*\tilde{K}L^{-*})^*.$$

**Algorithm 3.** Harmonic two-sided Krylov–Schur.

**Input:** Nonnormal $A \in \mathbb{C}^{n\times n}$, starting vectors $\boldsymbol{v}_1$ and $\boldsymbol{w}_1$, minimum and maximum dimensions $m$ and $\ell$, target $\tau$.

**Output:** $V_{m+1}$, $W_{m+1}$, $\underline{H}_m$, and $\underline{K}_m$ such that $AV_m = V_{m+1}\underline{H}_m \approx V_m H_m$ and $A^*W_m = W_{m+1}\underline{K}_m \approx W_m K_m$.

1. **for** *number of restarts* **do**
2. Expand the Krylov decompositions to

$$AV_\ell = V_\ell H_\ell + \boldsymbol{v}_{\ell+1}\boldsymbol{h}_\ell^*, \qquad A^*W_\ell = W_\ell K_\ell + \boldsymbol{w}_{\ell+1}\boldsymbol{k}_\ell^*,$$

   using the Arnoldi process, and update $M_{\ell+1} = W_{\ell+1}^* V_{\ell+1}$.
3. Compute $QR = \underline{K}_\ell - \bar{\tau}\underline{I}$, and set $\boldsymbol{p} = (Q^* M_{\ell+1,\ell})^{-1} Q^* M_{\ell+1}\boldsymbol{e}_{\ell+1}$.
4. Compute $QR = \underline{H}_\ell - \tau\underline{I}$, and set $\boldsymbol{q} = (Q^* M_{\ell,\ell+1}^*)^{-1} Q^* M_{\ell+1}^*\boldsymbol{e}_{\ell+1}$.
5. Let $H_\ell = H_\ell + \boldsymbol{p}\boldsymbol{h}_\ell^*$ and $\boldsymbol{v}_{\ell+1} = \boldsymbol{v}_{\ell+1} - V_\ell\boldsymbol{p}$.
6. Let $K_\ell = K_\ell + \boldsymbol{q}\boldsymbol{k}_\ell^*$ and $\boldsymbol{w}_{\ell+1} = \boldsymbol{w}_{\ell+1} - W_\ell\boldsymbol{q}$.
7. Compute the Schur decompositions $H_\ell = QSQ^*$ and $K_\ell = ZTZ^*$.
8. Partition $Q$, $S$, $Z$, and $T$ as in (7).
9. Set $V_m = V_\ell Q_1$, $H_m = S_{11}$, $\boldsymbol{h}_m = Q_1^*\boldsymbol{b}_\ell$.
10. Set $W_m = W_\ell Z_1$, $K_m = T_{11}$, $\boldsymbol{k}_m = Z_1^*\boldsymbol{c}_\ell$.
11. Set $M_m = Z_1^* M_\ell Q_1$.
12. Set $H_m = H_m + (V_m^*\boldsymbol{v}_{\ell+1})\boldsymbol{h}_m^*$, $\boldsymbol{v}_{m+1} = (I - V_m V_m^*)\boldsymbol{v}_{\ell+1}$,
    $\boldsymbol{h}_m = \|\boldsymbol{v}_{m+1}\|\boldsymbol{h}_m$, $\boldsymbol{v}_{m+1} = \boldsymbol{v}_{m+1}/\|\boldsymbol{v}_{m+1}\|$.
13. Set $K_m = K_m + (W_m^*\boldsymbol{w}_{\ell+1})\boldsymbol{k}_m^*$, $\boldsymbol{w}_{m+1} = (I - W_m W_m^*)\boldsymbol{w}_{\ell+1}$,
    $\boldsymbol{k}_m = \|\boldsymbol{w}_{m+1}\|\boldsymbol{k}_m$, $\boldsymbol{w}_{m+1} = \boldsymbol{w}_{m+1}/\|\boldsymbol{w}_{m+1}\|$.
14. **if** *converged (see the discussion after Algorithm* 2*)* **then break**
15. **end**

Using this identity, (4) can be written as

$$\begin{aligned} A\widehat{V}_\ell &= \widehat{V}_\ell T + \tilde{\boldsymbol{v}}_{\ell+1}\boldsymbol{h}_\ell^* U^{-1}, \\ A^*\widehat{W}_\ell &= \widehat{W}_\ell T^* + \tilde{\boldsymbol{w}}_{\ell+1}\boldsymbol{k}_\ell^* L^{-*}, \end{aligned} \tag{13}$$

where $T$ is tridiagonal since both $T = U\tilde{H}U^{-1}$ and $T^* = L^*\tilde{K}L^{-*}$ are upper-Hessenberg. The decompositions in (13) coincide with two-sided Lanczos. Assume for harmonic two-sided Arnoldi that $W_\ell^*(A-\tau I)V_\ell$ is nonsingular, let $W_\ell^*(A-\tau I)V_\ell = LU$ be a decomposition into lower and upper triangular factors, and define $\widehat{V}_\ell = V_\ell U^{-1}$ and $\widehat{W}_\ell = (A - \tau I)^* W_\ell L^{-*}$. Then Proposition 2 can be utilized to show that $T = U\tilde{H}U^{-1} = (L^*\tilde{K}L^{-*})^*$ is tridiagonal.

To summarize, two-sided Lanczos and two-sided Arnoldi generate bases for the same subspaces, although two-sided Lanczos uses biorthonormal bases and short recursions, while two-sided Arnoldi uses orthonormal bases and (full) reorthogonalization. The option to use short recursions with two-sided Lanczos makes it a computationally appealing method in situations when the computational cost or the memory requirements for (full) reorthogonalization would be prohibitive. On the other hand, using biorthonormal bases without (full) reorthogonalization may lead to numerical stability issues. Methods that were developed to handle these issues include look-ahead techniques, selective reorthogonalization, and the detection of spurious Ritz values; see, for instance, Ruhe [15, sect. 4.4.4]. However, increases in memory capacity and computational power of computer hardware have diminished the necessity of

such methods, and the modern approach is to favor (full) reorthogonalization when stability and accuracy are crucial. Even though two-sided Lanczos can be implemented with full re-biorthogonalization, Stewart [22] provides convincing reasons for preferring orthogonal bases. For example, orthogonal bases tend to be less sensitive to perturbations than biorthogonal bases. Furthermore, if $X$ and $Y$ are biorthonormal, then the computation of

$$(I - XY^*)\boldsymbol{a}$$

for some vector $\boldsymbol{a}$ may incur a relative error up to $\gamma\|XY^*\|\epsilon$. Here, $\epsilon$ is the machine epsilon, and $\gamma$ is a constant that depends on the accuracy of $X$ and $Y$. The error bound implies that even if re-biorthogonalization is used, accuracy may be lost, especially if $\|XY^*\|$ is large and if errors accumulate.

Two-sided Krylov–Schur uses orthonormal bases and applies only orthonormal transformations to the bases. Some of the accuracy and stability issues are avoided as a result, especially if two-sided Krylov–Schur is implemented with full reorthogonalization. Unfortunately, we are not entirely clear of all stability issues associated with oblique projections, or more specifically, the terms $(W_\ell^* V_\ell)^{-1}$ for standard extraction and $(W_\ell^*(A-\tau I)V_\ell)^{-1}$ for harmonic extraction. The vectors $\tilde{\boldsymbol{v}}_{\ell+1}$ and $\tilde{\boldsymbol{w}}_{\ell+1}$ and the matrices $\tilde{H}_\ell$ and $\tilde{K}_\ell$ will depend on the previous matrix inverses, and therefore the computed Schur decompositions do, too.

As it turns out, it is possible to avoid the explicit use of $(W_\ell^* V_\ell)^{-1}$ and improve the accuracy of the computations in Algorithms 2 and 3. For simplicity we consider only two-sided Rayleigh–Ritz extraction and note that the results can be adapted to two-sided harmonic Ritz. Suppose, for the moment, that we are given the orthonormal matrices $Q$ and $Z$. The objective is to obtain the decompositions in (9) from (6) without using $(W_\ell^* V_\ell)^{-1}$. The update $\widehat{V}_1 = V_\ell Q_1$ can clearly be computed without using a matrix inverse; now

$$\begin{aligned}
A\widehat{V}_1 &= \widehat{V}_1 S_{11} + \tilde{\boldsymbol{v}}_{\ell+1}\tilde{\boldsymbol{h}}_1^* \\
&= \widehat{V}_1 Q_1^* \tilde{H}_\ell Q_1 + (I - V_\ell (W_\ell^* V_\ell)^{-1} W_\ell^*)\boldsymbol{v}_{\ell+1}\boldsymbol{h}_\ell^* Q_1 \\
&= \widehat{V}_1 Q_1^* H_\ell Q_1 + V_\ell Q_1 Q_1^* (W_\ell^* V_\ell)^{-1} W_\ell^* \boldsymbol{v}_{\ell+1}\boldsymbol{h}_\ell^* Q_1 + (I - V_\ell (W_\ell^* V_\ell)^{-1} W_\ell^*)\boldsymbol{v}_{\ell+1}\boldsymbol{h}_\ell^* Q_1 \\
&= \widehat{V}_1 Q_1^* H_\ell Q_1 + (I - V_\ell Q_2 Q_2^* (W_\ell^* V_\ell)^{-1} W_\ell^*)\boldsymbol{v}_{\ell+1}\boldsymbol{h}_\ell^* Q_1.
\end{aligned}$$

It is straightforward to verify that $\widehat{H} = Q_1^* H_\ell Q_1$ and

$$\widehat{\boldsymbol{v}}_{\ell+1}\widehat{\boldsymbol{h}}_1^* = (I - V_\ell Q_2 Q_2^* (W_\ell^* V_\ell)^{-1} W_\ell^*)\boldsymbol{v}_{\ell+1}\boldsymbol{h}_\ell^* Q_1.$$

On the other hand,

$$\begin{aligned}
AV_\ell Q_1 - V_\ell Q_1 Q_1^* H_\ell Q_1 &= V_\ell Q_2 Q_2^* H_\ell Q_1 + \boldsymbol{v}_{\ell+1}\boldsymbol{h}_\ell^* Q_1 \\
&= [V_\ell\, \boldsymbol{v}_{\ell+1}] \begin{bmatrix} Q_2 & \\ & 1 \end{bmatrix} \begin{bmatrix} Q_2^* & \\ & 1 \end{bmatrix} \begin{bmatrix} H_\ell \\ \boldsymbol{h}_\ell^* \end{bmatrix} Q_1.
\end{aligned}$$

It follows that it is possible to determine $\widehat{\boldsymbol{v}}_{\ell+1}\widehat{\boldsymbol{h}}_1^*$ by computing a rank-1 approximation $\boldsymbol{a}\boldsymbol{b}^*$ of

$$\begin{bmatrix} Q_2^* & \\ & 1 \end{bmatrix} \begin{bmatrix} H_\ell \\ \boldsymbol{h}_\ell^* \end{bmatrix} Q_1,$$

with $\|\boldsymbol{a}\| = 1$, and setting $\widehat{\boldsymbol{h}}_1 = \boldsymbol{b}$ and

$$\widehat{\boldsymbol{v}}_{\ell+1} = [V_\ell \, \boldsymbol{v}_{\ell+1}] \begin{bmatrix} Q_2 & \\ & 1 \end{bmatrix} \boldsymbol{a}.$$

An alternative is to use the relation

$$(I - V_\ell Q_2 Q_2^* (W_\ell^* V_\ell)^{-1} W_\ell^*) \boldsymbol{v}_{\ell+1} \|\boldsymbol{h}_\ell^* Q_1\|^2 = (V_\ell Q_2 Q_2^* H_\ell Q_1 + \boldsymbol{v}_{\ell+1} \boldsymbol{h}_\ell^* Q_1) Q_1^* \boldsymbol{h}_\ell$$

to determine $\widehat{\boldsymbol{v}}_{\ell+1}$, which is particularly appealing from a computational point of view. In our tests we found that the latter approach was faster and provided the best numerical performance. The vector $\tilde{\boldsymbol{v}}_{\ell+1}$ is no longer needed with the above approaches, and its computation can be omitted. To summarize, the inverse of $W_\ell^* V_\ell$ can be bypassed once $Q$ is known.

Computing $Q$ without using $(W_\ell^* V_\ell)^{-1}$ is the remaining step. It is possible to avoid the explicit use of the inverse with the QZ decomposition

$$W_\ell^* A V_\ell = P S_\alpha Q^* \qquad \text{and} \qquad W_\ell^* V_\ell = P S_\beta Q^*$$

of the matrix pencil $(W_\ell^* A V_\ell, W_\ell^* V_\ell)$. Here $P$ and $Q$ are orthonormal, $S_\alpha$ and $S_\beta$ are upper triangular, and $S = S_\beta^{-1} S_\alpha$. The QZ decomposition can be reordered if necessary. In our tests we found that the QZ approach did not improve the accuracy with sufficient significance and reliability to justify the increased computational cost.

In this section we have investigated the relation between two-sided Lanczos and two-sided Krylov–Schur and argued how most of the problems with the former are solved by a proper implementation of the latter.

**6. Error bounds for Ritz values and Ritz vectors.** In previous sections we have discussed the computation of Ritz values, Ritz vectors, and their harmonic counterparts. In this section we investigate the convergence of Ritz values and Ritz vectors with respect to the convergence of the search space to an invariant subspace. We will first focus on the convergence of Ritz values and address the convergence of the Ritz vectors later.

To investigate the convergence of a Ritz value $\theta$ to an eigenvalue of $A$, we could invoke, for instance, the Bauer–Fike theorem (cf., e.g., [16, Thm. 3.6]). The Bauer–Fike theorem is a key result in perturbation theory, and below we present a new two-sided version.

THEOREM 3 (two-sided Bauer–Fike). *Suppose that $A$ is diagonalizable such that*

$$A = X \Lambda X^{-1}.$$

*Let $(\theta, \boldsymbol{v}, \boldsymbol{w})$ be an approximate eigentriplet of $A$ with $\|\boldsymbol{v}\| = \|\boldsymbol{w}\| = 1$, and define the residuals*

$$\boldsymbol{r} = A\boldsymbol{v} - \theta \boldsymbol{v} \qquad \text{and} \qquad \boldsymbol{s}^* = \boldsymbol{w}^* A - \theta \boldsymbol{w}^*.$$

*Assume $\boldsymbol{w}^* \boldsymbol{v} \neq 0$ and define $\kappa_\theta = |\boldsymbol{w}^* \boldsymbol{v}|^{-1}$. If the condition number of $X$ is denoted by $\kappa(X)$, then there exists an eigenvalue $\lambda$ of $A$ such that*

$$|\lambda - \theta| \leq \sqrt{\kappa(X) \kappa_\theta \, \|\boldsymbol{r}\| \, \|\boldsymbol{s}\|}.$$

*Proof.* If $\theta$ is an eigenvalue of $A$, the result is clear. Otherwise $A - \theta I$ is nonsingular and

$$|\boldsymbol{w}^* \boldsymbol{v}| = |\boldsymbol{s}^* (A - \theta I)^{-2} \boldsymbol{r}| = |\boldsymbol{s}^* X (\Lambda - \theta I)^{-2} X^{-1} \boldsymbol{r}| \leq \kappa(X) \, \|\boldsymbol{r}\| \, \|\boldsymbol{s}\| \, \|(\Lambda - \theta I)^{-2}\|.$$

Rearranging the terms gives

$$\min_{\mu\in\Lambda(A)} |\mu-\theta|^2 \le \kappa(X)\kappa_\theta \,\|\boldsymbol{r}\|\,\|\boldsymbol{s}\|. \qquad \square$$

In particular, if $\max\{\|\boldsymbol{r}\|,\|\boldsymbol{s}\|\}\to 0$, then $\theta$ converges to some eigenvalue $\lambda$ of $A$, and $\kappa_\theta$ converges to the condition number $\kappa(\lambda)$ of $\lambda$. Theorem 3 can be used with Ritz vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ to match Ritz values to eigenvalues of $A$ one at a time.

An alternate approach for studying the convergence of Ritz values is through Elsner's theorem [19, p. 38].

THEOREM 4 (Elsner's theorem [19]). *Let the eigenvalues of $B$ be $\lambda_1,\ldots,\lambda_n$, and let the eigenvalues of $B+E$ be $\theta_1,\ldots,\theta_n$. Then there is a permutation $j_1,\ldots,j_n$ of the integers $1,\ldots,n$ such that*

$$|\lambda_i-\theta_{j_i}| \le 4(\|B\|+\|B+E\|)^{1-1/n}\|E\|^{1/n} \qquad (i=1,\ldots,n).$$

Hence, if the eigenvalues of $B$ are in the spectrum of $A$ and the eigenvalues of $B+E$ are the computed Ritz values, then $\theta_{j_1},\ldots,\theta_{j_n}$ converge to $\lambda_1,\ldots,\lambda_n$ when $\|E\|\to 0$. An advantage of using Elsner's theorem is that we can match multiple $\theta$'s to eigenvalues simultaneously.

At this point it is helpful to introduce notation that allows the uniform treatment of the remainder of this section. Let

$$V=\begin{bmatrix} V_1 & V_2 & V_3\end{bmatrix} \qquad \text{and} \qquad W=\begin{bmatrix} W_1 & W_2 & W_3\end{bmatrix}$$

be full-rank orthonormal matrices, and introduce the shorthand notation $V_{1,2}$ and $W_{1,2}$ for the first two blocks of $V$ and $W$, respectively. In two-sided Krylov–Schur, the columns of $V_1$ and $W_1$ could, for instance, correspond either to the basis vectors retained after truncation or to a subset thereof. The next step is to make $V$ and $W$ biorthonormal, which is where the following proposition comes into play.

PROPOSITION 5. *If $W_1^*V_1$ and $W_{1,2}^*V_{1,2}$ are nonsingular, then the $3\times 3$ block LU decomposition of $W^*V$ is given by*

$$L=\begin{bmatrix} W_1^*V_1 & & \\ W_2^*V_1 & W_2^*(I-P_1)V_2 & \\ W_3^*V_1 & W_3^*(I-P_1)V_2 & W_3^*(I-P_{1,2})V_3\end{bmatrix},$$

$$U=\begin{bmatrix} I & (W_1^*V_1)^{-1}W_1^*V_2 & (W_1^*V_1)^{-1}W_1^*V_3 \\ & I & (W_2^*(I-P_1)V_2)^{-1}W_2^*(I-P_1)V_3 \\ & & I\end{bmatrix},$$

*where $P_1=V_1(W_1^*V_1)^{-1}W_1^*$ and $P_{1,2}=V_{1,2}(W_{1,2}^*V_{1,2})^{-1}W_{1,2}^*$.*

*Proof.* Suppose for the moment that $W_2^*(I-P_1)V_2$ is nonsingular so that $U$ is well defined. For most of the blocks it is straightforward to verify by direct computation that $LU=W^*V$. The only difficult block is

$$W_3^*V_3 = W_3^*P_1V_3 + W_3^*(I-P_1)V_2(W_2^*(I-P_1)V_2)^{-1}W_2^*(I-P_1)V_3 + W_3^*(I-P_{1,2})V_3.$$

To show that equality holds, it suffices to show that $P_{1,2}=Q$, where $Q$ is the projector defined by

$$Q=P_1+(I-P_1)V_2(W_2^*(I-P_1)V_2)^{-1}W_2^*(I-P_1).$$

From its definition we see that the range of $Q$ must be a subset of the range of $V_{1,2}$, that is, $\mathcal{R}(Q) \subset \mathcal{R}(V_{1,2})$, and likewise $\mathcal{R}(Q^*) \subset \mathcal{R}(W_{1,2})$. Furthermore, notice that $QV_1 = V_1$, $QV_2 = V_2$, $Q^*W_1 = W_1$, and $Q^*W_2 = W_2$. Since a projector is uniquely defined by its column space and its row space, it follows from [22, Thm. 2.2] that $Q = P_{1,2}$.

To prove the ansatz that $W_2^*(I - P_1)V_2$ is nonsingular, let $L_{1,2}$ and $U_{1,2}$ be the upper-left $2 \times 2$ blocks of $L$ and $U$ so that

$$\begin{aligned}\det(W_{1,2}^* V_{1,2}) = \det(L_{1,2}U_{1,2}) &= \det(L_{1,2}) \det(U_{1,2}) \\ &= \det(W_1^* V_1) \det(W_2^*(I - P_1)V_2) \neq 0. \quad \square\end{aligned}$$

Suppose that $W_1^*V_1$ and $W_{1,2}^*V_{1,2}$ are nonsingular and that $L$ and $U$ are given by Proposition 5; then the matrices defined by

$$\tilde{V} = VU^{-1} = \begin{bmatrix}\tilde{V}_1 & \tilde{V}_2 & \tilde{V}_3\end{bmatrix} \qquad \text{and} \qquad \tilde{W} = WL^{-*} = \begin{bmatrix}\tilde{W}_1 & \tilde{W}_2 & \tilde{W}_3\end{bmatrix}$$

are biorthonormal. Furthermore,

$$\tilde{V}_1 = V_1, \qquad \tilde{W}_1^* = (W_1^*V_1)^{-1}W_1^*, \qquad I - V_1(W_1^*V_1)^{-1}W_1^* = I - \tilde{V}_1\tilde{W}_1^*,$$

and

$$I - V_{1,2}(W_{1,2}^*V_{1,2})^{-1}W_{1,2}^* = I - \tilde{V}_{1,2}\tilde{W}_{1,2}^* = \tilde{V}_3\tilde{W}_3^*.$$

Assume that

$$S = (W_{1,2}^*V_{1,2})^{-1}W_{1,2}^*AV_{1,2} \qquad \text{and} \qquad T = (V_{1,2}^*W_{1,2})^{-1}V_{1,2}^*A^*W_{1,2}$$

are upper triangular; then an argument similar to the one at the beginning of section 5 shows that

$$\tilde{W}_{1,2}^*A\tilde{V}_{1,2} = U_{1,2}SU_{1,2}^{-1} = (L_{1,2}^*TL_{1,2}^{-*})^*$$

is block diagonal, with

$$\begin{aligned}\tilde{W}_{1,2}^*A\tilde{V}_{1,2} &= \begin{bmatrix}S_{11} & \\ & S_{22}\end{bmatrix} \\ &= \begin{bmatrix}(W_1^*V_1)^{-1}T_{11}^*(W_1^*V_1) & \\ & (W_2^*(I - P_1)V_2)^{-1}T_{22}^*(W_2^*(I - P_1)V_2\end{bmatrix}.\end{aligned}$$

Finally, we have the following definition and assume for the remainder of this section that $\operatorname{rank}(X) \leq \operatorname{rank}(V_1)$.

Definition 6. *Let $\mathcal{X}$ be an invariant subspace of $A$ such that $A\mathcal{X} \subseteq \mathcal{X}$, and suppose that $[X\, X_\perp]$ is orthonormal, $X$ is a basis of $\mathcal{X}$, and $B$ is such that $AX = XB$. If the spectra of $B$ and $X_\perp^*AX_\perp$ are disjoint, then $(B, X)$ is called a simple orthonormal eigenpair of $A$.*

Using the new notation, we are ready to state a generalization of Jia and Stewart [8, Thm. 4.1], which allows the application of Elsner's theorem to two-sided Arnoldi. The value $\delta$ can be interpreted as a measure of the angle between subspaces and will be analyzed in Theorem 11 and Proposition 13.

THEOREM 7. *Let $(B, X)$ be a simple orthonormal eigenpair of $A$. Define $Z = \tilde{W}_1^* X$, and orthonormalize the columns of $Z$ by setting*

$$\tilde{Z} = ZQ, \qquad \text{where} \qquad Q = (Z^*Z)^{-1/2}.$$

*Then there exists a matrix $E$ satisfying*

$$\|E\| = \|\tilde{W}_1^* A(I - \tilde{V}_1\tilde{W}_1^*)XQ\|$$

*such that $(Q^{-1}BQ, \tilde{Z})$ is an eigenpair of $S_{11} - E$. Furthermore, if $\delta = \|(I - \tilde{V}_1\tilde{W}_1^*)X\| < 1$, then*

$$\|E\| \leq \|\tilde{W}_1^* A\| \frac{\delta}{1-\delta}.$$

*Proof.* For the first part of the proof we multiply $AX = XB$ from the left by $\tilde{W}_1^*$ and obtain

$$\tilde{W}_1^* A(\tilde{V}_1\tilde{W}_1^* + (I - \tilde{V}_1\tilde{W}_1^*))X = \tilde{W}_1^* XB.$$

Since $\tilde{W}_1^* A\tilde{V}_1 = S_{11}$, we can rearrange the terms to get

$$S_{11}\tilde{W}_1^* X - \tilde{W}_1^* XB = -\tilde{W}_1^* A(I - \tilde{V}_1\tilde{W}_1^*)X,$$

which we use to define the residual

$$R = S_{11}\tilde{Z} - \tilde{Z}Q^{-1}BQ = -\tilde{W}_1^* A(I - \tilde{V}_1\tilde{W}_1^*)XQ$$

and the perturbation matrix $E = R\tilde{Z}^*$. Then $S_{11} - E$ satisfies

$$(S_{11} - E)\tilde{Z} = \tilde{Z}Q^{-1}BQ$$

and

$$\|E\| = \|R\| = \|\tilde{W}_1^* A(I - \tilde{V}_1\tilde{W}_1^*)XQ\|,$$

which concludes the first part of the proof. For the second part of the proof we use the relation

$$\|Q\| = \sigma_{\min}^{-1}(Z) = \sigma_{\min}^{-1}(\tilde{V}_1\tilde{W}_1^* X).$$

To compute the smallest singular value of $\tilde{V}_1\tilde{W}_1^* X$, observe that

$$1 \leq \min_{\|\boldsymbol{z}\|=1} (\|\tilde{V}_1\tilde{W}_1^* X\boldsymbol{z}\| + \|(I - \tilde{V}_1\tilde{W}_1^*)X\boldsymbol{z}\|).$$

Therefore,

$$\sigma_{\min}(\tilde{V}_1\tilde{W}_1^* X) = \min_{\|\boldsymbol{z}\|=1} \|\tilde{V}_1\tilde{W}_1^* X\boldsymbol{z}\| \geq 1 - \max_{\|\boldsymbol{z}\|=1} \|(I - \tilde{V}_1\tilde{W}_1^*)X\boldsymbol{z}\| = 1 - \delta,$$

and

$$\|E\| \leq \|\tilde{W}_1^* A\| \, \|(I - \tilde{V}_1\tilde{W}_1^*)\| \, \|Q\| \leq \|\tilde{W}_1^* A\| \frac{\delta}{1-\delta}. \qquad \square$$

The key insight from Theorem 7 is that, under mild conditions, there exists a matrix $E$ such that the eigenvalues of $\tilde{Z}^*(S_{11} - E)\tilde{Z} = Q^{-1}BQ$ are eigenvalues of $A$. By subsequently applying Theorem 4, the following corollary may be obtained.

COROLLARY 8. *Assume* $r = \operatorname{rank}(X) = \operatorname{rank}(V_1)$*, let the eigenvalues of* $B$ *be* $\lambda_1, \ldots, \lambda_r$*, and let the eigenvalues of* $S_{11}$ *be* $\theta_1, \ldots, \theta_r$*. Then there are integers* $j_1, \ldots, j_r$ *such that*

$$|\lambda_i - \theta_{j_i}| \leq 4(2\|(W_1^* V_1)^{-1}\| \, \|A\| + \|E\|)^{1-1/r} \|E\|^{1/r} \qquad (i = 1, \ldots, r).$$

Hence, if $\|(W_1^* V_1)^{-1}\|$ is asymptotically uniformly bounded and $\|E\| \to 0$, then there are Ritz values that converge to eigenvalues of $A$. In practice, the assumption on $(W_1^* V_1)^{-1}$ means that the corollary cannot be applied to defective eigenvalues.

The next proof relates the separation between Ritz values and eigenvalues to the convergence of the subspace $V_1$ to the invariant subspace $X$ of $A$. The proof uses the following definition of the separation operator.

DEFINITION 9. *The separation between an* $n \times n$ *matrix* $N$ *and an* $m \times m$ *matrix* $M$ *is defined by*

$$\operatorname{sep}(N, M) = \min_{\|Z\|=1} \|NZ - ZM\|.$$

For more information on the separation operator, see, for example, [19, p. 256].

THEOREM 10. *Let* $(B, X)$ *be a simple orthonormal eigenpair of* $A$*; then*

$$\operatorname{sep}(\tilde{V}_1 S_{11} \tilde{W}_1^*, B) \leq \frac{\|\tilde{V}_1 \tilde{W}_1^* A \tilde{V}_3 \tilde{W}_3^* X\|}{\|\tilde{V}_1 \tilde{W}_1^* X\|}.$$

*Proof.* Since $AX = XB$, we have that

$$\tilde{W}_{1,2}^* A \tilde{V} \tilde{W}^* X = \tilde{W}_{1,2}^* X B.$$

Rearranging the terms gives

$$\begin{bmatrix} S_{11} & \\ & S_{22} \end{bmatrix} \begin{bmatrix} \tilde{W}_1^* X \\ \tilde{W}_2^* X \end{bmatrix} - \begin{bmatrix} \tilde{W}_1^* X \\ \tilde{W}_2^* X \end{bmatrix} B = - \begin{bmatrix} \tilde{W}_1^* A \tilde{V}_3 \tilde{W}_3^* X \\ \tilde{W}_2^* A \tilde{V}_3 \tilde{W}_3^* X \end{bmatrix}. \tag{14}$$

From the first block row we see that

$$S_{11} \tilde{W}_1^* X - \tilde{W}_1^* X B = -\tilde{W}_1^* A \tilde{V}_3 \tilde{W}_3^* X,$$

and hence

$$(\tilde{V}_1 S_{11} \tilde{W}_1^*) \tilde{V}_1 \tilde{W}_1^* X - \tilde{V}_1 \tilde{W}_1^* X B = -\tilde{V}_1 \tilde{W}_1^* A \tilde{V}_3 \tilde{W}_3^* X.$$

Using the definition of the separation operator, we can now derive the bound

$$\operatorname{sep}(\tilde{V}_1 S_{11} \tilde{W}_1^*, B) \|\tilde{V}_1 \tilde{W}_1^* X\| \leq \|\tilde{V}_1 \tilde{W}_1^* A \tilde{V}_3 \tilde{W}_3^* X\|,$$

which concludes the proof. ∎

Theorem 10 tells us that the separation between $\tilde{V}_1 S_{11} \tilde{W}_1^*$ and $B$ must go to zero as the span of $X$ becomes contained in the span of $V_{1,2}$; this is true in particular if $V_1$ converges to $X$.

It is instructive to determine what can be said of $\|(I - P_1)X\|$ if it is known that $\|(I - P_{1,2})X\| \to 0$. Saad provides a bound in the case of Hermitian matrices; see, for example, [16, Thm. 4.6]. Saad's theorem was generalized by Stewart for general matrices in [18]. In [6, Thm. 3] the theorem was further generalized to a two-sided result, but using $\|[\tilde{W}_2 \ \tilde{W}_3]^* X\|$ instead of $\|(I - P_1)X\|$, and restricted by the assumption that $X$ is a vector. We therefore state a new two-sided Saad-type theorem.

THEOREM 11. *Let $(B, X)$ be a simple orthonormal eigenpair of $A$. If*

$$\operatorname{sep}(\tilde{V}_2 S_{22} \tilde{W}_2^*, B) > 0,$$

*then*

$$\begin{aligned}\|(I - \tilde{V}_1 \tilde{W}_1^*)X\| &\leq \frac{\|\tilde{V}_2 \tilde{W}_2^* A \tilde{V}_3 \tilde{W}_3^* X\|}{\operatorname{sep}(\tilde{V}_2 S_{22} \tilde{W}_2^*, B)} + \|\tilde{V}_3 \tilde{W}_3^* X\| \\ &\leq \left(1 + \frac{\|\tilde{V}_2 \tilde{W}_2^* A\|}{\operatorname{sep}(\tilde{V}_2 S_{22} \tilde{W}_2^*, B)}\right) \|\tilde{V}_3 \tilde{W}_3^* X\|.\end{aligned}$$

*Proof.* From the second block row of (14) it follows that

$$S_{22} \tilde{W}_2^* X - \tilde{W}_2^* X B = -\tilde{W}_2^* A \tilde{V}_3 \tilde{W}_3^* X.$$

Using the fact that $\tilde{V}_2^* \tilde{W}_2 = I$, we can write

$$(\tilde{V}_2 S_{22} \tilde{W}_2^*) \tilde{V}_2 \tilde{W}_2^* X - \tilde{V}_2 \tilde{W}_2^* X B = -\tilde{V}_2 \tilde{W}_2^* A \tilde{V}_3 \tilde{W}_3^* X,$$

so that

$$\operatorname{sep}(\tilde{V}_2 S_{22} \tilde{W}_2^*, B) \|\tilde{V}_2 \tilde{W}_2^* X\| \leq \|\tilde{V}_2 \tilde{W}_2^* A \tilde{V}_3 \tilde{W}_3^* X\|.$$

Therefore, we acquire the bound

$$\|(I - \tilde{V}_1 \tilde{W}_1^*)X\| = \|\tilde{V}_2 \tilde{W}_2^* X + \tilde{V}_3 \tilde{W}_3^* X\| \leq \frac{\|\tilde{V}_2 \tilde{W}_2^* A \tilde{V}_3 \tilde{W}_3^* X\|}{\operatorname{sep}(\tilde{V}_2 S_{22} \tilde{W}_2^*, B)} + \|\tilde{V}_3 \tilde{W}_3^* X\|,$$

which concludes the proof. ∎

If there exists a positive constant $\alpha$ such that

$$\operatorname{sep}(\tilde{V}_2 S_{22} \tilde{W}_{22}^*, B) \geq \alpha > 0$$

as $\|(I - P_{1,2})X\| \to 0$, then the bound

$$\|(I - P_1)X\| \lesssim \left(1 + \frac{\|\tilde{V}_2 \tilde{W}_2^* A\|}{\alpha}\right) \|(I - P_{1,2})X\|$$

is asymptotically satisfied and $\|(I - P_1)X\| \to 0$ when $\|(I - P_{1,2})X\| \to 0$. The intuitive interpretation of the lower bound $\alpha$ is that there must be a gap between the spectra of $B$ and $S_{22}$ as $V_1$ and $S_{11}$ converge.

Applying Theorems 7, 10, and 11 to two-sided Krylov–Schur yields the following bounds.

COROLLARY 12. *Suppose the relations in* (7) *are satisfied with $V_1 = \widehat{V}_1$, $V_2 = \widehat{V}_2$, $W_1 = \widehat{W}_1$, and $W_2 = \widehat{W}_2$; then the bound in Theorem* 7 *can be written as*

$$\|E\| \leq \|P_1\| \, \|\tilde{\boldsymbol{k}}_1\| \, \|(I - P_{1,2})XQ\|,$$

*the bound in Theorem* 10 *as*

$$\operatorname{sep}(\tilde{V}_1 S_{11} \tilde{W}_1^*, B) \leq \|P_1\| \, \|\tilde{\boldsymbol{k}}_1\| \, \frac{\|(I - P_{1,2})X\|}{\|P_1 X\|},$$

*and the bound in Theorem* 11 *as*

$$\|(I - P_1)X\| \leq \left(1 + \frac{\|P_{1,2} - P_1\| \, \|\tilde{\boldsymbol{k}}_2\|}{\operatorname{sep}(\tilde{V}_2 S_{22} \tilde{W}_2^*, B)}\right) \|(I - P_{1,2})X\|.$$

The corollary shows that bounds of the form $\|\tilde{V}_j\tilde{W}_j^*\|\,\|\tilde{\boldsymbol{k}}_j\|$ are obtained instead of $\|\tilde{V}_j\tilde{W}_j^*A\|$ when two-sided Krylov–Schur is used. This is an attractive result since $\|\tilde{\boldsymbol{k}}_1\|$ can be expected to go to zero as $V_1$ converges to $X$.

It is possible to bound the norm of the oblique projections from the present section in terms of more common orthogonal projections; see, for example, the following proposition.

PROPOSITION 13. *Suppose that $X$, $V$, and $W$ have orthonormal columns and that $W^*V$ is nonsingular; then for the 2-norm we have*

$$\|(I-VV^*)X\|_2 \le \|(I-V(W^*V)^{-1}W^*)X\|_2 \le \|(W^*V)^{-1}\|_2\,\|(I-VV^*)X\|_2,$$

*and for the Frobenius norm*

$$\|(I-VV^*)X\|_F \le \|(I-V(W^*V)^{-1}W^*)X\|_F \le \sqrt{1+\|(W^*V)^{-1}\|_F^2}\,\|(I-VV^*)X\|_F.$$

*Proof.* Define $Z=(I-VV^*)X$ and $P=V(W^*V)^{-1}W^*$; then

$$\begin{aligned}\|Z\|_F^2 &= \|(I-VV^*)(I-P)X\|_F^2\\ &= \operatorname{tr}(X^*(I-P)^*(I-VV^*)(I-P)X)\\ &= \|(I-P)X\|_F^2 - \|V^*(I-P)X\|_F^2 \le \|(I-P)X\|_F^2\end{aligned}$$

and

$$\begin{aligned}\|(I-P)X\|_F^2 = \|(I-P)Z\|_F^2 = \|Z-PZ\|_F^2 = \|Z\|_F^2 + \|PZ\|_F^2\\ \le \|Z\|_F^2(1+\|(W^*V)^{-1}\|_F^2).\end{aligned}$$

For the 2-norm we give a simplified and block version of the first part of the proof found in Chaturantabut and Sorensen [2, Lem. 3.2]. For a nontrivial projector $P$ it holds that $\|I-P\|_2=\|P\|_2$; see, for example, Szyld [24]. Therefore

$$\|Z\|_2 = \|(I-VV^*)(I-P)X\|_2 \le \|(I-P)X\|_2$$

and

$$\|(I-P)X\|_2 = \|(I-P)Z\|_2 \le \|I-P\|_2\,\|Z\|_2 = \|P\|_2\,\|Z\|_2 \le \|(W^*V)^{-1}\|_2\,\|Z\|_2. \qquad \square$$

Consequently, if $\|(W^*V)^{-1}\|$ is sufficiently small, then the norms

$$\|(I-V(W^*V)^{-1}W^*)X\| \qquad \text{and} \qquad \|V(W^*V)^{-1}W^*X\|$$

can be seen as a generalization of $\sin\angle(X,V)$ and $\cos\angle(X,V)$, respectively; see Figure 1 for an illustration.

**7. Two-sided distance properties.** In the previous section we have considered the convergence of subspaces to invariant subspaces. The focus of this section is on the minimum distance between a given matrix $A$ and a matrix with given invariant subspaces. Given a subspace $\mathcal{V}$, Noschese and Reichel [12] consider the problem of finding the matrix $M$ closest to $A$ satisfying

$$M\mathcal{V} \subseteq \mathcal{V}. \tag{15}$$

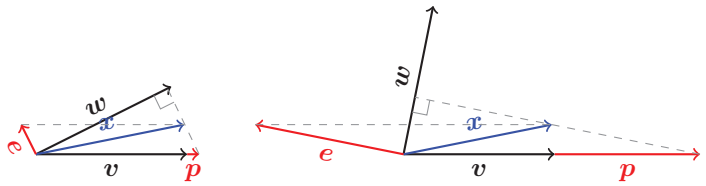


FIG. 1. *Consider the approximation* $\boldsymbol{v}$ *to* $\boldsymbol{x}$*, the projection* $\boldsymbol{p} = \boldsymbol{v}(\boldsymbol{w}^*\boldsymbol{v})^{-1}\boldsymbol{w}^*\boldsymbol{x}$*, and the complementary part* $\boldsymbol{e} = \boldsymbol{x} - \boldsymbol{p}$*. In the left diagram the angle between* $\boldsymbol{v}$ *and* $\boldsymbol{w}$ *is small, and* $\|\boldsymbol{p}\| \approx \|\boldsymbol{v}^*\boldsymbol{x}\| = \cos\angle(\boldsymbol{x},\boldsymbol{v})$ *and* $\|\boldsymbol{e}\| \approx \|(I - \boldsymbol{v}\boldsymbol{v}^*)\boldsymbol{x}\| = \sin\angle(\boldsymbol{x},\boldsymbol{v})$*. In the right diagram the angle between* $\boldsymbol{v}$ *and* $\boldsymbol{w}$ *is large, and* $\|\boldsymbol{p}\|$ *and* $\|\boldsymbol{e}\|$ *are no longer satisfactory approximations to the cosine and sine.*

In the two-sided case we impose the additional constraint

$$M^*\mathcal{W} \subseteq \mathcal{W} \tag{16}$$

for a given subspace $\mathcal{W}$. Alternatively, this is the problem of finding the backward error $E = A - M$, where the norm of $E$ can be seen as a measure for the quality of the subspaces as approximate invariant subspaces. Consider the following well-known theorem [9, Main Thm.].

THEOREM 14 (Kahan, Parlett, and Jiang [9]). *Let* $A$ *be an* $n \times n$ *matrix, and let two* $n \times m$ *matrices* $V$ *and* $W$ *having orthonormal columns be given. Suppose that* $W^*V$ *is nonsingular. Let*

$$R = AV - VC, \qquad S^* = W^*A - DW^*,$$

*where* $C$ *and* $D$ *are Rayleigh quotients*

$$C = (W^*V)^{-1}W^*AV, \qquad D = W^*AV(W^*V)^{-1}.$$

*Then the solution* $E$ *of*

$$(A - E)V = VC \quad \text{and} \quad W^*(A - E) = DW^*$$

*that simultaneously minimizes both*

$$\|E\|_2 = \min_E \|E\|_2 = \max\{\|R\|_2, \|S\|_2\}$$

*and*

$$\|E\|_F = \min_E \|E\|_F = \sqrt{\|R\|_F^2 + \|S\|_F^2}$$

*is given by*

$$E = RV^* + WS^*.$$

Using the theorem we readily find the following result.

COROLLARY 15. *Suppose that* $V$ *and* $W$ *are orthonormal bases of the subspaces* $\mathcal{V}$ *and* $\mathcal{W}$*, respectively, and that* $W^*V$ *is nonsingular. Then, the matrix* $M$ *closest to* $A$ *that satisfies*

$$M\mathcal{V} \subseteq \mathcal{V} \qquad \text{and} \qquad M^*\mathcal{W} \subseteq \mathcal{W}$$

*is given by*

$$M = A - (I - V(W^*V)^{-1}W^*)AVV^* + WW^*A(I - V(W^*V)^{-1}W^*).$$

*Furthermore, if two-sided Arnoldi is used to compute $V$, $W$, $\tilde{H}$, and $\tilde{K}$ so that*

$$R = AV - V\tilde{H} = \tilde{\boldsymbol{v}}\tilde{\boldsymbol{h}}^* \qquad \textit{and} \qquad S = A^*W - W\tilde{K} = \tilde{\boldsymbol{w}}\tilde{\boldsymbol{k}}^*,$$

*where $\tilde{\boldsymbol{v}}$, $\tilde{\boldsymbol{w}}$, $\tilde{\boldsymbol{h}}$, and $\tilde{\boldsymbol{k}}$ are as in* (8)*, then*

$$\|E\|_2 = \max\{\|\tilde{\boldsymbol{v}}\|\,\|\tilde{\boldsymbol{h}}\|, \|\tilde{\boldsymbol{w}}\|\,\|\tilde{\boldsymbol{k}}\|\} \qquad \textit{and} \qquad \|E\|_F = \sqrt{\|\tilde{\boldsymbol{v}}\|^2\|\tilde{\boldsymbol{h}}\|^2 + \|\tilde{\boldsymbol{w}}\|^2\|\tilde{\boldsymbol{k}}\|^2}$$

*for $E = A - M$.*

The matrix $M$ from Corollary 15 satisfies the additional constraint

$$W^*(A - zI)V = W^*(M - zI)V$$

for all scalars $z$. This kind of shift-invariance allows us to interpret $\|E\|_2$ and $\|E\|_F$ as a backward error for the approximation of pseudospectra in section 8.2, where we compute

$$\sigma_{\min}(W^*(A - zI)V)$$

for a large number of complex shifts $z$ near a region of interest. The matrix $M$ is in general not of low rank, and instead we might be interested in the two-sided Arnoldi approximation

$$A_m = V(W^*V)^{-1}W^*AV(W^*V)^{-1}W^* = V\tilde{H}(W^*V)^{-1}W = V(W^*V)^{-1}\tilde{K}^*W^*,$$

which is the unique rank-$m = \operatorname{rank}(W^*AV)$ matrix satisfying

$$A_m\mathcal{V} \subseteq \mathcal{V}, \qquad A_m^*\mathcal{W} \subseteq \mathcal{W}, \qquad \text{and} \qquad W^*AV = W^*A_mV.$$

An alternative for the singular value problem is to consider the problem of finding the matrix $N$ closest to $A$ satisfying

$$N\mathcal{V} \subseteq \mathcal{W} \qquad \text{and} \qquad N^*\mathcal{W} \subseteq \mathcal{V},$$

as opposed to $M$ satisfying (15) and (16). Noschese and Reichel [12, sect. 3] show that

$$N = (I - WW^*)A(I - VV^*) + WW^*AVV^*$$

minimizes the distance to $A$ with

$$\|A - N\|_F^2 = \|AV\|_F^2 + \|A^*W\|_F^2 - 2\|W^*AV\|_F^2.$$

As before, $N$ satisfies the additional constraint

$$W^*(A - zI)V = W^*(N - zI)V$$

for any scalar $z$, making $\|A - N\|_F$ another backward error. The unique rank-$m$ approximation $B_m$ satisfying

$$B_m\mathcal{V} \subseteq \mathcal{W}, \qquad B_m^*\mathcal{W} \subseteq \mathcal{V}, \qquad \text{and} \qquad W^*AV = W^*B_mV$$

is given by the two-sided Arnoldi approximation

$$B_m = WW^*AVV^* = W(W^*V)\tilde{H}V^* = W\tilde{K}^*(W^*V)V^*.$$

The proposition below gives the distance between $A$ and $B_m$; the proof closely follows the arguments in [12, Prop. 3.3] but uses general left-orthonormal $V$ and $W$.

PROPOSITION 16 (generalization of [12, Prop. 3.3]). *Let $V$ and $W$ have orthonormal columns, and define the matrix $B_m = WW^*AVV^*$; then*

$$\|A - B_m\|_F^2 = \|A\|_F^2 - \|B_m\|_F^2.$$

*Proof.* Using the cyclic property of the trace, we obtain

$$\begin{aligned}\|A - WW^*AVV^*\|_F^2 &= \operatorname{tr}(A^*A - A^*WW^*AVV^* - VV^*A^*WW^*A \\ &\qquad + VV^*A^*WW^*AVV^*) \\ &= \operatorname{tr}(A^*A) - \operatorname{tr}(V^*A^*WW^*AV) - \operatorname{tr}(V^*A^*WW^*AV) \\ &\quad + \operatorname{tr}(V^*A^*WW^*AV) \\ &= \|A\|_F^2 - \|B_m\|_F^2. \end{aligned}$$

∎

We have given a two-sided analogue of Noschese and Reichel's result for (right) invariant subspaces. The bounds in Corollary 15 are particularly elegant and efficient to compute in the context of two-sided Krylov–Schur. Distances in the case of invariant singular subspaces are efficiently computable as well, assuming that the Frobenius norm is used and $A$ is explicitly available. To summarize, the distance properties from this section can be used to gain insight into the quality of approximate invariant subspaces.

## 8. Applications and numerical experiments.

**8.1. Eigenvalue condition numbers.** Suppose we wish to compute the best-conditioned eigenvalues of a nonnormal matrix $A$, which is effectively the opposite of the goal of the sensitive pole algorithm [13]. For instance, $A$ might be constructed from uncertain data, making the best-conditioned eigenvalues the most reliable ones. Alternatively, one may be focused on the least sensitive eigenvalues of some $A = A(\boldsymbol{p}_0)$ obtained from a parameterized problem for a specific set of parameters given by $\boldsymbol{p}_0$. Since the eigenvalue condition numbers are essential quantities, the choice of a two-sided method over a one-sided method may be appropriate.

In Table 1 we compare one-sided and two-sided Krylov–Schur for the computation of the best-conditioned eigenvalues. That is, we are looking for an approximation $\theta$ to an eigenvalue $\lambda$, and an approximation $\kappa_\theta$ to $\kappa(\lambda)$, where $\kappa(\lambda)$ is as small as possible. We recognize that it may be more useful in practice to restrict the search to the best-conditioned eigenvalue near a target, but we make no such restriction here for the sake of simplicity. We measure the relative errors

$$\operatorname{err}_\lambda = \left|\frac{\lambda - \theta}{\lambda}\right| \qquad \text{and} \qquad \operatorname{err}_{\kappa(\lambda)} = \left|\frac{\kappa(\lambda) - \kappa_\theta}{\kappa(\lambda)}\right|,$$

as well as the number of matrix-vector products executed before the algorithms are terminated. We use (10) as a stopping criterion and terminate one-sided and two-sided Krylov–Schur when

$$\frac{\|A\boldsymbol{v} - \theta\boldsymbol{v}\|}{|\theta|} \leq \epsilon 2^{10} \qquad \text{and} \qquad \frac{\max\{\|A\boldsymbol{v} - \theta\boldsymbol{v}\|, \|\boldsymbol{w}^*A - \theta\boldsymbol{w}^*\|\}}{|\theta\boldsymbol{w}^*\boldsymbol{v}|} \leq \epsilon 2^{10},$$

respectively, where the $\theta$'s are Ritz values, $\boldsymbol{v}$ and $\boldsymbol{w}$ are right and left Ritz vectors with unit norm, and $\epsilon$ is the machine epsilon. For example, $\epsilon \approx 2.22 \cdot 10^{-16}$ and $\epsilon 2^{10} \approx 2.27 \cdot 10^{-13}$ for IEEE double precision floating point numbers. We run the algorithms with minimum subspace dimension $m = 25$ and maximum subspace dimension $\ell = 50$

TABLE 1
*Median results over* 1000 *runs with different random initial vectors for computing the best-conditioned eigenvalues of nonnormal matrices with one-sided and two-sided Krylov–Schur.*

| | | | One-sided | | | Two-sided | | |
|---|---|---|---|---|---|---|---|---|
| Name | $n$ | $\kappa(\lambda)$ | $\mathrm{err}_\lambda$ | $\mathrm{err}_{\kappa(\lambda)}$ | MVs | $\mathrm{err}_\lambda$ | $\mathrm{err}_{\kappa(\lambda)}$ | MVs |
| randn* | 1024 | 3.34 | 1.25e + 00 | 9.82e − 01 | 800 | 1.01e − 14 | 3.97e − 14 | 1100 |
| bfw782a* | 782 | 1.00 | 4.00e − 02 | 9.96e − 01 | 100 | 3.55e − 15 | 3.09e − 14 | 200 |
| ck656 | 656 | 1.02 | 9.75e − 01 | 8.70e − 01 | 1275 | 4.84e − 16 | 3.35e − 03 | 50 |
| pde900 | 900 | 4.04 | 2.35e − 01 | 1.00e + 00 | 1575 | 2.67e − 15 | 1.89e − 14 | 125 |
| rdb1250l | 1250 | 1.05 | 9.55e − 01 | 8.51e − 01 | 400 | 7.20e − 15 | 3.69e − 15 | 150 |
| olm1000 | 1000 | 1.00 | 1.00e + 00 | 9.94e − 01 | 3300 | 2.99e − 14 | 2.94e − 14 | 7525 |
| qh1484 | 1484 | 1.00 | 1.00e + 00 | 8.77e − 01 | 825 | 2.87e − 04 | 4.27e − 10 | 75 |
| rdb1250 | 1250 | 1.01 | 9.88e − 01 | 8.41e − 01 | 350 | 5.32e − 15 | 2.65e − 15 | 150 |
| qc2534 | 2534 | 1.01 | 1.41e + 00 | 1.00e + 00 | 10300 | 7.36e − 15 | 2.42e − 15 | 75 |
| af23560* | 23560 | 1.10 | 9.37e − 01 | 9.96e − 01 | 2700 | 4.98e − 14 | 6.80e − 07 | 350 |

by default, and with $m = 50$ and $\ell = 100$ for problems marked by an asterisk (*). All the matrices, except randn, are from a test matrix collection of non-Hermitian eigenvalue problems [1] and are balanced first [3]. The matrix randn is generated using the identically named MATLAB function, and we use the same function to generate random starting vectors.

The results in Table 1 show that two-sided Krylov–Schur computes more accurate approximations to both $\lambda$ and $\kappa(\lambda)$ in every case, and does so using fewer matrix-vector products in seven out of 10 cases. In particular, the total number of matrix-vector products used by one-sided Krylov–Schur is 21625, versus 9800 used by two-sided Krylov–Schur. The high relative error of the one-sided approximations can be explained by the fact that one-sided Krylov–Schur converges to incorrect eigenvalues, a problem not shared by its two-sided counterpart. Evidently, two-sided Krylov–Schur benefits from the improved accuracy of the two-sided condition number estimates and the two-sided Rayleigh quotient.

**8.2. Pseudospectra.** When studying nonnormal matrices, computing pseudospectra rather than eigenvalues and condition numbers may be more insightful [25]. In particular, pseudospectra provide more detailed information regarding the behavior of the eigenvalues under matrix perturbations in the nonnormal case. Indeed, one possible definition of the $\varepsilon$-pseudospectrum of $A$ that clearly shows its relation with matrix perturbations is

$$\Lambda_\varepsilon(A) = \{z \in \mathbb{C} \colon z \in \Lambda(A + E) \text{ for some } E \text{ with } \|E\| < \varepsilon\},$$

where $\Lambda(A + E)$ denotes the spectrum of $A + E$. An alternate definition that is more fitting for the computation of pseudospectra is

$$\Lambda_\varepsilon(A) = \{z \in \mathbb{C} \colon \sigma_{\min}(A - zI) < \varepsilon\}.$$

Ergo, one can simply compute $\sigma_{\min}(A - zI)$ for $z \in \mathbb{C}$ and plot $\varepsilon$-level curves; unfortunately, doing so for many grid points and large $A$ is generally time- and memory-consuming. One method to improve performance is to use one-sided Krylov–Schur to obtain

$$AV_m = V_{m+1}\underline{H}_m,$$

with orthonormal $V_{m+1}$, and compute the approximation

$$\sigma_{\min}(A - zI) \approx \sigma_{\min}((A - zI)V_m) = \sigma_{\min}(V_{m+1}^*(A - zI)V_m) = \sigma_{\min}(\underline{H}_m - z\underline{I}); \tag{17}$$

see Wright and Trefethen [26]. Since the right and left singular subspaces differ for nonnormal matrices, it seems natural to project onto a subspace distinct from $\mathcal{V}_{m+1} = \operatorname{span}\{V_{m+1}\}$. At the same time, a shift-invariant subspace is ideal if the goal is to reduce computational effort. This suggests that the left Krylov subspace $\mathcal{W}_m$ belonging to $A^*$ may be an excellent choice, especially if $\mathcal{W}_m$ approximates an invariant subspace belonging to the eigenvalues of interest. Hence, by using two-sided Krylov–Schur we can compute the approximation

$$\sigma_{\min}(A - zI) \approx \sigma_{\min}(W_m^*(A - zI)V_m)$$

or, if $\boldsymbol{w}_{m+1}$ and $\boldsymbol{v}_{m+1}$ have also been computed,

$$\sigma_{\min}(A - zI) \approx \min\{\sigma_{\min}(W_{m+1}^*(A - zI)V_m), \sigma_{\min}(W_m^*(A - zI)V_{m+1})\}. \tag{18}$$

The key idea is to use the shift-invariant subspaces $\mathcal{V}_m = \mathcal{K}_m(A, \boldsymbol{v}_1)$ and $\mathcal{W}_m = \mathcal{K}_m(A^*, \boldsymbol{w}_1)$ to compute the smallest singular values in a region surrounding the eigenvalues of interest by imposing the Galerkin conditions

$$\begin{aligned}(A - zI)V_m\boldsymbol{c} - \theta W_m\boldsymbol{d} &\perp \mathcal{W}_m,\\ (A - zI)^*W_m\boldsymbol{d} - \theta V_m\boldsymbol{c} &\perp \mathcal{V}_m\end{aligned}$$

for a large number of complex shifts $z$. Furthermore, the two-sided approach is symmetric in the sense that the same results are obtained if $A$ is replaced by $A^*$ and the starting vectors are swapped, which is not the case for the one-sided approximation.

We compute the pseudospectra of three disparate matrices in specific regions. The first matrix, randn, is generated using the identically named MATLAB function. The second and third matrices, rdb800l and pipe, are taken from Wright and Trefethen [26, sect. 5]. For the Krylov–Schur algorithms we use minimum dimension $m = 25$ and maximum dimension $\ell = 50$. Table 2 lists additional details, including the number of restarts, which are hand-picked to achieve near-optimal results. Because of the conditioning of the eigenvalues, we recompute $V_{m+1}^*AV_m$ and $W_{m+1}^*AV_m$ before computing the pseudospectra, as opposed to working with $(W_{m+1}^*V_{m+1})\underline{H}_m$ and $(V_{m+1}^*W_{m+1})\underline{K}_m$.

TABLE 2

*The dimension size, region of interest, target, use of harmonic extraction, and number of restarts (RS) for one-sided and two-sided Krylov–Schur for each matrix.*

| Name | $n$ | Region | Target | Harmonic | RS-1 | RS-2 |
|---|---|---|---|---|---|---|
| randn | 1024 | $[-27, -17] \times [17, 25]$ | $-22 + 21i$ | yes | 100 | 25 |
| rdb800l | 800 | $[-1.1, 1.1] \times [-0.25, 2.75]$ | $+1.25i$ | no | 125 | 50 |
| pipe | 402 | $[-0.15, 0.05] \times [-0.05, 0.05]$ | $+0.05$ | yes | 1500 | 1000 |

The pseudospectra of the test matrices can be seen in Figure 2, and their approximations with one-sided and two-sided Krylov–Schur in Figures 3 and 4, respectively. The latter two figures also include heat maps of the quantity

$$z \mapsto \log_{10}\left|\frac{\sigma_{\min}(A - zI) - \theta}{\sigma_{\min}(A - zI)}\right|, \tag{19}$$

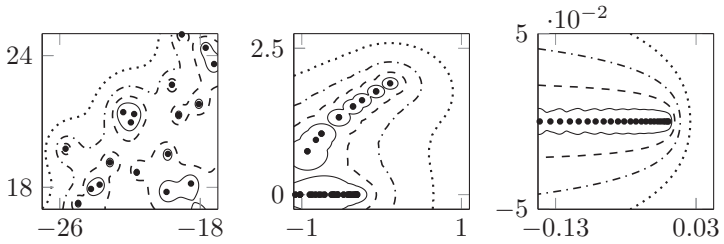


FIG. 2. *Pseudospectra of* randn *(left),* rdb800l *(middle), and* pipe *(right). The level curves range from* $10^{-1.7}$ *to* $10^{-0.5}$, $10^{-1.4}$ *to* $10^{-0.5}$, *and* $10^{-5}$ *to* $10^{-3.5}$, *respectively.*

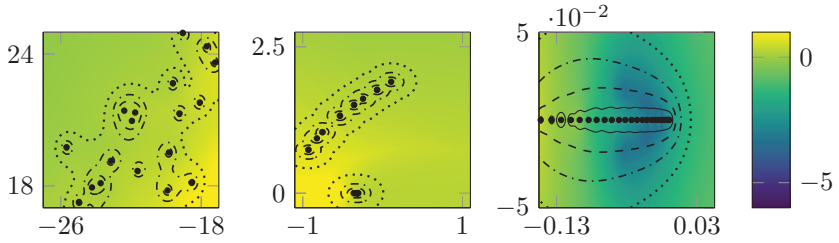


FIG. 3. *Level curves for the pseudospectra approximations obtained with one-sided Krylov–Schur, with* randn *(left),* rdb800l *(middle), and* pipe *(right). The heat maps show the value of the error measure defined in* (19) *and have the average values* +0.249, +0.366, *and* −1.372, *respectively.*

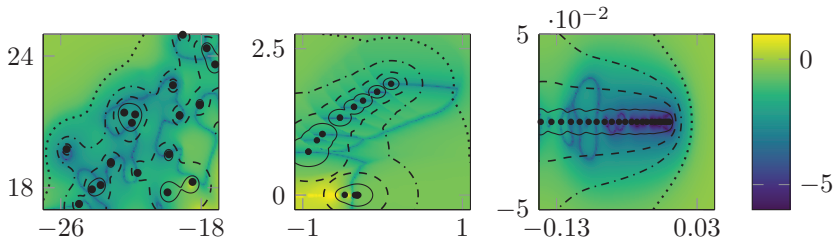


FIG. 4. *Level curves for the pseudospectra approximations obtained with one-sided Krylov–Schur, with* randn *(left),* rdb800l *(middle), and* pipe *(right). The heat maps show the value of the error measure defined in* (19) *and have the average values* −1.464, −0.920, *and* −1.597, *respectively.*

where $\theta$ is the approximation from either (17) for one-sided Krylov–Schur or (18) for two-sided Krylov–Schur. The first two subplots in Figure 3 show that one-sided Krylov–Schur is capable of capturing the qualitative behavior of the pseudospectrum reasonably well, although the level curves appear to be "shifted." For instance, the outermost level curve in the rdb800l approximation corresponds to $\varepsilon = 10^{-0.5}$, while the true pseudospectrum has the level curve for $\varepsilon = 10^{-0.8}$ at approximately the same position. The displacement of the level curves is presumably caused by the high relative errors in the approximate singular values. Indeed, for the same two examples, two-sided Krylov–Schur achieves lower relative errors and has better contour approximations. The approximation quality in the last example is comparable for both methods, with the two-sided Krylov–Schur approximation being more accurate near the eigenvalues, and the one-sided Krylov–Schur approximation being better further away. This contrast might be explained by the two-sided Rayleigh quotient

having faster asymptotic convergence than the one-sided Rayleigh quotient. Finally, we remark that only the one-sided approximation of the singular value is monotonic in the sense that

$$\sigma_{\min}(A - zI) \leq \sigma_{\min}(V_{m+1}^*(A - zI)V_m),$$

and as a result, there are areas where

$$\sigma_{\min}(A-zI) \leq \sigma_{\min}(W_m^*(A-zI)V_m) \qquad \text{or} \qquad \sigma_{\min}(A-zI) \geq \sigma_{\min}(W_m^*(A-zI)V_m),$$

separated by curves where

$$\sigma_{\min}(A - zI) = \sigma_{\min}(W_m^*(A - zI)V_m).$$

These "zero-error" curves tend to connect accurate Ritz values and show up as dark(er) lines in the heat maps in Figure 4.

**9. Conclusions.** We have presented a two-sided Krylov–Schur method for nonnormal matrices as a natural generalization of the one-sided Krylov–Schur approach by Stewart. An advantage of two-sided Krylov–Schur over two-sided Lanczos is the use of orthonormal bases, and an advantage over one-sided Krylov–Schur is the simultaneous approximation of left and right eigenvectors or eigenspaces. The two-sided approximations may already give useful information concerning eigenvalue conditioning during the iterations. Furthermore, for some applications, the two-sided method may converge with fewer matrix-vector products than the standard Krylov–Schur method.

Primary disadvantages of the new method are the computational cost per iteration, which is roughly twice that of the one-sided Krylov–Schur method, and potential numerical stability and accuracy issues in the computation of the Ritz values. However, the numerical issues caused by oblique projections can mainly be avoided with a proper implementation, as discussed in section 5.

The two-sided Krylov–Schur method may be combined with either the standard two-sided Rayleigh–Ritz extraction or the harmonic two-sided Rayleigh–Ritz extraction. We have seen that the implementation of the standard two-sided extraction is relatively straightforward, while the implementation of the harmonic extraction is more complicated.

Theoretical convergence properties have been investigated and generalized and show when and how well we can expect two-sided methods to converge. Furthermore, numerical experiments demonstrate that two-sided Krylov–Schur may excel in finding the best-conditioned eigenvalues of nonnormal matrices. Additional numerical experiments show that the shift-invariant left and right Krylov spaces computed with two-sided Krylov–Schur may be useful for the approximation of pseudospectra.

**Acknowledgments.** We would like to thank the editor and the referees for their excellent feedback and helpful suggestions.

REFERENCES

[1] Z. Bai, D. Day, J. Demmel, and J. Dongarra, *A Test Matrix Collection for Non-Hermitian Eigenvalue Problems*, Technical report, University of Tennessee, Knoxville, TN, 1996.

[2] S. Chaturantabut and D. C. Sorensen, *Nonlinear model reduction via discrete empirical interpolation*, SIAM J. Sci. Comput., 32 (2010), pp. 2737–2764, https://doi.org/10.1137/090766498.

[3] T.-Y. Chen and J. W. Demmel, *Balancing sparse matrices for computing eigenvalues*, Linear Algebra Appl., 309 (2000), pp. 261–287, https://doi.org/10.1016/S0024-3795(00)00014-8.

[4] J. Cullum and T. Zhang, *Two-sided Arnoldi and nonsymmetric Lanczos algorithms*, SIAM J. Matrix Anal. Appl., 24 (2002), pp. 303–319, https://doi.org/10.1137/S0895479898339013.
[5] R. J. A. David and D. S. Watkins, *An inexact Krylov–Schur algorithm for the unitary eigenvalue problem*, Linear Algebra Appl., 429 (2008), pp. 1213–1228, https://doi.org/10.1016/j.laa.2007.09.034.
[6] M. Hochbruck and M. E. Hochstenbach, *Subspace Extraction for Matrix Functions*, preprint, 2005.
[7] I. M. Jaimoukha and E. M. Kasenally, *Implicitly restarted Krylov subspace methods for stable partial realizations*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 633–652, https://doi.org/10.1137/S0895479895279873.
[8] Z. Jia and G. W. Stewart, *An analysis of the Rayleigh–Ritz method for approximating eigenspaces*, Math. Comp., 70 (2001), pp. 637–647, https://doi.org/10.1090/S0025-5718-00-01208-4.
[9] W. Kahan, B. N. Parlett, and E. Jiang, *Residual bounds on approximate eigensystems of nonnormal matrices*, SIAM J. Numer. Anal., 19 (1982), pp. 470–484, https://doi.org/10.1137/0719030.
[10] D. Kressner, *Numerical Methods and Software for General and Structured Eigenvalue Problems*, Ph.D. thesis, TU Berlin, Berlin, Germany, 2004, https://doi.org/10.14279/depositonce-983.
[11] D. Kressner, *A periodic Krylov–Schur algorithm for large matrix products*, Numer. Math., 103 (2006), pp. 461–483, https://doi.org/10.1007/s00211-006-0682-1.
[12] S. Noschese and L. Reichel, *Inverse subspace problems with applications*, Numer. Linear Algebra Appl., 21 (2014), pp. 589–603, https://doi.org/10.1002/nla.1914.
[13] J. Rommes and N. Martins, *Computing large-scale system eigenvalues most sensitive to parameter changes, with applications to power system small-signal stability*, IEEE Trans. Power Syst., 23 (2008), pp. 434–442, https://doi.org/10.1109/TPWRS.2008.920050.
[14] A. Ruhe, *The two-sided Arnoldi algorithm for nonsymmetric eigenvalue problems*, in Matrix Pencils, B. Kågström and A. Ruhe, eds., Springer, Berlin, Heidelberg, 1983, pp. 104–120, https://doi.org/10.1007/BFb0062097.
[15] A. Ruhe, *Lanczos method*, in Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide, Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, eds., SIAM, Philadelphia, 2000, pp. 56–66 (section 4.4), https://doi.org/10.1137/1.9780898719581.ch4.
[16] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, rev. ed., SIAM, Philadelphia, 2011, https://doi.org/10.1137/1.9781611970739.
[17] G. L. G. Sleijpen, H. A. van der Vorst, and E. Meijerink, *Efficient expansion of subspaces in the Jacobi–Davidson method for standard and generalized eigenproblems*, Electron. Trans. Numer. Anal., 7 (1998), pp. 75–89.
[18] G. W. Stewart, *A generalization of Saad's theorem on Rayleigh–Ritz approximations*, Linear Algebra Appl., 327 (2001), pp. 115–119, https://doi.org/10.1016/S0024-3795(00)00324-4.
[19] G. W. Stewart, *Matrix Algorithms. Volume* II: *Eigensystems*, SIAM, Philadelphia, 2001, https://doi.org/10.1137/1.9780898718058.
[20] G. W. Stewart, *A Krylov–Schur algorithm for large eigenproblems*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 601–614, https://doi.org/10.1137/S0895479800371529.
[21] G. W. Stewart, *Addendum to "A Krylov–Schur algorithm for large eigenproblems,"* SIAM J. Matrix Anal. Appl., 24 (2002), pp. 599–601, https://doi.org/10.1137/S0895479802403150.
[22] G. W. Stewart, *On the numerical analysis of oblique projectors*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 309–348, https://doi.org/10.1137/100792093.
[23] M. Stoll, *A Krylov–Schur approach to the truncated SVD*, Linear Algebra Appl., 436 (2012), pp. 2795–2806, https://doi.org/10.1016/j.laa.2011.07.022.
[24] D. Szyld, *The many proofs of an identity on the norm of oblique projections*, Numer. Algorithms, 42 (2006), pp. 309–323, https://doi.org/10.1007/s11075-006-9046-2.
[25] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, Princeton, NJ, 2005.
[26] T. G. Wright and L. N. Trefethen, *Large-scale computation of pseudospectra using ARPACK and eigs*, SIAM J. Sci. Comput., 23 (2001), pp. 591–605, https://doi.org/10.1137/S106482750037322X.
[27] Y. Zhou and Y. Saad, *Block Krylov–Schur method for large symmetric eigenvalue problems*, Numer. Algorithms, 47 (2008), pp. 341–359, https://doi.org/10.1007/s11075-008-9192-9.