Received: 12 August 2019 | Revised: 2 May 2020 | Accepted: 7 June 2020

DOI: 10.1002/nla.2324

RESEARCH ARTICLE

WILEY

New strategies for determining backward perturbation bound of approximate two-sided Krylov subspaces

Yunjie Wang[1,2] | **Gang Wu**[1]

[1]School of Mathematics, China University of Mining and Technology, Xuzhou, P.R. China

[2]Kewen Institute, Jiangsu Normal University, Xuzhou, P.R. China

Correspondence
Gang Wu, School of Mathematics, China University of Mining and Technology, Xuzhou, Jiangsu 221116, P.R. China.
Email: gangwu@cumt.edu.cn

Funding information
the Fundamental Research Funds for the Central Universities, Grant/Award Number: 2019XKQYMS89

Summary

Given a nonsymmetric matrix $A \in \mathbb{R}^{n\times n}$ and two unit norm vectors, the two-sided Krylov subspace methods construct a pair of bases for two Krylov subspaces with respect to A and A^{T}, respectively. In practical calculations, however, the two subspaces spanned by the computed bases may not be Krylov subspaces. Given two subspaces $\mathcal{K}$ and $\mathcal{L}$, in [G. Wu et al, *Toward backward perturbation bounds for approximate dual Krylov subspaces*, BIT, 53 (2013), pp. 225-239], the authors considered how to determine a backward perturbation E whose norm is as small as possible, such that $\mathcal{K}$ and $\mathcal{L}$ are Krylov subspaces of $A+E$ and $(A+E)^{\mathrm{T}}$, respectively. However, as the two bases used are biorthonormal, their results are nonoptimal in terms of unitarily invariant norms, and the perturbation bound can be greatly overestimated. In this work, we revisit this problem and use *orthonormal* bases instead of *biorthonormal* bases to derive new perturbation bounds. We propose two new strategies, the first one focuses on choosing optimal orthonormal basis matrices, and the second one resorts to solving small-sized generalized Sylvester matrix equations. Numerical experiments show that our bounds improve the existing one substantially.

KEYWORDS

backward perturbation, Krylov subspace, Lagrange multiplier method, two-sided Arnoldi method, two-sided Krylov subspaces, two-sided Lanczos method

1 | INTRODUCTION

Krylov subspace methods play a very important role in large-scale scientific computations.[1-6] For instance, they are especially important in solving large-scale linear systems,[4,7-10] approximating a few eigenpairs of large matrices,[5,6,11,12] solving large-scale least squares problems;[2,3] and they are also popular for large-scale model reduction problems[13-16] and for large matrix functions.[17-19]

Given a nonsymmetric matrix $A \in \mathbb{R}^{n\times n}$, let $\mathbf{u}$ be a vector of unit norm, then

$$\mathcal{K}_{m+1}(A, \mathbf{u}) = \mathrm{span}\{\mathbf{u}, A\mathbf{u}, A^2\mathbf{u}, \ \dots \ , A^m\mathbf{u}\}$$

is called an $(m+1)$-dimensional Krylov subspace with respect to A and $\mathbf{u}$, if and only if the grade of $\mathbf{u}$ associated with A is not less than $m+1$ [4, p. 152]. However, the vectors in the Krylov sequence will become increasingly dependent as m increases. To circumvent this difficulty, one often builds an orthonormal basis for the Krylov subspace. The Arnoldi

method is one of the most popular (one-sided) Krylov subspace methods.[11] The m-step Arnoldi process can be formulated as follows References 4-6

$$AU_m = U_{m+1}G_{m+1,m},$$

where $U_{m+1} \in \mathbb{R}^{n\times(m+1)}$ is an orthonormal basis matrix for $\mathcal{K}_{m+1}(A, \mathbf{u})$, $G_{m+1,m} = U_{m+1}^{\mathrm{T}}AU_m \in \mathbb{R}^{(m+1)\times m}$ is an upper Hessenberg matrix, with $m \ll n$. In particular, when A is symmetric, the reduced matrix $G_m = U_m^{\mathrm{T}}AU_m$ is symmetric and tridiagonal, and the Arnoldi decomposition reduces to the Lanczos decomposition.[9]

The two-sided (or the nonsymmetric) Lanczos method and the two-sided Arnoldi method are two representative two-sided Krylov subspaces methods. The two-sided Lanczos method is a commonly used technique for computing both the right and the left eigenpairs *simultaneously* of a large nonsymmetric matrix.[1,9,12,20,21] It exploits three-term recurrences with biorthonormal bases, at the expense of doubling the number of matrix-vector multiplications compared with the Arnoldi method. More precisely, given two vectors $\mathbf{v}_1$ and $\mathbf{w}_1$ satisfying $\mathbf{w}_1^{\mathrm{T}}\mathbf{v}_1 = 1$, the two-sided Lanczos method builds a pair of *biorthonormal* basis matrices W_{m+1} and V_{m+1} (ie, $W_{m+1}^{\mathrm{T}}V_{m+1} = I$) for two-sided Krylov subspaces[4-7,10,20-22]

$$\mathcal{K}_{m+1}(A^{\mathrm{T}}, \mathbf{w}_1) = \mathrm{span}\{\mathbf{w}_1, A^{\mathrm{T}}\mathbf{w}_1, \ldots, (A^{\mathrm{T}})^m\mathbf{w}_1\} \quad \text{and} \quad \mathcal{K}_{m+1}(A, \mathbf{v}_1) = \mathrm{span}\{\mathbf{v}_1, A\mathbf{v}_1, \ldots, A^m\mathbf{v}_1\}. \tag{1}$$

The m-step two-sided Lanczos decomposition has the following form

$$\begin{cases} AV_m = V_mT_m + \mathbf{v}_{m+1}\mathbf{b}_m^{\mathrm{T}} \\ A^{\mathrm{T}}W_m = W_mT_m^{\mathrm{T}} + \mathbf{w}_{m+1}\mathbf{c}_m^{\mathrm{T}}, \end{cases} \tag{2}$$

where T_m is nonsymmetric and tridiagonal.

However, the two-sided Lanczos method may suffer from numerical instability because nonorthogonal basis matrices are utilized.[1,23] The accuracy and stability of the computed bases can be improved by using the two-sided Arnoldi method,[17,24-28] which replaces the *biorthonormal* basis matrices by using two *orthonormal* basis matrices V_{m+1} and W_{m+1} (ie, $V_{m+1}^{\mathrm{T}}V_{m+1} = W_{m+1}^{\mathrm{T}}W_{m+1} = I$). The two-sided Arnoldi method proposed by Ruhe,[26,27] and later as a block method by Cullum and Zhang,[24] independently generates orthonormal bases for the right search space $\mathcal{K}$ and the left search space $\mathcal{L}$. This can be performed by applying the regular Arnoldi method twice. That is, given two unit norm vectors $\mathbf{v}_1$ and $\mathbf{w}_1$, one performs m steps of the Arnoldi method on A to compute an orthonormal basis matrix V_{m+1} for the Krylov subspace $\mathcal{K}_{m+1}(A, \mathbf{v}_1)$, and then perform m steps of the Arnoldi method on A^{T} to compute an orthonormal basis matrix W_{m+1} for the Krylov subspace $\mathcal{K}_{m+1}(A^{\mathrm{T}}, \mathbf{w}_1)$. The following two-sided Arnoldi relations hold

$$\begin{cases} AV_m = V_{m+1}H_{m+1,m} \\ A^{\mathrm{T}}W_m = W_{m+1}K_{m+1,m}, \end{cases} \tag{3}$$

where $H_{m+1,m}$,$K_{m+1,m}$ are two $(m+1)$-by-m matrices. The two-sided Arnoldi method can find approximations to both the left and the right eigenvectors of A simultaneously, and this method can be used to estimate the condition numbers of the eigenvalues of A. Moreover, the two-sided Arnoldi method can be utilized to evaluate the bilinear form $\mathbf{c}^{\mathrm{H}}f(A)\mathbf{b}$,[15,16,29] where $f(A)$ is a matrix function on A and $\mathbf{b}$,$\mathbf{c}$ are two given vectors, with $\mathbf{c}^{\mathrm{H}}$ being the conjugate transpose of $\mathbf{c}$. Recently, the two-sided Arnoldi method was considered in the task of updating matrix functions $f(A)$, where the matrix $A \in \mathbb{R}^{n\times n}$ is subject to a low-rank modification.[17] A two-sided Arnoldi algorithm with Krylov-Schur restarting was investigated in Reference 28, and a compact two-sided Krylov method was applied to solve nonlinear eigenvalue problems in Reference 25.

In practical calculations, however, the subspaces spanned by the computed bases may not be Krylov subspaces.[22,30] For instance, the application of A may be inexact in practice,[31-34] such that the constructed subspaces are not Krylov subspaces with respect to A. Moreover, it is well known that the bases built in the Arnoldi method, the Lanczos method, and the two-sided Lanczos method may lose orthogonality or biorthogonality during iterations.[4] Indeed, loss of orthogonality is a sufficient and necessary condition for the convergence of Ritz pairs.[35] In this case, the subspaces spanned by the bases are not Krylov subspaces with respect to A or A^T any longer. Although some reorthogonalization techniques can be applied to deal with these problems,[4,20,36] the subspaces spanned by the new bases are still not Krylov subspaces in general, and it is desirable to evaluate the quality of the computed subspaces.

For the one-sided Krylov subspace methods such as the Arnoldi method, let $\mathcal{U}$ be a subspace of dimension $m+1$, Stewart[30] determined a matrix E of *minimal norm* such that $\mathcal{U}$ is a Krylov subspace of $A+E$. The proof consists of two steps. First, it is shown that there is a basis for $\mathcal{U}$ that satisfies an approximate Krylov relation for A with a minimal residual. Second, standard techniques are used to throw the residual back onto A. Given a matrix $A \in \mathbb{R}^{n\times n}$ and two subspaces $\mathcal{K}$ and $\mathcal{L}$ of dimension $m+1$, Wu et al[22] generalized this result and considered how to determine a *quasioptimal* perturbation E, such that $\mathcal{K}, \mathcal{L}$ are Krylov subspaces of $A+E$ and $(A+E)^{\mathrm{T}}$, respectively. This problem arises from evaluating the quality of the computed two-sided Krylov subspaces. However, the two bases used in their strategy are *biorthonormal*, and the results given in Reference 22 are *nonoptimal* in terms of unitarily invariant norms when A is nonsymmetric. Consequently, the perturbation bound can be greatly overestimated in practice. Thus, it is interesting to improve the result and to give better perturbation bounds for approximate two-sided Krylov subspaces, and it is challenging to find the perturbation matrix that is optimal under unitarily invariant norms.[22]

In this work, we revisit this problem and aim to derive better perturbation bounds than the one given in Reference 22. Two strategies are proposed. The first one relies on choosing appropriate orthonormal bases for $\mathcal{K}$ and $\mathcal{L}$, instead of biorthonormal bases as in Reference 22. The key of the second strategy resorts to solving small-sized generalized Sylvester matrix equations. As the norm of E can be utilized to evaluate the quality of the computed two-sided Krylov subspaces, and the norms of a matrix are equivalent,[3] it suffices to consider the Frobenius norm of the backward perturbation. In the proposed methods, the norms of the perturbation matrices can be evaluated by using the bases chosen and the Krylov residuals involved, which are easy to compute in practice. Note that the backward perturbations for approximate invariant subspace or singular value subspace are considered in References 37-39, which are different problems from ours.

The framework of this article is as follows. Given two subspaces $\mathcal{K}$ and $\mathcal{L}$, in Section 2, we consider how to choose appropriate orthonormal basis matrices for determining a backward perturbation bound. For the chosen orthonormal basis matrices, in Section 3, we focus on computing two optimal matrices H and K, and construct a backward perturbation matrix whose Frobenius norm is as small as possible. Numerical experiments are performed in Section 4 to demonstrate the efficacy of our new strategies. Some concluding remarks are given in Section 5.

We use the following notations. Let $A \in \mathbb{R}^{n\times n}$ be a nonsymmetric matrix, let $\mathcal{K}$, $\mathcal{L}$ be given subspaces of dimension $m+1$ with $m \ll n$, and let V_{m+1}, W_{m+1} be two real orthonormal basis matrices whose columns are bases for $\mathcal{K}$ and $\mathcal{L}$, respectively. Throughout this article, we make the assumption that the matrix $W_{m+1}^{\mathrm{T}} V_{m+1}$ is nonsingular, where W_{m+1}^{T} denotes the transpose of W_{m+1}. Let V_m, W_m be matrices consisting of the first m columns of V_{m+1} and W_{m+1}, respectively. We denote by $\mathrm{tr}(A)$ the trace of a matrix A, and by $\|A\|_F$ the Frobenius norm (ie, the F-norm) of A. Let $F(X)$ be a matrix function on X, then $\frac{\partial F(X)}{\partial X}$ stands for the elementwise derivative of $F(X)$ on the matrix X. The symbol I stands for the identity matrix and the symbol $\mathbf{0}$ denotes a zero matrix or vector, whose order are clear from the context.

2 | CHOOSING OPTIMAL ORTHONORMAL BASIS MATRICES FOR GIVEN H AND K

The *backward error* E is an n-by-n matrix whose norm is as small as possible, such that the following *Krylov decompositions* [6, chapter 4] hold

$$\begin{cases} (A+E)V_m = V_{m+1}H_{m+1,m} \\ (A+E)^{\mathrm{T}}W_m = W_{m+1}K_{m+1,m}, \end{cases} \tag{4}$$

where $V_{m+1}^{\mathrm{T}} V_{m+1} = W_{m+1}^{\mathrm{T}} W_{m+1} = I_{m+1}$, and $H_{m+1,m}, K_{m+1,m} \in \mathbb{R}^{(m+1)\times m}$ has no particular form. Indeed, $H_{m+1,m}$ and $K_{m+1,m}$ can easily be made upper Hessenberg if necessary. We drop the subscripts of H and K for ease of notation. In this section, by choosing optimal orthonormal basis matrices for given H and K, we aim to seek a backward perturbation E whose norm is as small as possible. Note that (4) is different from the one discussed in Reference 22, where the two matrices V_{m+1} and W_{m+1} are *biorthonormal* rather than *orthonormal*. This problem is also different from the one discussed in Reference 30, where one-sided Krylov subspace methods are considered. The following theorem considers the existence of the solutions to (4), whose proof is along the lines of [22, theorem 2.2]. For completeness, we present a proof here.

Theorem 1. *Let V_{m+1} and W_{m+1} be two orthonormal basis matrices for $\mathcal{K}$ and $\mathcal{L}$, respectively, and let $V_m^\perp$ and $W_m^\perp$ be orthonormal basis matrices of the orthogonal complement of* $\mathrm{span}\{V_m\}$ *and* $\mathrm{span}\{W_m\}$*, respectively. Denote $V = [V_m,\ V_m^\perp]$ and $W = [W_m,\ W_m^\perp]$, then the solutions to (4) exist if and only if*

$$(W_m^{\mathrm{T}} V_{m+1})H = K^{\mathrm{T}}(W_{m+1}^{\mathrm{T}} V_m). \tag{5}$$

Moreover, the solutions have the following form

$$E = -W \begin{bmatrix} W_m^{\mathrm{T}} R & S^{\mathrm{T}} V_m^\perp \\ (W_m^\perp)^{\mathrm{T}} R & Z \end{bmatrix} V^{\mathrm{T}}, \tag{6}$$

where the matrix $Z \in \mathbb{R}^{(n-m)\times(n-m)}$ is arbitrary, and

$$\begin{cases} R = AV_m - V_{m+1}H \\ S = A^{\mathrm{T}} W_m - W_{m+1}K. \end{cases}$$

Proof. To prove the necessary condition, suppose that E is a solution to (4). Premultiplying the two equations in (4) from the left by W_m^{T} and V_m^{T} gives

$$\begin{cases} W_m^{\mathrm{T}}(A+E)V_m = W_m^{\mathrm{T}} V_{m+1} H \\ V_m^{\mathrm{T}}(A+E)^{\mathrm{T}} W_m = V_m^{\mathrm{T}} W_{m+1} K. \end{cases}$$

So we get

$$W_m^{\mathrm{T}} V_{m+1} H = K^{\mathrm{T}} W_{m+1}^{\mathrm{T}} V_m,$$

and (5) is satisfied. Thus, (5) is equivalent to $W_m^{\mathrm{T}} R = S^{\mathrm{T}} V_m$, that is,

$$W_m^{\mathrm{T}} R = W_m^{\mathrm{T}} A V_m - W_m^{\mathrm{T}} V_{m+1} H = W_m^{\mathrm{T}} A V_m - K^{\mathrm{T}} W_{m+1}^{\mathrm{T}} V_m = S^{\mathrm{T}} V_m.$$

Recall that both V and W are orthonormal, so any matrix $E \in \mathbb{R}^{n\times n}$ can be expressed as $E = -WMV^{\mathrm{T}}$, where $M \in \mathbb{R}^{n\times n}$ is to be determined. Postmultiplying E from the right by V_m yields

$$EV_m = -WMV^{\mathrm{T}} V_m = -WM \begin{bmatrix} I \\ \mathbf{0} \end{bmatrix}. \tag{7}$$

Decompose

$$M = \begin{bmatrix} B & C \\ D & Z \end{bmatrix},$$

where $B \in \mathbb{R}^{m\times m}$, $C \in \mathbb{R}^{m\times(n-m)}$, $D \in \mathbb{R}^{(n-m)\times m}$, and $Z \in \mathbb{R}^{(n-m)\times(n-m)}$. By (4), $EV_m = -R$, and it follows from (7) that

$$\begin{bmatrix} B \\ D \end{bmatrix} = M \begin{bmatrix} I \\ \mathbf{0} \end{bmatrix} = W^{\mathrm{T}} R = \begin{bmatrix} W_m^{\mathrm{T}} R \\ (W_m^\perp)^{\mathrm{T}} R \end{bmatrix}. \tag{8}$$

Similarly, we have

$$\begin{bmatrix} B, C \end{bmatrix} = \begin{bmatrix} S^{\mathrm{T}} V_m, & S^{\mathrm{T}} V_m^\perp \end{bmatrix}, \tag{9}$$

and the solutions to (4) have the form of (6).

On the other hand, assume that (5) holds. Let E be the form of (6), then

$$EV_m = -W \begin{bmatrix} W_m^{\mathrm{T}} R & S^{\mathrm{T}} V_m^{\perp} \\ (W_m^{\perp})^{\mathrm{T}} R & Z \end{bmatrix} V^{\mathrm{T}} V_m = -W \begin{bmatrix} W_m^{\mathrm{T}} R & S^{\mathrm{T}} V_m^{\perp} \\ (W_m^{\perp})^{\mathrm{T}} R & Z \end{bmatrix} \begin{bmatrix} I \\ \mathbf{0} \end{bmatrix} = -R,$$

and we get

$$(A + E)V_m = V_{m+1}H.$$

Moreover,

$$E^{\mathrm{T}} W_m = -V \begin{bmatrix} R^{\mathrm{T}} W_m & R^{\mathrm{T}} W_m^{\perp} \\ (V_m^{\perp})^{\mathrm{T}} S & Z^{\mathrm{T}} \end{bmatrix} W^{\mathrm{T}} W_m = -V \begin{bmatrix} V_m^{\mathrm{T}} S & R^{\mathrm{T}} W_m^{\perp} \\ (V_m^{\perp})^{\mathrm{T}} S & Z^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} I \\ \mathbf{0} \end{bmatrix} = -S,$$

where we used $W_m^{\mathrm{T}} R = S^{\mathrm{T}} V_m$. Thus, it follows that

$$(A + E)^{\mathrm{T}} W_m = W_{m+1}K,$$

and E is a solution to (4). □

In view of (1) and (6), the perturbation matrix E is not only related to V_{m+1} and W_{m+1}, but also to R, S, and Z. Given V_{m+1} and W_{m+1}, the norms of R and S are related to the choices of H and K. Therefore, it is interesting to consider how to choose H and K in practice. A natural idea is to seek two matrices $\underline{H}$ and $\underline{K}$, such that

$$\underline{H} = \arg \min_{H \in \mathbb{R}^{(m+1)\times m}} \|R\|_F = \arg \min_{H \in \mathbb{R}^{(m+1)\times m}} \|AV_m - V_{m+1}H\|_F$$

and

$$\underline{K} = \arg \min_{K \in \mathbb{R}^{(m+1)\times m}} \|S\|_F = \arg \min_{K \in \mathbb{R}^{(m+1)\times m}} \|A^{\mathrm{T}} W_m - W_{m+1}K\|_F.$$

As both V_{m+1} and W_{m+1} are orthonormal, the solutions of the above problems are Reference 6

$$\underline{H} = V_{m+1}^{\mathrm{T}} A V_m \quad \text{and} \quad \underline{K} = W_{m+1}^{\mathrm{T}} A^{\mathrm{T}} W_m.$$

Unfortunately, the above two matrices do not satisfy the restriction (5) in general.

If $W_{m+1}^{\mathrm{T}} V_{m+1}$ is nonsingular, we define the two matrices $\hat{H}$ and $\hat{K}$ as

$$\hat{H} = (W_{m+1}^{\mathrm{T}} V_{m+1})^{-1} W_{m+1}^{\mathrm{T}} A V_m, \quad \text{and} \quad \hat{K} = (V_{m+1}^{\mathrm{T}} W_{m+1})^{-1} V_{m+1}^{\mathrm{T}} A^{\mathrm{T}} W_m, \tag{10}$$

then it follows that

$$\begin{aligned} \hat{H} &= \arg \min_{H \in \mathbb{R}^{(m+1)\times m}} \|W_{m+1} W_{m+1}^{\mathrm{T}} R\|_F \\ &= \arg \min_{H \in \mathbb{R}^{(m+1)\times m}} \|W_{m+1} W_{m+1}^{\mathrm{T}} (AV_m - V_{m+1}H)\|_F \\ &= \arg \min_{H \in \mathbb{R}^{(m+1)\times m}} \|W_{m+1}^{\mathrm{T}} (AV_m - V_{m+1}H)\|_F, \end{aligned} \tag{11}$$

and

$$\begin{aligned} \hat{K} &= \arg \min_{K \in \mathbb{R}^{(m+1)\times m}} \|V_{m+1} V_{m+1}^{\mathrm{T}} S\|_F \\ &= \arg \min_{K \in \mathbb{R}^{(m+1)\times m}} \|V_{m+1}^{\mathrm{T}} (A^{\mathrm{T}} W_m - W_{m+1}K)\|_F. \end{aligned} \tag{12}$$

Thus, $\hat{H}$ and $\hat{K}$ are *optimal* in the sense of (11) and (12). We have the following theorem for the solutions E:

Theorem 2. *Let $\hat{H}$ and $\hat{K}$ be the matrices defined in (10), and let*

$$\hat{R}_m = AV_m - V_{m+1}\hat{H}, \quad \hat{S}_m = A^{\mathrm{T}} W_m - W_{m+1}\hat{K}. \tag{13}$$

Then

$$W_m^{\mathrm{T}} V_{m+1} \hat{H} = \hat{K}^{\mathrm{T}} W_{m+1}^{\mathrm{T}} V_m, \tag{14}$$

and

$$W_{m+1}^{\mathrm{T}} \hat{R}_m = \mathbf{0}, \quad \hat{S}_m^{\mathrm{T}} V_{m+1} = \mathbf{0}. \tag{15}$$

Moreover,

$$E = -W \begin{bmatrix} \mathbf{0} & \hat{S}_m^{\mathrm{T}} V_m^{\perp} \\ (W_m^{\perp})^{\mathrm{T}} \hat{R}_m & Z \end{bmatrix} V^{\mathrm{T}} \tag{16}$$

are solutions to (4) for any matrix $Z \in \mathbb{R}^{(n-m)\times(n-m)}$. Specifically, if we let

$$\hat{E} = -\hat{R}_m V_m^{\mathrm{T}} - W_m \hat{S}_m^{\mathrm{T}} \tag{17}$$

then it is a special solution to (4), with

$$\|\hat{E}\|_F = \min_{Z \in \mathbb{R}^{(n-m)\times(n-m)}} \|E\|_F = \sqrt{\|\hat{R}_m\|_F^2 + \|\hat{S}_m\|_F^2}. \tag{18}$$

Proof. We notice that (see also Reference 24)

$$W_{m+1}^{\mathrm{T}} A V_{m+1} = W_{m+1}^{\mathrm{T}} V_{m+1} \left((W_{m+1}^{\mathrm{T}} V_{m+1})^{-1} W_{m+1}^{\mathrm{T}} A V_{m+1} \right) \tag{19}$$

$$= \left(W_{m+1}^{\mathrm{T}} A V_{m+1} (W_{m+1}^{\mathrm{T}} V_{m+1})^{-1} \right) W_{m+1}^{\mathrm{T}} V_{m+1}. \tag{20}$$

Taking the first m rows and m columns of the right-hand sides of (19) and (20), we get

$$W_m^{\mathrm{T}} V_{m+1} \left((W_{m+1}^{\mathrm{T}} V_{m+1})^{-1} W_{m+1}^{\mathrm{T}} A V_m \right) = \left(W_m^{\mathrm{T}} A V_{m+1} (W_{m+1}^{\mathrm{T}} V_{m+1})^{-1} \right) W_{m+1}^{\mathrm{T}} V_m,$$

and (14) follows from (10). Moreover, we have that

$$\begin{aligned} W_{m+1}^{\mathrm{T}} \hat{R}_m &= W_{m+1}^{\mathrm{T}} A V_m - W_{m+1}^{\mathrm{T}} V_{m+1} (W_{m+1}^{\mathrm{T}} V_{m+1})^{-1} W_{m+1}^{\mathrm{T}} A V_m \\ &= W_{m+1}^{\mathrm{T}} A V_m - W_{m+1}^{\mathrm{T}} A V_m = \mathbf{0}, \end{aligned}$$

and thus $W_m^{\mathrm{T}} \hat{R}_m = \mathbf{0}$. Similarly, we can prove that $\hat{S}_m^{\mathrm{T}} V_{m+1} = \mathbf{0}$ and $\hat{S}_m^{\mathrm{T}} V_m = \mathbf{0}$. Combining (14) and (15) with Theorem 1, we know that (16) is a solution to (4) for any matrix $Z \in \mathbb{R}^{(n-m)\times(n-m)}$.

As both V and W are orthonormal, $\|E\|_F$ is minimized as Z is zero, and

$$\hat{E} = -W \begin{bmatrix} \mathbf{0} & \hat{S}_m^{\mathrm{T}} V_m^{\perp} \\ (W_m^{\perp})^{\mathrm{T}} \hat{R}_m & \mathbf{0} \end{bmatrix} V^{\mathrm{T}} = -\hat{R}_m V_m^{\mathrm{T}} - W_m \hat{S}_m^{\mathrm{T}}$$

is a special solution to (4). Thus, we have from $W_m^{\mathrm{T}} \hat{R}_m = \hat{S}_m^{\mathrm{T}} V_m = \mathbf{0}$ that

$$\begin{aligned} \min_{Z \in \mathbb{R}^{(n-m)\times(n-m)}} \|E\|_F = \|\hat{E}\|_F &= \sqrt{\|(W_m^{\perp})^{\mathrm{T}} \hat{R}_m\|_F^2 + \|\hat{S}_m^{\mathrm{T}} V_m^{\perp}\|_F^2} \\ &= \sqrt{\|W_m^{\perp} (W_m^{\perp})^{\mathrm{T}} \hat{R}_m\|_F^2 + \|\hat{S}_m^{\mathrm{T}} V_m^{\perp} (V_m^{\perp})^{\mathrm{T}}\|_F^2} \\ &= \sqrt{\|(I - W_m W_m^{\mathrm{T}}) \hat{R}_m\|_F^2 + \|\hat{S}_m^{\mathrm{T}} (I - V_m V_m^{\mathrm{T}})\|_F^2} \\ &= \sqrt{\|\hat{R}_m\|_F^2 + \|\hat{S}_m^{\mathrm{T}}\|_F^2} = \sqrt{\|\hat{R}_m\|_F^2 + \|\hat{S}_m\|_F^2}, \end{aligned}$$

which completes the proof. ∎

Remark 1. Given an n-by-n matrix A and two n-by-m orthonormal matrices X_m and Y_m. Suppose that $Y_m^{\mathrm{H}}X_m$ is nonsingular. The Kahan-Parlett-Jiang theorem[37] considers a matrix E for two *given* m-by-m matrices M_m and $L_m = (Y_m^{\mathrm{H}}X_m)^{-1}M_m^{\mathrm{H}}(Y_m^{\mathrm{H}}X_m)$, such that

$$\min_{Z\in\mathbb{C}^{(n-m)\times(n-m)}} \|E\|_{2,F} = \min_{Z\in\mathbb{C}^{(n-m)\times(n-m)}} \left\| Y \begin{bmatrix} Y_m^{\mathrm{H}}R & S^{\mathrm{H}}X_m^{\perp} \\ (Y_m^{\perp})^{\mathrm{H}}R & Z \end{bmatrix} X^{\mathrm{H}} \right\|_{2,F}, \tag{21}$$

where $R = AX_m - X_mL_m$ and $S = A^{\mathrm{H}}Y_m - Y_mM_m$, and $\|\cdot\|_{2,F}$ denotes 2-norm or Frobenius norm of a matrix. We stress that (18) and (21) are different problems from each other: the former is for the two-sided Krylov subspace problem, and the latter is for the two-sided invariant subspace problem. Specifically, both the two matrices $\hat{H}$ and $\hat{K}$ are rectangle and of size $(m+1)$-by-m in (18), while the two matrices M_m and L_m are square and of size m-by-m in (21).

When A is symmetric, it is desirable to construct a *symmetric* perturbation matrix $\hat{E}$ with minimal F-norm, such that $\mathcal{K}$ is a Krylov subspace of $A + \hat{E}$. In this situation, (17) reduces to [22, theorem 2.3]:

$$\hat{E} = -\hat{R}_mV_m^{\mathrm{T}} - V_m\hat{R}_m^{\mathrm{T}}, \tag{22}$$

which minimizes the Frobenius norm for given V,W,H, and K.

As was pointed out in References 22,30, R and S are optimal only for specific choices of bases, and large Krylov residuals may lead to large perturbations. Next, we aim to choose appropriate orthonormal bases, so that the Krylov residuals can be *minimized* in terms of F-norm as much as possible. The idea is to choose a pair of *orthonormal* matrices $P, Q \in \mathbb{R}^{(m+1)\times(m+1)}$ such that the Krylov residuals are minimized according to the new basis matrices $V_{m+1}P$ and $W_{m+1}Q$.

Let $P, Q \in \mathbb{R}^{(m+1)\times(m+1)}$ be any orthonormal matrices, and let $P_m, Q_m \in \mathbb{R}^{(m+1)\times m}$ be the orthonormal matrices composed of the first m columns of P and Q, respectively. Suppose that V_{m+1} and W_{m+1} are two given orthonormal basis matrices for the Krylov subspaces $\mathcal{K}$ and $\mathcal{L}$, respectively, we define

$$R_{m+1} = AV_{m+1} - V_{m+1}(W_{m+1}^{\mathrm{T}}V_{m+1})^{-1}W_{m+1}^{\mathrm{T}}AV_{m+1} \tag{23}$$

and

$$S_{m+1} = A^{\mathrm{T}}W_{m+1} - W_{m+1}(V_{m+1}^{\mathrm{T}}W_{m+1})^{-1}V_{m+1}^{\mathrm{T}}A^{\mathrm{T}}W_{m+1}. \tag{24}$$

Notice that the $\hat{R}_m$ and $\hat{S}_m$ that appeared in (13) are composed of the first m columns of R_{m+1} and S_{m+1}, respectively. Denote by

$$\tilde{V}_{m+1} = V_{m+1}P, \quad \tilde{W}_{m+1} = W_{m+1}Q, \tag{25}$$

and by $\tilde{V}_m = V_{m+1}P_m$, $\tilde{W}_m = W_{m+1}Q_m$ the matrices composed of the first m columns of $\tilde{V}_{m+1}$ and $\tilde{W}_{m+1}$, and let

$$\tilde{H} = (\tilde{W}_{m+1}^{\mathrm{T}}\tilde{V}_{m+1})^{-1}\tilde{W}_{m+1}^{\mathrm{T}}A\tilde{V}_m, \quad \tilde{K} = (\tilde{V}_{m+1}^{\mathrm{T}}\tilde{W}_{m+1})^{-1}\tilde{V}_{m+1}^{\mathrm{T}}A^{\mathrm{T}}\tilde{W}_m, \tag{26}$$

then it is easy to see that

$$A\tilde{V}_m - \tilde{V}_{m+1}\tilde{H} = R_{m+1}P_m, \quad A^{\mathrm{T}}\tilde{W}_m - \tilde{W}_{m+1}\tilde{K} = S_{m+1}Q_m. \tag{27}$$

Similar to the proof of Theorem 2, we have that

$$\tilde{W}_m^{\mathrm{T}}\tilde{V}_{m+1}\tilde{H} = \tilde{K}^{\mathrm{T}}\tilde{W}_{m+1}^{\mathrm{T}}\tilde{V}_m. \tag{28}$$

Let $\tilde{V} = [\tilde{V}_m,\ \tilde{V}_m^{\perp}]$, $\tilde{W} = [\tilde{W}_m,\ \tilde{W}_m^{\perp}]$ be two n-by-n orthonormal matrices, where $\tilde{V}_m^{\perp}$ and $\tilde{W}_m^{\perp}$ are orthonormal basis matrices for the orthogonal complement subspaces of the subspaces span$\{\tilde{V}_m\}$ and span$\{\tilde{W}_m\}$, respectively. If there is no ambiguity, we still use the notation E to denote the perturbation matrix. By Theorem 2,

$$E = -\tilde{W} \begin{bmatrix} \mathbf{0} & (S_{m+1}Q_m)^{\mathrm{T}}\tilde{V}_m^{\perp} \\ (\tilde{W}_m^{\perp})^{\mathrm{T}}R_{m+1}P_m & Z \end{bmatrix} \tilde{V}^{\mathrm{T}} \tag{29}$$

are solutions to the following problem for any matrix $Z \in \mathbb{R}^{(n-m)\times(n-m)}$:

$$\begin{cases} (A+E)\tilde{V}_m = \tilde{V}_{m+1}\tilde{H} \\ (A+E)^{\mathrm{T}}\tilde{W}_m = \tilde{W}_{m+1}\tilde{K}. \end{cases} \tag{30}$$

Specifically, if we let

$$\begin{aligned} \underline{E} &= -R_{m+1}P_m\tilde{V}_m^{\mathrm{T}} - \tilde{W}_m(S_{m+1}Q_m)^{\mathrm{T}} \\ &= -R_{m+1}P_mP_m^{\mathrm{T}}V_{m+1}^{\mathrm{T}} - W_{m+1}Q_mQ_m^{\mathrm{T}}S_{m+1}^{\mathrm{T}}, \end{aligned}$$

then it is a solution to (30), with

$$\|\underline{E}\|_F = \min_{Z\in\mathbb{R}^{(n-m)\times(n-m)}} \|E\|_F = \sqrt{\|R_{m+1}P_m\|_F^2 + \|S_{m+1}Q_m\|_F^2}. \tag{31}$$

Indeed, if we set P,Q to be the identity matrix, then $R_{m+1}P_m$ and $S_{m+1}Q_m$ reduce to the $\hat{R}_m$ and $\hat{S}_m$ that appeared in Theorem 2, respectively.

Next we consider how to choose new orthonormal basis matrices for $\mathcal{K}, \mathcal{L}$, such that $\|E\|_F$ is as small as possible and $\mathcal{K}, \mathcal{L}$ are Krylov subspaces with respect to $A+E$ and $(A+E)^{\mathrm{T}}$, respectively. We first need the following lemma.

Lemma 1 (30, corollary 3.3). *An orthonormal matrix U spans a Krylov subspace of A if and only if the matrix*

$$S = AU - U(U^{\mathrm{H}}AU).$$

has rank not greater than 1.

The following theorem shows how to choose the new basis matrices appropriately:

Theorem 3. *Let $\rho_1 \geq \rho_2 \geq \ldots \geq \rho_{m+1}$ and let $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{m+1}$ be the singular values of R_{m+1} and S_{m+1} defined in (23) and (24), respectively. Denote by $\tilde{P} = [\tilde{P}_m,\ \tilde{\mathbf{p}}_1]$ and by $\tilde{Q} = [\tilde{Q}_m,\ \tilde{\mathbf{q}}_1]$, where $\tilde{\mathbf{p}}_1, \tilde{\mathbf{q}}_1$ are the right singular vectors corresponding to the largest singular values ρ_1 and σ_1 of R_{m+1} and S_{m+1}, and $\tilde{P}_m$ and $\tilde{Q}_m$ are composed of the right singular vectors corresponding to the second to the $(m+1)$th singular values of R_{m+1} and S_{m+1}, respectively. Let*

$$\tilde{E} = -(R_{m+1}\tilde{P}_m)(V_{m+1}\tilde{P}_m)^{\mathrm{T}} - (W_{m+1}\tilde{Q}_m)(S_{m+1}\tilde{Q}_m)^{\mathrm{T}}, \tag{32}$$

then

$$\|\tilde{E}\|_F = \min_{\substack{P,Q\in\mathbb{R}^{(m+1)\times(m+1)} \\ P^{\mathrm{T}}P=I,\ Q^{\mathrm{T}}Q=I}} \|\underline{E}\|_F = \sqrt{(\rho_2^2 + \ldots + \rho_{m+1}^2) + (\sigma_2^2 + \ldots + \sigma_{m+1}^2)}. \tag{33}$$

Moreover, $\mathcal{K}$ and $\mathcal{L}$ are Krylov subspaces with respect to $A+\tilde{E}$ and $(A+\tilde{E})^{\mathrm{T}}$, respectively.

Proof. Recall that $\tilde{P}_m$ and $\tilde{Q}_m$ are composed of the right singular vectors corresponding to the second to the $(m+1)$th singular values of R_{m+1} and S_{m+1}. For any orthonormal matrices P,Q with P_m,Q_m being their first m columns, we have from the interlacing theorem for singular values that [40, lemma 3.3.1]

$$\|R_{m+1}P_m\|_F \geq \|R_{m+1}\tilde{P}_m\|_F = \sqrt{\rho_2^2 + \ldots + \rho_{m+1}^2}, \quad \|S_{m+1}Q_m\|_F \geq \|S_{m+1}\tilde{Q}_m\|_F = \sqrt{\sigma_2^2 + \ldots + \sigma_{m+1}^2}.$$

From (31), we have

$$\|\underline{E}\|_F \geq \|\tilde{E}\|_F = \sqrt{(\rho_2^2 + \ldots + \rho_{m+1}^2) + (\sigma_2^2 + \ldots + \sigma_{m+1}^2)}.$$

So it follows that

$$\|\tilde{E}\|_F = \min_{\substack{P,Q\in\mathbb{R}^{(m+1)\times(m+1)} \\ P^{\mathrm{T}}P=I,\ Q^{\mathrm{T}}Q=I}} \|\underline{E}\|_F, \tag{34}$$

and the last columns of P and Q are uniquely determined by the first m columns of P and Q up to scaling.

To show that $\mathcal{K}$ and $\mathcal{L}$ are Krylov subspaces of $A+\tilde{E}$ and $(A+\tilde{E})^{\mathrm{T}}$, respectively, we consider

$$\tilde{R}_{m+1} = (A+\tilde{E})\tilde{V}_{m+1} - \tilde{V}_{m+1}(\tilde{W}_{m+1}^{\mathrm{T}}\tilde{V}_{m+1})^{-1}\tilde{W}_{m+1}^{\mathrm{T}}(A+\tilde{E})\tilde{V}_{m+1} \tag{35}$$

and

$$\tilde{S}_{m+1} = (A+\tilde{E})^{\mathrm{T}}\tilde{W}_{m+1} - \tilde{W}_{m+1}(\tilde{V}_{m+1}^{\mathrm{T}}\tilde{W}_{m+1})^{-1}\tilde{V}_{m+1}^{\mathrm{T}}(A+\tilde{E})^{\mathrm{T}}\tilde{W}_{m+1}, \tag{36}$$

where $\tilde{V}_{m+1} = V_{m+1}\tilde{P}$ and $\tilde{W}_{m+1} = W_{m+1}\tilde{Q}$. According to Lemma 1, it is only necessary to show that $\tilde{R}_{m+1}$ and $\tilde{S}_{m+1}$ have rank not greater than one. We only prove that $\mathrm{rank}(\tilde{R}_{m+1}) \le 1$, and the proof of $\mathrm{rank}(\tilde{S}_{m+1}) \le 1$ is similar. Note that

$$\begin{aligned}
\tilde{R}_{m+1} &= \left(I - \tilde{V}_{m+1}(\tilde{W}_{m+1}^{\mathrm{T}}\tilde{V}_{m+1})^{-1}\tilde{W}_{m+1}^{\mathrm{T}}\right)(A+\tilde{E})\tilde{V}_{m+1} \\
&= \left(I - \tilde{V}_{m+1}(\tilde{W}_{m+1}^{\mathrm{T}}\tilde{V}_{m+1})^{-1}\tilde{W}_{m+1}^{\mathrm{T}}\right)[\tilde{V}_{m+1}\tilde{H},\ (A+\tilde{E})\tilde{\mathbf{v}}_{m+1}] \\
&= \left[\mathbf{0}, \left(I - \tilde{V}_{m+1}(\tilde{W}_{m+1}^{\mathrm{T}}\tilde{V}_{m+1})^{-1}\tilde{W}_{m+1}^{\mathrm{T}}\right)(A+\tilde{E})\tilde{\mathbf{v}}_{m+1}\right],
\end{aligned}$$

where $\tilde{\mathbf{v}}_{m+1}, \tilde{\mathbf{w}}_{m+1}$ are the last column of $\tilde{V}_{m+1}$, $\tilde{W}_{m+1}$, respectively, and we used

$$\left(I - \tilde{V}_{m+1}(\tilde{W}_{m+1}^{\mathrm{T}}\tilde{V}_{m+1})^{-1}\tilde{W}_{m+1}^{\mathrm{T}}\right)\tilde{V}_{m+1}\tilde{H} = \tilde{V}_{m+1}\tilde{H} - \tilde{V}_{m+1}\tilde{H} = \mathbf{0}.$$

Thus, $\mathrm{rank}(\tilde{R}_{m+1}) \le 1$, and $\mathcal{K}$ is a Krylov subspace with respect to $A+\tilde{E}$. ∎

Remark 2. *In Reference* 22*, the authors used only one orthonormal matrix to quasiminimize the Krylov residuals. As a comparison, we make use of two different orthonormal matrices $\tilde{P}, \tilde{Q}$ such that R_{m+1},S_{m+1} are minimized over the two subspaces $\mathcal{K}$ and $\mathcal{L}$*. On the other hand, when A is symmetric, $\mathcal{K} = \mathcal{L}$ and $V_{m+1} = W_{m+1}$, we have $R_{m+1} = S_{m+1}$. In this case, the backward error (32) reduces to

$$\tilde{E} = -(R_{m+1}P_m)(V_{m+1}P_m)^{\mathrm{T}} - (V_{m+1}P_m)(R_{m+1}P_m)^{\mathrm{T}} = -\tilde{R}_m\tilde{V}_m^{\mathrm{T}} - \tilde{V}_m\tilde{R}_m^{\mathrm{T}},$$

which is optimal for the one-sided Krylov subspace method [30, p. 85].

In summary, we have the following algorithm for computing F-norm of the perturbation matrix $\tilde{E}$:

Algorithm 1 . An algorithm for evaluating F-norm of the perturbation matrix $\tilde{E}$ in two-sided Krylov subspace methods

1. *Given two subspaces $\mathcal{K}$ and $\mathcal{L}$ of dimension $m+1$, say, generated by running the m-step two-sided Lanczos process or the m-step two-sided Arnoldi process. Let V_{m+1} and W_{m+1} be any orthonormal basis matrices for $\mathcal{K}$ and $\mathcal{L}$, respectively, such that $W_{m+1}^{\mathrm{T}}V_{m+1}$ is nonsingular.*
2. *Compute R_{m+1} and S_{m+1} by using (20) and (21).*
3. *Compute the singular values $\rho_1 \ge \rho_2 \ge \ldots \ge \rho_{m+1}$ of R_{m+1}, and the singular values $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_{m+1}$ of S_{m+1}.*
4. *Let $\|\tilde{E}\|_F = \sqrt{\rho_2^2 + \ldots + \rho_{m+1}^2 + \sigma_2^2 + \ldots + \sigma_{m+1}^2}$.*

Recall that there is no need to compute or form the perturbation matrix $\tilde{E}$ explicitly in practice. The main overhead of this algorithm includes calculating the residuals R_{m+1} and S_{m+1}, as well as computing the singular values of the two $n\times(m+1)$ matrices. The total cost in $\mathcal{O}(nm^2)$ flops, where $m \ll n$. Notice that in the two-sided Lanczos method, a pair of biorthonormal basis matrices are constructed rather than orthonormal. We thus orthonormalize the columns of them and use the resulting matrices as the initial guess.

3 | PICKING OPTIMAL H AND K FOR CHOSEN BASIS MATRICES

As was pointed out in Section 2, the choices of H and K are closely related to the norm of the perturbation matrix E. From (11) and (12), we see that

$$\tilde{H} = (\tilde{W}_{m+1}^{\mathrm{T}}\tilde{V}_{m+1})^{-1}\tilde{W}_{m+1}^{\mathrm{T}}A\tilde{V}_m \quad \text{and} \quad \tilde{K} = (\tilde{V}_{m+1}^{\mathrm{T}}\tilde{W}_{m+1})^{-1}\tilde{V}_{m+1}^{\mathrm{T}}A^{\mathrm{T}}\tilde{W}_m$$

are not optimal for the two given basis matrices $\tilde{V}_{m+1}$ and $\tilde{W}_{m+1}$; refer to (35) and (36). In this section, we focus on picking optimal H and K for fixed basis matrices $\tilde{V}_{m+1}$ and $\tilde{W}_{m+1}$, such that the F-norm of the perturbation matrix E is as small as possible. From now on, we get rid of the tilde for ease of notation, say, we denote by V_{m+1} and W_{m+1} the two basis matrices $\tilde{V}_{m+1}$ and $\tilde{W}_{m+1}$, respectively.

Define the set $\mathbb{M}$ as

$$\mathbb{M} = \{(H,K) |\ H,K \in \mathbb{R}^{(m+1)\times m},\ (W_m^{\mathrm{T}} V_{m+1})H = K^{\mathrm{T}}(W_{m+1}^{\mathrm{T}} V_m)\}. \tag{37}$$

Then the problem considered in this section can be rewritten as finding a matrix $\check{E}$ of size n-by-n, such that

$$\|\check{E}\|_F = \arg \min_{\substack{(H,K)\in\mathbb{M},\\ Z\in\mathbb{R}^{(n-m)\times(n-m)}}} \|E\|_F, \tag{38}$$

where E has the form of (6). We solve (38) by using the tool of matrix derivative.[40,41] To do this, the following lemma is needed:

Lemma 2. *Let A,B,C, and M be real valued matrices of compatible sizes, and suppose furthermore that M is symmetric positive semidefinite. Then the following statements hold* [*]*:*

$$\frac{\partial}{\partial X}\mathrm{tr}(AXB) = A^{\mathrm{T}}B^{\mathrm{T}} \tag{39}$$

$$\frac{\partial}{\partial X}\mathrm{tr}((C-AXB)^{\mathrm{T}}(C-AXB)) = 2A^{\mathrm{T}}(AXB-C)B^{\mathrm{T}} \tag{40}$$

$$\frac{\partial}{\partial X}\mathrm{tr}((C-AXB)^{\mathrm{T}}M(C-AXB)) = 2A^{\mathrm{T}}M(AXB-C)B^{\mathrm{T}}. \tag{41}$$

Proof. The Equations (39) and (40) are from the cyclic property of the trace and equations (101) and (119) of Reference 41, respectively, and (41) can be reduced to (40) by using the matrix square root $M = (M^{\frac{1}{2})T}M^{\frac{1}{2}}$. ∎

The theorem as follows establishes an optimal bound for the backward perturbation matrix for two-sided Krylov subspace methods. The key technology is to make use of the Lagrange multiplier method.

Theorem 4. *Let V_{m+1} and W_{m+1} be the chosen orthonormal basis matrices of $\mathcal{K}$ and $\mathcal{L}$, respectively, and suppose that $W_{m+1}^{\mathrm{T}}V_{m+1}$ is nonsingular. Denote by $C = V_{m+1}^{\mathrm{T}}\left(A-(I-W_mW_m^{\mathrm{T}})A(I-V_mV_m^{\mathrm{T}})\right)W_{m+1}$ and by $B_{i,j} = W_i^{\mathrm{T}}V_j$ for $i,j\in\{m,m+1\}$. Then* [†]

(i) The solution $\check{H}$ to the matrix equation

$$H-(I-B_{m,m+1}^{\mathrm{T}}B_{m,m+1})H(I-B_{m+1,m}^{\mathrm{T}}B_{m+1,m}) = CB_{m+1,m} \tag{42}$$

is unique;

(ii) The solution $\check{K}$ to the matrix equation

$$K-(I-B_{m+1,m}B_{m+1,m}^{\mathrm{T}})K(I-B_{m,m+1}B_{m,m+1}^{\mathrm{T}}) = (B_{m,m+1}C)^{\mathrm{T}} \tag{43}$$

is unique;

(iii) Denote $\check{R} = AV_m - V_{m+1}\check{H}$ and $\check{S} = A^{\mathrm{T}}W_m - W_{m+1}\check{K}$, then

$$\check{E} = -\check{R}V_m^{\mathrm{T}} - W_m\check{S}^{\mathrm{T}} + W_m\check{S}^{\mathrm{T}}V_mV_m^{\mathrm{T}} \tag{44}$$

is an optimal solution to (4) for F-norm, with

$$\|\check{E}\|_F = \min_{\substack{(H,K)\in\mathbb{M},\\ Z\in\mathbb{R}^{(n-m)\times(n-m)}}} \|E\|_F = \sqrt{\|\check{R}\|_F^2 + \|\check{S}\|_F^2 - \|W_m^{\mathrm{T}}\check{R}\|_F^2}. \tag{45}$$

[*]We thank a reviewer for the concise proof.
[†]We thank a reviewer for the concise proof of (i).

Proof. (i) and (ii): We only prove (i), and the proof of (ii) is in a similar way. Let $E(H,K)$ be a solution to (4) under the chosen basis matrices, where $(H,K) \in \mathbb{M}$. To get the minimum $E(H,K)$ in terms of F-norm, we note from (6) that Z must be a zero matrix, moreover,

$$\|E(H,K)\|_F^2 = \|R(H)\|_F^2 + \|S(K)\|_F^2 - \|W_m^{\mathrm{T}} R(H)\|_F^2, \tag{46}$$

where $R(H) = AV_m - V_{m+1}H$ and $S(K) = A^{\mathrm{T}}W_m - W_{m+1}K$. Next we seek the optimal H and K by using the Lagrange multiplier method. Define an auxiliary matrix function $\mathcal{L}(H,K,\Lambda)$ as

$$\begin{aligned}\mathcal{L}(H,K,\Lambda) &= \frac{1}{2}(\|R(H)\|_F^2 + \|S(K)\|_F^2 - \|W_m^{\mathrm{T}} R(H)\|_F^2) + \mathrm{tr}\left(\Lambda(B_{m,m+1}H - K^{\mathrm{T}}B_{m+1,m})\right) \\ &= \frac{1}{2}\mathrm{tr}(R(H)^{\mathrm{T}}(I - W_m W_m^{\mathrm{T}})R(H) + S(K)^{\mathrm{T}}S(K)) + \mathrm{tr}\left(\Lambda(B_{m,m+1}H - K^{\mathrm{T}}B_{m+1,m})\right).\end{aligned}$$

Using Lemma 2 to compute the partial derivatives of $\mathcal{L}$, and setting the derivatives to zero, gives

$$(I - B_{m,m+1}^{\mathrm{T}}B_{m,m+1})H + B_{m,m+1}^{\mathrm{T}}\Lambda^{\mathrm{T}} = V_{m+1}^{\mathrm{T}}(I - W_m W_m^{\mathrm{T}})AV_m \tag{47}$$

$$K = W_{m+1}^{\mathrm{T}}A^{\mathrm{T}}W_m + B_{m+1,m}\Lambda \tag{48}$$

$$B_{m,m+1}H = K^{\mathrm{T}}B_{m+1,m}. \tag{49}$$

By combining (48) and (49), we see that

$$\begin{aligned}(\Lambda^{\mathrm{T}}B_{m+1,m}^{\mathrm{T}})B_{m+1,m} &= K^{\mathrm{T}}B_{m+1,m} - W_m^{\mathrm{T}}AW_{m+1}B_{m+1,m} \\ &= B_{m,m+1}H - W_m^{\mathrm{T}}AW_{m+1}B_{m+1,m}.\end{aligned}$$

Postmultiplying (47) by the (nonsingular) matrix $B_{m+1,m}^{\mathrm{T}}B_{m+1,m}$ and plugging in the above equation, yields

$$\begin{aligned}&(I - B_{m,m+1}^{\mathrm{T}}B_{m,m+1})HB_{m+1,m}^{\mathrm{T}}B_{m+1,m} + B_{m,m+1}^{\mathrm{T}}B_{m,m+1}H \\ &\quad = V_{m+1}^{\mathrm{T}}(I - W_m W_m^{\mathrm{T}})AV_m B_{m+1,m}^{\mathrm{T}}B_{m+1,m} + B_{m,m+1}^{\mathrm{T}}W_m^{\mathrm{T}}AW_{m+1}B_{m+1,m},\end{aligned}$$

which simplifies to

$$H - (I - B_{m,m+1}^{\mathrm{T}}B_{m,m+1})H(I - B_{m+1,m}^{\mathrm{T}}B_{m+1,m}) = CB_{m+1,m}. \tag{50}$$

The above matrix equation is a generalized Sylvester matrix equation; one refers to Reference 41 for this type of problem. Next we show that (50) has a unique solution by using singular value decomposition. Let $B_{m,m+1} = U_1\Sigma_1 V_1^{\mathrm{T}}$ and let $B_{m+1,m} = U_2\Sigma_2 V_2^{\mathrm{T}}$ be their (full) singular value decompositions, then (50) is equivalent to

$$H - (I - V_1\Sigma_1^{\mathrm{T}}\Sigma_1 V_1^{\mathrm{T}})H(I - V_2\Sigma_2^{\mathrm{T}}\Sigma_2 V_2^{\mathrm{T}}) = CU_2\Sigma_2 V_2^{\mathrm{T}}.$$

Premultiplying the above equality by V_1^{T} and postmultiplying by V_2, we get

$$(V_1^{\mathrm{T}}HV_2) - (I - \Sigma_1^{\mathrm{T}}\Sigma_1)(V_1^{\mathrm{T}}HV_2)(I - \Sigma_2^{\mathrm{T}}\Sigma_2) = V_1^{\mathrm{T}}CU_2\Sigma_2.$$

The above matrix equation can be rewritten as

$$\left(I - (I - \Sigma_2^{\mathrm{T}}\Sigma_2) \otimes (I - \Sigma_1^{\mathrm{T}}\Sigma_1)\right)\mathrm{vec}(V_1^{\mathrm{T}}HV_2) = \mathrm{vec}(V_1^{\mathrm{T}}CU_2\Sigma_2), \tag{51}$$

where $\otimes$ is the Kronecker product and $\mathrm{vec}(X)$ represents the vector obtained from stacking the columns of a matrix X.

As $B_{m,m+1}^{\mathrm{T}}B_{m,m+1}$ and $B_{m+1,m}^{\mathrm{T}}B_{m+1,m}$ are real symmetric, their eigenvalues are real. Recall that both V_{m+1} and W_{m+1} are orthonormal, so all the eigenvalues lie in [0,1]. Moreover, $B_{m+1,m+1}$ is nonsingular, which implies $B_{m+1,m}$ is full rank and $B_{m+1,m}^{\mathrm{T}}B_{m+1,m}$ is nonsingular. Consequently, all diagonal elements of $(I - \Sigma_2^{\mathrm{T}}\Sigma_2)$ line in [0,1), hence the matrix

$I-(I-\Sigma_2^{\mathrm{T}}\Sigma_2)\otimes(I-\Sigma_1^{\mathrm{T}}\Sigma_1)$ is nonsingular, and the solution to (51) is unique. In other words, the solution to (42) is unique. Notice that $I-(I-\Sigma_2^{\mathrm{T}}\Sigma_2)\otimes(I-\Sigma_1^{\mathrm{T}}\Sigma_1)$ is diagonal, one can solve the above nonsingular system without forming the coefficient matrix explicitly.

(iii) Let $\check{H}$ and $\check{K}$ be the solutions to (42) and (43), respectively. Recall that they satisfy the relation

$$W_m^{\mathrm{T}}V_{m+1}\check{H}=\check{K}^{\mathrm{T}}W_{m+1}^{\mathrm{T}}V_m.$$

Thus,

$$\check{E}=-W\begin{bmatrix}W_m^{\mathrm{T}}\check{R} & \check{S}^{\mathrm{T}}V_m^{\perp}\\ (W_m^{\perp})^{\mathrm{T}}\check{R} & \mathbf{0}\end{bmatrix}V^{\mathrm{T}}=-\check{R}V_m^{\mathrm{T}}-W_m\check{S}^{\mathrm{T}}+W_m\check{S}^{\mathrm{T}}V_mV_m^{\mathrm{T}}$$

is a solution to (4) and

$$\|\check{E}\|_F^2=\|\check{R}\|_F^2+\|\check{S}\|_F^2-\|W_m^{\mathrm{T}}\check{R}\|_F^2.$$

As $\|\cdot\|_F^2$ is a convex function, we have

$$\|\check{E}\|_F=\min_{\substack{(H,K)\in\mathbb{M}_{\perp}\\ Z\in\mathbb{R}^{(n-m)\times(n-m)}}}\|E\|_F.$$

■

In summary, we have the following algorithm for evaluating the quality of the computed two-sided Krylov subspaces. The key is to first choose appropriate basis matrices by using Algorithm 1, and then minimize $\|E\|_F$ by picking "optimal" H and K.

Algorithm 2 . An algorithm for computing F-norm of the perturbation matrix $\check{E}$ in two-sided Krylov subspace methods

1. Given two subspaces $\mathcal{K}$ and $\mathcal{L}$ of dimension $m+1$, say, generated by the m-step two-sided Lanczos process or the m-step two-sided Arnoldi process. Let V_{m+1} and W_{m+1} be orthonormal basis matrices for $\mathcal{K}$ and $\mathcal{L}$, respectively, such that $V_{m+1}^{\mathrm{T}}W_{m+1}$ is nonsingular;

2. Form the new basis matrices: Calculate R_{m+1} and S_{m+1} by using (20) and (21), and compute their economized singular value decompositions. Let $P=[P_m,\mathbf{p}_1]$ and $Q=[Q_m,\mathbf{q}_1]$, where $\mathbf{p}_1,\mathbf{q}_1$ are the right singular vectors corresponding to the largest singular values ρ_1 and σ_1 of R_{m+1} and S_{m+1}, and P_m and Q_m are composed of the right singular vectors corresponding to the second to the $(m+1)$th singular values of R_{m+1} and S_{m+1}, respectively. Let $V_{m+1}=V_{m+1}P$ and $W_{m+1}=W_{m+1}Q$;

3. Solve the matrix Equations (39) and (40) for $\check{H}$ and $\check{K}$, say, by using singular value decomposition;

4. Compute $\check{R}=AV_m-V_{m+1}\check{H}$ and $\check{S}=A^{\mathrm{T}}W_m-W_{m+1}\check{K}$;

5. Let $\|\check{E}\|_F=\sqrt{\|\check{R}\|_F^2+\|\check{S}\|_F^2-\|W_m^{\mathrm{T}}\check{R}\|_F^2}$.

Remark 3. Let us briefly explain the differences between the proposed algorithms. Indeed, Algorithms 1 and 2 share the same (chosen) basis matrices. The difference is that they use different H and K: Algorithm 1 uses $\tilde{H}$ and $\tilde{K}$ defined in (26), while Algorithm 2 uses the "optimal" matrices $\check{H}$ and $\check{K}$ which are solution of (42) and (43). Thus, we have from (38) that $\|\check{E}\|_F\le\|\tilde{E}\|_F$, that is, $\check{E}$ is at least as good as $\tilde{E}$ theoretically. Algorithm 2 first does a few steps from Algorithm 1, and then computes the optimal H and K. One can also do it the other way around. For example, first picking H and K, and then trying to adjust V_{m+1} and W_{m+1} to reduce $\|E\|_F$. Indeed, this is just the scheme used in Algorithm 1.

On the other hand, one can also combine Algorithm 1 with Algorithm 2 together. More precisely, given two initial basis matrices V_{m+1} and W_{m+1}, we run Algorithm 1 to get a pair of "optimal" basis matrices, and then run Algorithm 2 to seek the "optimal" H and K, and turn to Algorithm 1. The algorithm will be stopped as soon as the perturbation bounds from Algorithm 1 to 2 are close to each other. However, as the two subspaces are fixed, we find experimentally that the improvement is marginal.

The main overhead for Algorithm 2 is $\mathcal{O}(nm^2)$ flops, and it is a little higher than that of Algorithm 1. More precisely, we need $\mathcal{O}(nm^2)$ flops for choosing the basis matrices, and another $\mathcal{O}(nm^2)$ flops for forming $B_{m,m+1}$, $B_{m+1,m}$, $\check{R}$, and $\check{S}$; as well as $\mathcal{O}(m^3)$ for the singular value decompositions and $\mathcal{O}(m^2)$ for solving (42) and (43). As $m\ll n$, the main cost of this algorithm is $\mathcal{O}(nm^2)$ flops.

TABLE 1 The test matrices used in numerical experiments

Matrix	n	Type	Background
b2_ss	1089	Nonsymmetric real	Chemical process simulation problem
mahindas	1258	Nonsymmetric real	Economic problem
olm1000	1000	Nonsymmetric real	Computational fluid dynamics problem
radfr1	1048	Nonsymmetric real	Chemical process simulation problem
sherman2	1080	Nonsymmetric real	Computational fluid dynamics problem
west0989	989	Nonsymmetric real	Chemical process simulation problem

On the other hand, recall that in some two-sided Krylov subspace methods such as the two-sided Lanczos method, a pair of biorthonormal basis matrices V_{m+1} and W_{m+1} are constructed rather than orthonormal.[7,9,10,12,20-22] In this situation, we can orthonormalize the columns of V_{m+1} and W_{m+1} (say, by applying the economized QR decomposition), respectively, to get a pair of orthonormal basis matrices for Algorithms 1 and 2.

4 | NUMERICAL EXPERIMENTS

The aim of this article is to provide an improved way for determining backward perturbation bound of approximate two-sided Krylov subspaces $\mathcal{K}$ and $\mathcal{L}$. In this section, we perform some numerical experiments to illustrate the effectiveness of the proposed strategies. To this aim, the two subspaces $\mathcal{K}$ and $\mathcal{L}$ are generated by using the two-sided Lanczos method[9,12] or the two-sided Arnoldi method[26] on $A+E_0$ and $(A+E_0)^{\mathrm{T}}$, where A is a test matrix and E_0 is a random matrix constructed by the following MATLAB commands:

```
normA=norm(A,'fro');
E0=sprandn(A);
E0=E0/norm(E0,'fro');
E0=t*normA*E0;
```

where $t = \|E_0\|_F / \|A\|_F$ are chosen as 10^{-2},10^{-6},10^{-10}, and 10^{-14}, respectively. Thus, in exact arithmetic, the subspaces $\mathcal{K}$ and $\mathcal{L}$ are the same for the three algorithms, provided that the same initial vectors are utilized.

Both the work of Wu et al[22] and that of this article consider how to determine a perturbation E whose norm is as small as possible, such that $\mathcal{K}, \mathcal{L}$ are Krylov subspaces of $A+E$ and $(A+E)^{\mathrm{T}}$, respectively. Although they are different algorithms altogether, we compare the algorithm presented in Reference 22, and the two algorithms proposed in this article, for evaluating the quality of the computed two-sided Krylov subspaces. In all the tables below, we denote by $\|E_W\|_F$, $\|E_1\|_F$, and $\|E_2\|_F$ the F-norm of the perturbation matrices obtained from running Algorithm 2.1 in Reference 22, Algorithms 1 and 2, respectively. Note that the smaller the norm, the better the algorithm will be. Thus, the performance of the algorithms are evaluated by the perturbation bounds in all the experiments.

There are six test matrices that are from the SuiteSparse Matrix Collection [3]. Table 1 lists the size, type and background of these matrices in detail. All the numerical experiments are run on a PC with Intel(R) Core(TM) i3-7100 CPU@3.90 GHz and 4 GB RAM using MATLAB R2015b.

Example 1. In this example, we apply the proposed strategies on the approximate Krylov subspaces generated by using the two-sided Lanczos method. The two subspaces $\mathcal{K}$ and $\mathcal{L}$ are built by running the two-sided Lanczos procedure on $A+E_0$ and $(A+E_0)^{\mathrm{T}}$ with steps $m=20$, where the (unit) initial vectors $\mathbf{v}_1 = \mathbf{w}_1$ are generated sssby the MATLAB command *randn*(n,1). The test matrices are `mahindas`, `radfr1`, and `sherman2`. Recall that in the two-sided Lanczos method, a pair of biorthonormal basis matrices V_{m+1} and W_{m+1} are constructed rather than orthonormal basis matrices. We thus orthonormalize the columns of them and use the resulting matrices as the initial guess for Algorithms 1 and 2. Table 2 lists the numerical results obtained, where κ_1 and κ_2 denote the 2-norm condition numbers of $W_{m+1}^{\mathrm{T}} V_{m+1}$ (where

[3]https://sparse.tamu.edu/

TABLE 2 Example 1. A comparison of $\|E_W\|_F$, $\|E_1\|_F$ and $\|E_2\|_F$ obtained from running algorithm 2.1 in Reference 22, Algorithms 1 and 2

Matrix	$t = \frac{\|E_0\|_F}{\|A\|_F}$	10^{-2}	10^{-6}	10^{-10}	10^{-14}
	$\frac{\|E_W\|_F}{\|A\|_F}$	1.57×10^{-1}	7.67×10^{-2}	4.13×10^{-4}	4.12×10^{-8}
	$\frac{\|E_1\|_F}{\|A\|_F}$	1.28×10^{-5}	2.47×10^{-5}	5.52×10^{-10}	5.56×10^{-14}
mahindas	$\frac{\|E_2\|_F}{\|A\|_F}$	$\mathbf{2.80\times10^{-6}}$	$\mathbf{2.70\times10^{-7}}$	$\mathbf{3.72\times10^{-11}}$	$\mathbf{2.40\times10^{-14}}$
	κ_1	2.66×10^{3}	6.53×10^{4}	1.97×10^{3}	1.97×10^{3}
	κ_2	5.39×10^{6}	1.99×10^{9}	2.91×10^{6}	2.90×10^{6}
	$\frac{\|E_W\|_F}{\|A\|_F}$	2.54×10^{1}	3.13×10^{-3}	3.37×10^{-6}	6.02×10^{-6}
	$\frac{\|E_1\|_F}{\|A\|_F}$	4.19×10^{-2}	1.36×10^{-5}	1.35×10^{-9}	1.41×10^{-10}
radfr1	$\frac{\|E_2\|_F}{\|A\|_F}$	$\mathbf{7.62\times10^{-3}}$	$\mathbf{8.80\times10^{-7}}$	$\mathbf{8.60\times10^{-11}}$	$\mathbf{8.05\times10^{-12}}$
	κ_1	1.45×10^{3}	1.20×10^{3}	1.20×10^{3}	1.20×10^{3}
	κ_2	3.60×10^{5}	1.44×10^{6}	1.44×10^{6}	1.44×10^{6}
	$\frac{\|E_W\|_F}{\|A\|_F}$	4.04×10^{1}	1.86×10^{-2}	2.12×10^{-6}	2.05×10^{-10}
	$\frac{\|E_1\|_F}{\|A\|_F}$	1.96×10^{-2}	1.79×10^{-5}	1.96×10^{-9}	2.00×10^{-13}
sherman2	$\frac{\|E_2\|_F}{\|A\|_F}$	$\mathbf{9.14\times10^{-3}}$	$\mathbf{3.50\times10^{-6}}$	$\mathbf{2.14\times10^{-10}}$	$\mathbf{6.18\times10^{-14}}$
	κ_1	5.40×10^{2}	1.41×10^{3}	1.59×10^{3}	1.59×10^{3}
	κ_2	5.83×10^{4}	1.71×10^{6}	2.28×10^{6}	2.28×10^{6}

Note: In all the algorithms, the subspaces $\mathcal{K}$ and $\mathcal{L}$ are generated by running the two-sided Lanczos process on $A+E_0$ and $(A+E_0)^{\mathrm{T}}$ with step size $m=20$, and the values in bold are the smallest one in the same column. Here, κ_1 and κ_2 denote the 2-norm condition numbers of $W_{m+1}^{\mathrm{T}}V_{m+1}$ (where V_{m+1} and W_{m+1} are the chosen basis matrices) and the (maximal) 2-norm condition numbers of the linear systems arising from the generalized Sylvester Equations (42) and (43), respectively.

V_{m+1} and W_{m+1} are the chosen basis matrices in Algorithms 1 and 2), and the (maximal) 2-norm condition numbers of the linear systems arising from the generalized Sylvester Equations (42) and (43), respectively.

It is seen from Table 2 that the norms of E obtained from Algorithms 1 and 2 are much smaller than those from algorithm 2.1 of Reference 22, while those from Algorithm 2 are the smallest. That is, for the same E_0, the values of $\|E_1\|_F$ and $\|E_2\|_F$ are much smaller than those of $\|E_W\|_F$, and $\|E_2\|_F$ is the smallest among the three. Therefore, the proposed methods are much better than the one advocated in [22, algorithm 2.1], and they are promising for evaluating the quality of the computed two-sided Krylov subspaces. Indeed, the strategy used in [22, algorithm 2.1] are based on a pair of biorthonormal basis matrices, which may result in a large backward perturbation bound in practice. As a comparison, our strategies work very well.

More interestingly, we observe that $\|E_1\|_F$ and $\|E_2\|_F$ are even smaller than $\|E_0\|_F$ as $t=\frac{\|E_0\|_F}{\|A\|_F}=10^{-2}$, 10^{-6}, and 10^{-10} for the `mahindas` and the `radfr1` matrices. The reason is that E_0 is not the globally optimal perturbation matrix, so the F-norms of E_1 and E_2 can be smaller than that of E_0. To show the effectiveness of our strategies more precisely, in Figure 1, we plot the curves of the relative errors $\frac{\|E\|_F}{\|A\|_F}$ obtained from algorithm 2.1 in Reference 22 (Wu et al), Algorithm 1 (Alg.1) and Algorithm 2 (Alg.2), with $t=10^{-1},10^{-2},\dots,10^{-14}$, for the `mahindas` matrix and the `radfr1` matrix. It is obvious to see that our backward error bounds are much sharper than those from Reference 22.

Example 2. In this example, the approximate Krylov subspaces are generated by running the two-sided Arnoldi method with step size $m=20$ for Algorithms 1 and 2. In algorithm 2.1 of Reference 22, the two-sided Lanczos procedure with step size $m=20$ is applied. For the sake of justification, we choose the same (unit) initial vector $\mathbf{v}_1=\mathbf{w}_1$ for the two-sided Arnoldi and the two-sided Lanczos procedures, generated by using the MATLAB command *randn*(n,1). Hence, the two spaces $\mathcal{K}$ and $\mathcal{L}$ used by the two-sided Arnoldi method and the two-sided Lanczos method will be the same in exact arithmetic. Note that the two-sided Arnoldi method generates a pair of orthonormal basis matrices, while the two-sided Lanczos method generates a pair of biorthonormal basis matrices. The test matrices are the `west0989`, `b2_ss` and the `olm1000` matrices. Table 3 list the numerical results, where κ_1 and κ_2 denote the 2-norm condition numbers of $W_{m+1}^{\mathrm{T}}V_{m+1}$

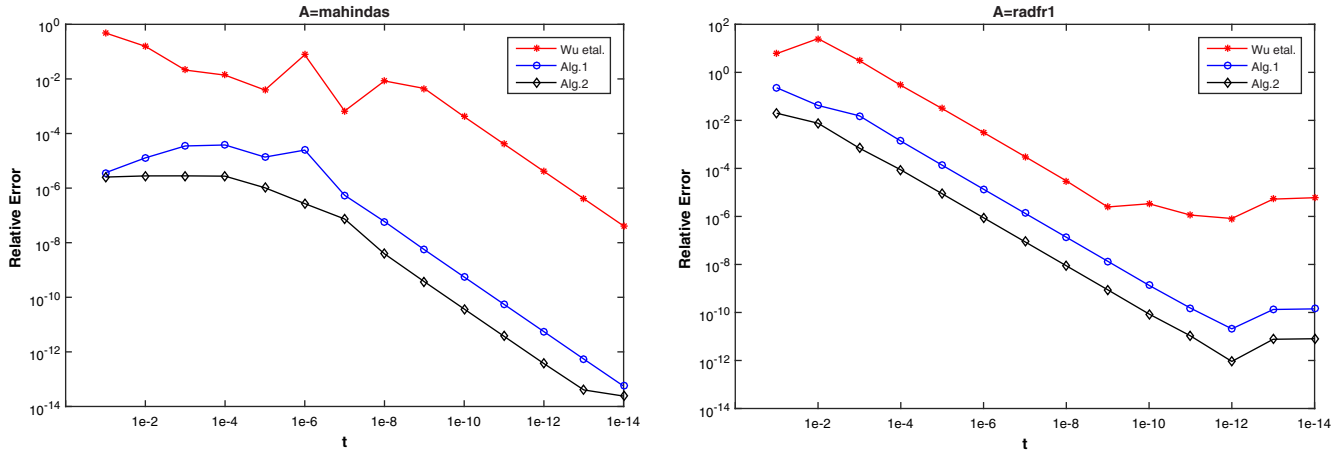

FIGURE 1 Example 1: Curves of the relative errors $\frac{\|E\|_F}{\|A\|_F}$ obtained from algorithm 2.1 in Reference 22 (Wu et al), Algorithm 1 (Alg.1) and Algorithm 2 (Alg.2), with $t = 10^{-1}, 10^{-2}, \ldots, 10^{-14}$, $m = 20$. The `mahindas` matrix (left) and the `radfr1` matrix (right)

TABLE 3 Example 2. A comparison of $\|E_W\|_F$, $\|E_1\|_F$, and $\|E_2\|_F$ obtained from running algorithm 2.1 in Reference 22, Algorithms 1 and 2

Matrix	$\frac{\|E_0\|_F}{\|A\|_F}$	10^{-2}	10^{-6}	10^{-10}	10^{-14}
	$\frac{\|E_W\|_F}{\|A\|_F}$	2.20×10^{0}	8.58×10^{1}	7.93×10^{-3}	8.01×10^{-7}
	$\frac{\|E_1\|_F}{\|A\|_F}$	1.04×10^{-2}	1.05×10^{-3}	1.03×10^{-7}	$\mathbf{1.04 \times 10^{-11}}$
west0989	$\frac{\|E_2\|_F}{\|A\|_F}$	$\mathbf{2.45 \times 10^{-3}}$	$\mathbf{5.68 \times 10^{-6}}$	$\mathbf{6.18 \times 10^{-9}}$	5.13×10^{-9}
	κ_1	1.46×10^{3}	3.58×10^{6}	3.36×10^{6}	3.36×10^{6}
	κ_2	1.43×10^{6}	3.07×10^{12}	3.49×10^{12}	3.49×10^{12}
	$\frac{\|E_W\|_F}{\|A\|_F}$	1.62×10^{2}	1.74×10^{4}	1.39×10^{0}	4.20×10^{-2}
	$\frac{\|E_1\|_F}{\|A\|_F}$	1.76×10^{-1}	3.61×10^{-3}	7.14×10^{-7}	$\mathbf{6.77 \times 10^{-11}}$
b2_ss	$\frac{\|E_2\|_F}{\|A\|_F}$	$\mathbf{4.67 \times 10^{-3}}$	$\mathbf{1.45 \times 10^{-3}}$	$\mathbf{1.92 \times 10^{-7}}$	9.56×10^{-10}
	κ_1	2.75×10^{4}	2.36×10^{7}	1.08×10^{7}	1.08×10^{7}
	κ_2	2.37×10^{8}	2.38×10^{13}	5.18×10^{13}	5.18×10^{13}
	$\frac{\|E_W\|_F}{\|A\|_F}$	6.59×10^{0}	5.58×10^{-3}	5.47×10^{-7}	9.65×10^{-11}
	$\frac{\|E_1\|_F}{\|A\|_F}$	3.32×10^{-2}	4.29×10^{-6}	4.33×10^{-10}	4.33×10^{-14}
olm1000	$\frac{\|E_2\|_F}{\|A\|_F}$	$\mathbf{6.43 \times 10^{-3}}$	$\mathbf{8.10 \times 10^{-7}}$	$\mathbf{8.20 \times 10^{-11}}$	$\mathbf{1.68 \times 10^{-14}}$
	κ_1	3.43×10^{2}	9.45×10^{2}	9.60×10^{2}	9.60×10^{2}
	κ_2	1.17×10^{5}	6.80×10^{5}	7.08×10^{5}	7.08×10^{5}

Note: Here, the subspaces $\mathcal{K}$ and $\mathcal{L}$ are generated by running the two-sided Arnoldi process (Algorithms 1 and 2) and the two-sided Lanczos process (algorithm 2.1 in Reference 22) on $A + E_0$ and $(A + E_0)^{\mathrm{T}}$ with step size $m = 20$, respectively, and the values in bold are the smallest one in the same column. Here, κ_1 and κ_2 denote the 2-norm condition numbers of $W_{m+1}^{\mathrm{T}} V_{m+1}$ (where V_{m+1} and W_{m+1} are the chosen basis matrices) and the (maximal) 2-norm condition numbers of the linear systems arising from the generalized Sylvester Equations (42) and (43), respectively.

(where V_{m+1} and W_{m+1} are the chosen basis matrices in Algorithms 1 and 2), and the (maximal) 2-norm condition numbers of the linear systems arising from the generalized Sylvester Equations (42) and (43), respectively. In Figure 2, we plot the curves of the relative errors $\frac{\|E\|_F}{\|A\|_F}$ obtained from algorithm 2.1 in Reference 22 (Wu et al), Algorithm 1 (Alg.1) and Algorithm 2 (Alg.2), with $t = 10^{-1}, 10^{-2}, \ldots, 10^{-14}$, for the `west0989` matrix and the `b2_ss` matrix.

Again, we observe from Table 3 that $\|E_1\|_F$ and $\|E_2\|_F$ are much smaller than $\|E_W\|_F$, and the proposed strategies are much better than the one given in Reference 22 for evaluating the quality of the computed two-sided Krylov subspaces. In

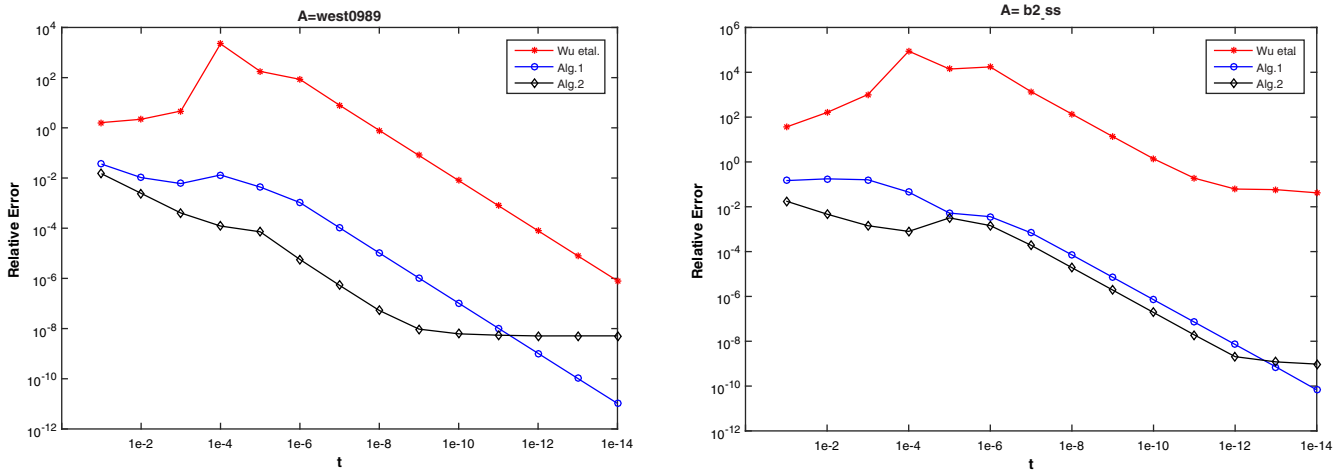

FIGURE 2 Example 2: Curves of the relative error $\frac{\|E\|_F}{\|A\|_F}$ obtained from algorithm 2.1 in Reference 22 (Wu et al), Algorithm 1 (Alg.1) and Algorithm 2 (Alg.2), with $t = 10^{-1}, 10^{-2}, \ldots, 10^{-14}$, $m = 20$. The `west0989` matrix (left) and the `b2_ss` matrix (right)

this experiment, we see that $\|E_2\|_F$ is the smallest one among the three relative errors in most cases, except for $t = 10^{-14}$. This is due to the fact that Algorithm 2 may suffer from the difficulty of ill-conditioning of the generalized Sylvester equations, however, its results are still much better than those from algorithm 2.1 in Reference 22. For instance, for the `west0989` and the `b2_ss` matrices, the condition numbers of the linear systems arising from the generalized Sylvester equations are in the order of 10^{13} as $t = 10^{-14}$. We see that the condition numbers of the linear systems are about square of the condition numbers of the matrices $W_{m+1}^{\mathrm{T}} V_{m+1}$, and we believe that the ill-conditioning of the linear systems is closely related to that of $W_{m+1}^{\mathrm{T}} V_{m+1}$. Thus, we suggest to use Algorithm 1 instead of Algorithm 2 when the generalized Sylvester matrix equations are ill-conditioned.

5 | CONCLUDING REMARKS

Two-sided Krylov subspace methods such as the two-sided Lanczos method and the two-sided Arnoldi method are popular techniques in large-scale matrix computations. However, the subspaces spanned by the computed bases in these methods may not be Krylov subspaces in practice, and it is desirable to estimate the quality of the computed Krylov subspaces efficiently. Given a matrix $A \in \mathbb{R}^{n\times n}$ and two subspaces $\mathcal{K}$ and $\mathcal{L}$ of dimension $m+1$, we consider how to determine a perturbation E whose norm is as small as possible, such that $\mathcal{K}, \mathcal{L}$ are Krylov subspaces of $A+E$ and $(A+E)^{\mathrm{T}}$, respectively.

The method proposed in Reference 22 is based on a pair of biorthonormal basis matrices of two-sided Krylov subspaces, and the perturbation bound established can be greatly overestimated in practice. In this work, we revisit this problem and make use of a pair of orthonormal basis matrices rather than biorthonormal basis matrices. Indeed, our strategies can be understood as designing new algorithms via perturbation analysis. More precisely, we propose two new algorithms for determining backward perturbation bounds of approximate two-sided Krylov subspaces. In the first one, for given H and K, we minimize $\|E\|_F$ by choosing the "optimal" basis matrices. In the second one, we first choose appropriate basis matrices, and then minimize $\|E\|_F$ by picking the "optimal" H and K. Numerical experiments show that the proposed algorithms are much better than the one given in Reference 22, and the second one is often superior to the first one.

Although Algorithm 2, which is based on the Lagrange multiplier method, is better than the first one theoretically, it may suffer from ill-conditioning of the reduced matrix equations, and Algorithm 1 is preferable in this case. How to deal with this problem effectively is definitely a part of our future work. Furthermore, as was pointed out in Section 3, the second strategy is only optimal for given orthonormal basis matrices, and it is not globally optimal when A is nonsymmetric. For two given subspaces $\mathcal{K}$ and $\mathcal{L}$, by "globally optimal," we mean that the Frobenius norm of E is minimized over all choices of Z, V_{m+1}, W_{m+1}, H, and K, where $Z \in \mathbb{R}^{(n-m)\times(n-m)}$ is arbitrary, V_{m+1}, W_{m+1} are orthonormal matrix bases of $\mathcal{K}$ and $\mathcal{L}$, and H, K satisfy (5). Indeed, it is very difficult to choose the optimal basis matrices in this approach, because

the corresponding residuals are nonlinear functions in P_m and Q_m. Thus, how to choose the optimal basis matrices in Algorithm 2, such that one can find the globally optimal backward perturbation, is still an open problem and deserves further investigation.

ACKNOWLEDGEMENTS

We would like to express our sincere thanks to the anonymous referees and our editor for insightful comments and suggestions that greatly improved the representation of this article.

CONFLICTS OF INTEREST

This work does not have any conflicts of interest.

ORCID

Gang Wu https://orcid.org/0000-0002-4936-437X

REFERENCES

1. Bai Z. Error analysis of the Lanczos algorithm for the nonsymmetric eigenvalue problem. Math Comput. 1994;62:209–226.
2. Björck Å. Numerical Methods for Least Squares Problems. Philadelphia, PA: SIAM, 1996.
3. Golub GH, Van Loan CF. Matrix Computations. 4th ed. Baltimore: The Johns Hopkins University Press, 2013.
4. Saad Y. Iterative Methods for Sparse Linear Systems. 2nd ed. Philadelphia, PA: SIAM, 2003.
5. Saad Y. Numerical Methods for Large Eigenvalue Problems. 2nd ed. Philadelphia, PA: SIAM, 2011.
6. Stewart GW. Matrix Algorithms. Vol II. Philadelphia, PA: Eigensystems, SIAM, 2001.
7. Astudillo R, de Gier J, van Gijzen M. Accelerating the induced dimension reduction method using spectral information. J Comput Appl Math. 2019;345:33–47.
8. Greenbaum A. Iterative Methods for Solving Linear Systems. Philadelphia, PA: SIAM, 1997.
9. Lanczos C. Solution of systems of linear equations by minimal iterations. J Res Natl Bur Stand. 1952;49:33–53.
10. van der Vorst HA. Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. SIAM J Sci Stat Comput. 1992;13:631–644.
11. Arnoldi WE. The principle of minimized iterations in the solution of the matrix eigenvalue problem. Quart Appl Math. 1951;9:17–29.
12. Lanczos C. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. J Res Natl Bur Stand. 1950;45:255–282.
13. Antoulas A, Beattie C, Gugercin S. Interpolatory model reduction of large-scale dynamical systems. In: Mohammadpour J, Grigoriadis K, editors. Efficient Modeling and Control of Large-Scale Systems. New York, NY: Springer-Verlag, 2010; p. 3–58.
14. Bai Z. Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. Appl Numer Math. 2002;43:9–44.
15. Strakos Z. Model reduction using the Vorobyev moment problem. Numer Alg. 2009;51:363–379.
16. Strakos Z, Tichý P. On efficient numerical approximation of the bilinear form $c^*A^{-1}b$. SIAM J Sci Comput. 2011;33:565–587.
17. Beckerman B, Kressner D, Schweitzer M. Low-rank updates of matrix functions. SIAM J Matrix Anal Appl. 2018;39:539–565.
18. Higham NJ. Functions of Matrices, Theory and Computation. Philadelphia, PA: SIAM, 2008.
19. Moler C, Van Loan CF. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM Rev. 2003;45:3–49.
20. Bai Z, Day D, Ye Q. An adaptive block Lanczos method for non-Hermitian eigenvalue problems. SIAM J Matrix Anal Appl. 1999;20:1060–1082.
21. Wu G. A dynamic thick restarted semi-refined ABLE algorithm for computing a few selected eigentriplets of large nonsymmetric matrices. Linear Algebra Appl. 2006;416:313–335.
22. Wu G, Wei Y, Jia ZG, Ling S, Zhang L. Towards backward perturbation bounds for approximate dual Krylov subspaces. BIT. 2013;53:225–239.
23. Stewart GW. On the numerical analysis of oblique projectors. SIAM J Matrix Anal Appl. 2011;32:309–348.
24. Cullum J, Zhang T. Two-sided Arnoldi and nonsymmetric Lanczos algorithms. SIAM J Matrix Anal Appl. 2002;24:303–319.
25. Lietaert P, Meerbergen K, Tisseur F. Compact two-sided Krylov methods for nonlinear eigenvalue problems. SIAM J Sci Comput. 2018;40:A2801–A2829.
26. Ruhe A. The two-sided Arnoldi algorithm for nonsymmetric eigenvalue problems. In: Pencils M, Kågstrom B, Ruhe A, editors. Berlin, Heidelberg: Springer, 1983; p. 104–120.
27. Ruhe A. The two-sided Arnoldi algorithm, Householder Symposium XIX on Numerical Linear Algebra. Vol 194. Belgium: SPA, 2014.
28. Zwaan I, Hochstenbach M. Krylov-Schur-type restarts for the two-sided Arnoldi method. SIAM J Matrix Anal Appl. 2017;38:297–321.
29. Golub GH, Meurant G. Matrices, Moments and Quadrature with Applications. Princeton, NJ: Princeton University Press, 2010.
30. Stewart GW. Backward error bounds for approximate Krylov subspaces. Linear Algebra Appl. 2002;340:81–86.
31. Bouras A, Frayssé V. Inexact matrix-vector products in Krylov methods for solving linear systems: a relaxation strategy. SIAM J Matrix Anal Appl. 2005;26:660–678.
32. Gaaf S, Simoncini V. Approximating the leading singular triplets of a large matrix function. Appl Numer Math. 2017;113:26–43.

33. Pozza S, Simoncini V. Inexact Arnoldi residual estimates and decay properties for functions of non-Hermitian matrices. BIT. 2019;59:969–986.
34. Simoncini V, Szyld D. Theory of inexact Krylov subspace methods and applications to scientific computing. SIAM J Sci Comput. 2003;25:454–477.
35. Paige CC. The Computation of Eigenvalues and Eigenvectors of Very Large Sparse Matrices [PhD thesis]. London, England: London University, Institute of Computer Science; 1971.
36. Paige CC. Practical use of the symmetric Lanczos process with reorthogonalization. BIT. 1971;10:183–195.
37. Kahan W, Parlett B, Jiang E. Residual bounds on approximate eigensystems of nonnormal matrices. SIAM J Numer Anal. 1982;19:470–484.
38. Noschese S, Reichel L. Inverse subspace problems with applications. Numer Linear Algebra Appl. 2014;21:589–603.
39. Sun J. Backward perturbation analysis of certain characteristic subspaces. Numer Math. 1993;65:357–382.
40. Horn RA, Johnson CR. Topics in Matrix Analysis. Cambridge: Cambridge University Press, 1991.
41. Petersen K, Pedersen M. The Matrix Cookbook, Version II: November 15, 2012, http://matrixcookbook.com.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Wang Y, Wu G. New strategies for determining backward perturbation bound of approximate two-sided Krylov subspaces. *Numer Linear Algebra Appl*. 2020;e2324. https://doi.org/10.1002/nla.2324