FAST INEXACT IMPLICITLY RESTARTED ARNOLDI METHOD FOR GENERALIZED EIGENVALUE PROBLEMS WITH SPECTRAL TRANSFORMATION[*]

FEI XUE[†] AND HOWARD C. ELMAN[‡]

Abstract. We study an inexact implicitly restarted Arnoldi (IRA) method for computing a few eigenpairs of generalized non-Hermitian eigenvalue problems with spectral transformation, where in each Arnoldi step (outer iteration) the matrix-vector product involving the transformed operator is performed by iterative solution (inner iteration) of the corresponding linear system of equations. We provide new perspectives and analysis of two major strategies that help reduce the inner iteration cost: a special type of preconditioner with "tuning", and gradually relaxed tolerances for the solution of the linear systems. We study a new tuning strategy constructed from vectors in both previous and the current IRA cycles, and we show how tuning is used in a new two-phase algorithm to greatly reduce inner iteration counts. We give an upper bound of the allowable tolerances of the linear systems and propose an alternative estimate of the tolerances. In addition, the inner iteration cost can be further reduced through the use of subspace recycling with iterative linear solvers. The effectiveness of these strategies is demonstrated by numerical experiments.

Key words. inexact implicitly restarted Arnoldi method, tuning, relaxation, subspace recycling

1. Introduction. Many scientific and engineering applications require a small group of eigenvalues closest to a specified shift or those with largest or smallest real parts. The shift-invert and the Cayley transformation [21] are the two most commonly used spectral transformations to map these eigenvalues to the dominant ones of the transformed operator, so that they can be readily computed by eigenvalue algorithms. The major challenge of this approach is that a linear system of equations involving a shifted matrix needs to be solved in each step (outer iteration) of the eigenvalue algorithm. For large-scale applications, for instance, finite element discretization of three-dimensional partial differential equations, this linear solve has to be done using iterative solvers (inner iteration) instead of factorization-based direct solvers. This offers the prospect of inexact eigenvalue algorithms with "inner-outer" structure, where the required solution of linear systems is computed only to a specified accuracy. This paper concerns efficient iterative solution of the linear systems of equations that arise when the inexact implicitly restarted Arnoldi (IRA) method with spectral transformation is used to detect a few eigenpairs of generalized non-Hermitian eigenvalue problems (GNHEP) $Av = \lambda Bv$.

In the past decade, considerable progress has been made in understanding inexact eigenvalue algorithms, especially the simplest one – inexact inverse iteration. Systematic study of this algorithm is mainly carried out by Spence and his collaborators (see [1, 2, 3, 12, 13, 14]). A major concern in these papers is the connection between the error of the inner solve and the convergence of the outer iteration, with different choices of variable shifts, tolerances and formulations of the linear systems. Meanwhile, there has been increasing interest in reducing the inner iteration cost to enhance the effectiveness of the algorithm. Reference [28] gives some new insights into preconditioning the linear systems arising in inexact Rayleigh quotient iteration by modifying the right hand side of the preconditioned system. This idea is extended in [1, 2, 3] and further refined in [13, 14] for inexact inverse iteration or Rayleigh quotient iteration, where a special type of preconditioner with "tuning" is constructed and analyzed. In [26], tuning is used in the iterative solution of the block linear systems arising in inexact subspace iteration. In all these algorithms, tuning makes the preconditioned right hand side of the linear system an approximate eigenvector (or invariant subspace) of the preconditioned system matrix, and hence the inner iteration counts are considerably reduced. This motivation of tuning has also recently been shown in [15] to bear an interesting relation to the Jacobi-Davidson method.

[†]Department of Mathematics, Temple University, Philadelphia, PA 19122.(`fxue@temple.edu`).

[‡]Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. (`elman@cs.umd.edu`)

In the meantime, some developments have been made in understanding inexact projection-based eigenvalue algorithms, such as the Lanczos and the Arnoldi methods. It was found in [18] and [4] that the matrix-vector product must be computed accurately in the initial Lanczos or Arnoldi steps, but the accuracy can be relaxed as the algorithm proceeds without obviously affecting the convergence of approximate eigenpairs. An analysis of this phenomenon is given in [27] for the Arnoldi method, using perturbation theory of invariant subspaces. It is shown there that the allowable errors of matrix-vector products in Arnoldi steps should be inversely proportional to the eigenvalue residual norm of the desired eigenpair. Therefore, as the Arnoldi method proceeds and converges to the eigenpair of interest, the accuracy of matrix-vector products can be relaxed. The use of inexact matrix-vector products has also been studied in the setting of Krylov subspace linear solvers; see [5, 6],[29, 30, 31] and [33, 38].

Further study of inexact Arnoldi methods is given in [16], where the tuning strategy and the relaxed accuracy of matrix-vector products are extended to inexact IRA with shift-invert transformation for standard eigenvalue problems. For the linear systems arising in Arnoldi steps (outer iterations) in a given IRA cycle, tuning is developed using all available Arnoldi vectors in that cycle. Numerical experiments show that for a test problem from Matrix Market [23], an ILU preconditioner with this tuning considerably reduces the inner iteration counts. It is observed there and confirmed in this paper that this improvement is mainly due to the fact that tuning helps cluster the eigenvalues of the preconditioned system matrix of the linear system in each Arnoldi step. In addition, [16] proposes a practical estimate of the allowable relaxed tolerances for the solution of the linear systems, using the distance between the spectra of two matrices containing the wanted and unwanted Ritz values; this is proposed as a simpler alternative to the separation between the two matrices [35]. Numerical experiments show that the total inner iteration counts of inexact IRA can be substantially reduced by the combined use of tuning and relaxed tolerances.

In this paper, we refine the tuning strategy and further study the allowable tolerances for inner solves of the inexact IRA method for generalized non-Hermitian eigenvalue problems. We first study a new tuning strategy constructed for a given Arnoldi step using the solutions of linear systems obtained in previous Arnoldi steps. In addition, we propose a two-phase strategy to solve the linear system in the current Arnoldi step. Specifically, we apply only one step of preconditioned GMRES with tuning to the current linear system to get an approximate solution, then solve the correction equation with any appropriate preconditioned linear solver; in particular, tuning is not needed for the correction equation. We show that the approximate solution obtained in the first phase can be a very good one if enough solution vectors from previous Arnoldi steps are used. With this special approximate solution, the correction equation can be solved with a relative tolerance much larger than that of the original linear system, and inner iteration counts can hence be reduced considerably. In addition, we use a special type of iterative linear solver with subspace recycling to solve the sequence of correction equations as the IRA method proceeds. We show that subspace recycling is cheap to use in this setting and can further reduce the inner iteration counts substantially.

A second goal of this paper is to present a refined analysis of the allowable tolerance for the linear systems in the inexact IRA method. We first give an upper bound of the allowable tolerance, showing that violation of this bound necessarily leads to contamination of the desired approximate invariant subspace by excessive errors from the inner solves. We then give a theoretically more accurate estimate of the allowable tolerance, which is between the upper bound and a conservative lower bound from [16]. As this estimate contains information not available until the end of the current IRA cycle, we use a computable substitute obtained at the end of the previous IRA cycle. We then compare this heuristic estimate with that from [16] and discuss the impact of the accuracy of the estimate on the inner solves.

The paper is organized as follows. In Section 2, we briefly review spectral transformations, the IRA method and some properties of the algorithm when exact shifts (unwanted Ritz values) are used in filter polynomials. We discuss a few strategies for the inner solves in Section 3, studying the properties of the new tuning strategy and the new two-phase strategy for solving

the linear system in each Arnoldi step. We also explain the effectiveness of the linear solver with subspace recycling applied to solve the correction equations. In Section 4, we study the allowable tolerances of the linear systems and give a necessary upper bound for the tolerance. A new heuristic estimate of the allowable tolerance is proposed and used in numerical experiments to corroborate the accuracy of the estimate from [16]. Numerical experiments in Section 5 show that the combined use of the new tuning, subspace recycling and relaxed tolerances greatly reduces the total inner iteration counts. In Section 6 we make some concluding remarks.

2. Review: spectral transformations and the IRA method. To make the exposition smooth, we briefly review two commonly used spectral transformations and the implicitly restarted Arnoldi (IRA) method.

The shift-invert and generalized Cayley transformations (see [21]) are usually used to detect interior eigenvalues or ones with large imaginary parts. They are defined as follows:

$$(2.1) \qquad Av = \lambda Bv \;\Leftrightarrow\; (A - \sigma B)^{-1} Bv = \left(\frac{1}{\lambda - \sigma} \right) v \quad \text{(shift-invert)}$$

$$Av = \lambda Bv \;\Leftrightarrow\; (A - \sigma_1 B)^{-1}(A - \sigma_2 B)v = \left(\frac{\lambda - \sigma_2}{\lambda - \sigma_1} \right) v \quad \text{(generalized Cayley)}$$

The shift-invert transformation maps eigenvalues near σ to dominant eigenvalues of $\mathcal{A} = (A - \sigma B)^{-1}B$; the Cayley transformation maps eigenvalues to the right of the line $Re(\lambda) = \frac{\sigma_1 + \sigma_2}{2}$ to eigenvalues of $\mathcal{A} = (A - \sigma_1 B)^{-1}(A - \sigma_2 B)$ outside the unit circle, and those to the left of this line to ones inside the unit circle (assuming that $\sigma_1 > \sigma_2$). The dominant eigenvalues of $\mathcal{A}$ can then be found by iterative eigenvalue algorithms. Once the eigenvalues of the transformed problem are obtained, they are transformed back to those of the original problem; the eigenvectors do not change with the transformation.

Without loss of generality, we use the generic notation $\mathcal{A} = A^{-1}B$ for which we seek the k eigenvalues of $Av = \lambda Bv$ with smallest magnitude (i.e., k dominant eigenvalues of $\mathcal{A}$). This notation in principle covers both types of operators in (2.1) with any shifts. For example, let $\widehat{A} = A - \sigma_1 B$ and $\widehat{B} = A - \sigma_2 B$, so that the Cayley operator is $\mathcal{A} = \widehat{A}^{-1}\widehat{B}$. This generic notation $\mathcal{A} = A^{-1}B$ is used throughout this paper, unless otherwise stated.

Both the shift-invert and the Cayley transformations have been implemented in ARPACK [20], a mathematical software package of high quality which has become the standard solver for large non-Hermitian eigenvalue problems. ARPACK is based on the implicitly restarted Arnoldi (IRA) method, a well-known and important method for eigenvalue computation developed by Sorensen [34] in 1992.

The key technique of the IRA method is the implicit application of a filter polynomial to a given Arnoldi decomposition to produce the effect of several steps of a restarted Arnoldi computation without any matrix-vector multiplications. Specifically, at the end of the ith IRA cycle we have an m-step Arnoldi decomposition

$$(2.2) \qquad \mathcal{A}U_m^{(i)} = U_m^{(i)} H_m^{(i)} + h_{m+1,m}^{(i)} u_{m+1}^{(i)} e_m^T.$$

Suppose $\kappa_1, \kappa_2, ..., \kappa_{m-k} \in \mathbb{C}$ are estimates of $m - k$ eigenvalues of $\mathcal{A}$ obtained from this process corresponding to a part of the spectrum we are not interested in. We can use these numbers as shifts to apply $m - k$ shifted QR steps to $H_m^{(i)}$ and get a Krylov decomposition

$$(2.3) \qquad \mathcal{A}\tilde{U}_m^{(i)} = \tilde{U}_m^{(i)} \tilde{H}_m^{(i)} + h_{m+1,m}^{(i)} u_{m+1}^{(i)} (e_m^T Q^{(i)}),$$

where $Q^{(i)} = Q_1 Q_2 ... Q_{m-k}$ is the product of $m - k$ upper Hessenberg unitary matrices, $\tilde{U}_m^{(i)} = U_m^{(i)} Q^{(i)}$, $\tilde{H}_m^{(i)} = Q^{(i)*} H_m^{(i)} Q^{(i)}$ is upper Hessenberg, and $e_m^T Q^{(i)}$ is the last row of $Q^{(i)}$ with $k - 1$ zero leading entries. For details, see [34], or [17, 35].

The restarted Arnoldi decomposition is then obtained from the first k columns of the above

4

Krylov decomposition as follows

$$\mathcal{A}\tilde{U}_k^{(i)} = \tilde{U}_k^{(i)}\tilde{H}_k^{(i)} + \tilde{h}_{k+1,k}^{(i)}\tilde{u}_{k+1}^{(i)}e_k^T + (h_{m+1,m}^{(i)}q_{m,k}^{(i)})u_{m+1}^{(i)}e_k^T, \qquad \text{or,} \tag{2.4}$$
$$\mathcal{A}U_k^{(i+1)} = U_k^{(i+1)}H_k^{(i+1)} + h_{k+1,k}^{(i+1)}u_k^{(i+1)}e_k^T.$$

Here $q_{m,k}^{(i)}$ is the (m,k) entry of $Q^{(i)}$, $U_k^{(i+1)} = \tilde{U}_k^{(i)}$, and $H_k^{(i+1)} = \tilde{H}_k^{(i)}$. Note that both $\tilde{u}_{k+1}^{(i)}$ and $u_{m+1}^{(i)}$ are orthogonal to $U_k^{(i+1)}$. Let $\hat{u}_k^{(i+1)} = \tilde{h}_{k+1,k}^{(i)}\tilde{u}_{k+1}^{(i)} + (h_{m+1,m}^{(i)}q_{m,k}^{(i)})u_{m+1}^{(i)}$; then $h_{k+1,k}^{(i+1)} = \|\hat{u}_k^{(i+1)}\|$, and $u_k^{(i+1)} = (h_{k+1,k}^{(i+1)})^{-1}\hat{u}_k^{(i+1)}$. Clearly, no additional matrix-vector product involving $\mathcal{A}$ is used for the restart. For the restarted Arnoldi decomposition, it can be shown that $u_1^{(i+1)} = (\mathcal{A} - \kappa_1 I)(\mathcal{A} - \kappa_2 I)...(\mathcal{A} - \kappa_{m-k}I)u_1^{(i)}$ up to a constant scaling factor. In other words, the eigenvector component corresponding to the unwanted spectrum in $u_1^{(i)}$ is filtered out by the filter polynomial.

An inexact implicitly restarted Arnoldi method is given as follows:

Algorithm 1: inexact implicitly restarted Arnoldi (IRA) method

---

Given a normalized $u_1^{(0)} \in \mathbb{C}^n$, $j = 1$

For $i = 0, 1, 2, ...$

    1. Compute $\mathcal{A}u_j^{(i)}$ by solving $Ay = Bu_j^{(i)}$ to a prescribed relative tolerance $\delta^{(i,j)}$

    2. Expand the Arnoldi decomposition by orthogonalizing y against $u_1^{(i)}, ..., u_j^{(i)}$ and normalizing; the new Arnoldi decomposition is $\mathcal{A}U_j^{(i)} = U_j^{(i)}H_j^{(i)} + h_{j+1,j}^{(i)}u_{j+1}^{(i)}e_j^T$. $j \leftarrow j+1$.

    3. If $j = m$, test for convergence. If not converged, invoke the implicit restart procedure to get $\mathcal{A}U_k^{(i+1)} = U_k^{(i+1)}H_k^{(i+1)} + h_{k+1,k}^{(i+1)}u_{k+1}^{(i+1)}e_k^T$, and $j \leftarrow k+1$.

End For

In this study, we choose the "exact shifts" strategy for the IRA method, which uses the unwanted eigenvalues of $H_m^{(i)}$ (Ritz values) as shifts for the implicit restart. This is the default choice in ARPACK and has proved successful in many applications. Some properties of the IRA method with the exact shifts strategy are given as follows.

PROPOSITION 2.1 (Corollary 2.3, Chapter 5 of [35]). *Suppose $\mu_1, ..., \mu_m$ are eigenvalues of $H_m^{(i)}$. If the implicit QR steps are performed with shifts $\{\mu_{k+1}, \mu_{k+2}, ..., \mu_m\}$, then*

$$\tilde{H}_m^{(i)} = Q^{(i)*}H_m^{(i)}Q^{(i)} = \begin{bmatrix} \tilde{H}_k^{(i)} & \tilde{H}_m^{12(i)} \\ 0 & \tilde{H}_m^{22(i)} \end{bmatrix}, \tag{2.5}$$

where $\tilde{H}_m^{22(i)}$ is an upper triangular matrix with $\mu_{k+1}, \mu_{k+2}, ..., \mu_m$ on its diagonal.

The proposition shows that $\tilde{h}_{k+1,1}^{(i)} = 0$ if exact shifts are used. This observation immediately leads to the following result.

PROPOSITION 2.2. *Let the Schur decomposition of $H_m^{(i)}$ be $H_m^{(i)} = W_m^{(i)}T_m^{(i)}W_m^{(i)*}$, where $W_m^{(i)} = \left[W_m^{1(i)}, W_m^{2(i)}\right]$ is unitary, and*

$$T_m^{(i)} = \begin{bmatrix} T_m^{11(i)} & T_m^{12(i)} \\ 0 & T_m^{22(i)} \end{bmatrix} \tag{2.6}$$

with $\lambda(T_m^{11(i)}) = \{\mu_1, \mu_2, ..., \mu_k\}$, $\lambda(T_m^{22(i)}) = \{\mu_{k+1}, \mu_{k+2}, ..., \mu_m\}$, and $\lambda(T_m^{11(i)}) \cap \lambda(T_m^{22(i)}) = \emptyset$. Then

$$\|\mathcal{A}U_m^{(i)}W_m^{1(i)} - U_m^{(i)}W_m^{1(i)}T_m^{11(i)}\| = \|\mathcal{A}U_k^{(i+1)} - U_k^{(i+1)}H_k^{(i+1)}\|, \quad and \tag{2.7}$$
$$\|h_{m+1,m}^{(i)}u_{m+1}^{(i)}e_m^T W_m^{1(i)}\| = \|h_{k+1,k}^{(i+1)}u_{k+1}^{(i+1)}e_k^T\|. \tag{2.8}$$

Proof. Let $Q^{(i)} = \left[Q^{1(i)}, Q^{2(i)}\right]$. From (2.5) and (2.6) we have $Q^{1(i)*}H_m^{(i)}Q^{1(i)} = \tilde{H}_k^{(i)}$ and $W_m^{1(i)*}H_m^{(i)}W^{1(i)} = T_m^{11(i)}$. Since $\lambda(\tilde{H}_k^{(i)}) = \lambda(T_m^{11(i)}) = \{\mu_1, \mu_2, ..., \mu_k\}$, there exists a $k \times k$ unitary matrix $V^{(i)}$ such that $V^{(i)*}\tilde{H}_k^{(i)}V^{(i)} = T_m^{11(i)}$ and $W_m^{1(i)} = Q^{1(i)}V^{(i)}$. Note from (2.3) and (2.4) that $U_m^{(i)}Q^{1(i)} = U_k^{(i+1)}$. Therefore

$$(2.9) \qquad \|\mathcal{A}U_m^{(i)}W_m^{1(i)} - U_m^{(i)}W_m^{1(i)}T_m^{11(i)}\|$$
$$= \|\mathcal{A}U_m^{(i)}Q^{1(i)}V^{(i)} - U_m^{(i)}Q^{1(i)}V^{(i)}T_m^{11(i)}V^{(i)*}V^{(i)}\|$$
$$= \|(\mathcal{A}U_k^{(i+1)} - U_k^{(i+1)}\tilde{H}_k^{(i)})V^{(i)}\| = \|\mathcal{A}U_k^{(i+1)} - U_k^{(i+1)}\tilde{H}_k^{(i)}\|.$$

Since $\tilde{h}_{k+1,k}^{(i)} = 0$, we have $h_{k+1,k}^{(i+1)}u_{k+1}^{(i+1)}e_k^T = (h_{m+1,m}^{(i)}q_{m,k}^{(i)})u_{m+1}^{(i)}e_k^T$ from (2.4), and therefore

$$(2.10) \qquad \|h_{m+1,m}^{(i+1)}u_{m+1}^{(i)}e_m^TW_m^{1(i)}\| = \|h_{m+1,m}^{(i+1)}u_{m+1}^{(i)}e_m^TQ^{1(i)}V^{(i)}\|$$
$$= \|(h_{m+1,m}^{(i+1)}q_{m,k}^{(i)})u_{m+1}^{(i)}e_k^T\| = \|h_{k+1,k}^{(i+1)}u_{k+1}^{(i+1)}e_k^T\|.$$

□

These results are applicable to the standard (exact) IRA method as well as the inexact version of Algorithm 1. For the exact IRA method (where the matrix-vector products involving $\mathcal{A}$ are computed exactly), (2.8) can be derived from (2.7). For inexact IRA, however, the "true eigenvalue residuals" in (2.7) and the "estimated eigenvalue residuals" in (2.8) are different. Proposition 2.2 shows that the two types of eigenvalue residual norms are "restart-invariant" if exact shifts are used: both quantities at the end of the ith IRA cycle are the same as those at the beginning of the $(i+1)$th IRA cycle.

3. New strategies for solving linear systems in inexact IRA. To improve the efficiency for solving the linear systems arising in inexact eigenvalue algorithms, a special type of preconditioner with "tuning" is studied in [13, 14, 26, 16]. An existing preconditioner P is modified using a special low-rank update of P to produce a tuned preconditioner $\mathbb{P}$ that behaves like the system matrix A on a certain set of vectors X. It is shown in these papers that the inner iteration counts needed to solve the linear system preconditioned by $\mathbb{P}$ are substantially smaller than those required to solve the system preconditioned by P.

For example, consider inexact subspace iteration with $\mathcal{A} = A^{-1}$ used to detect a few smallest eigenvalues of A. In each outer iteration, we approximately solve the block linear system $AY^{(i)} = X^{(i)}$, where $X^{(i)}$ contains the current approximate Schur vectors (therefore $X^{(i)*}X^{(i)} = I$). It is shown in [26] that a decreasing sequence of tolerances for the block systems is necessary to guarantee the linear convergence of $X^{(i)}$ to the desired invariant subspace. As a result, the block-GMRES iteration counts required to solve $AP^{-1}\tilde{Y}^{(i)} = X^{(i)}$ (with $Y^{(i)} = P^{-1}\tilde{Y}^{(i)}$) increases gradually as the outer iteration progresses. To resolve this difficulty, P is replaced by the tuned preconditioner

$$(3.1) \qquad \mathbb{P}^{(i)} = P + (A - P)X^{(i)}X^{(i)*},$$

for which $\mathbb{P}^{(i)}X^{(i)} = AX^{(i)}$, or equivalently, $A(\mathbb{P}^{(i)})^{-1}(AX^{(i)}) = AX^{(i)}$. That is, $AX^{(i)}$ spans an invariant subspace of the tuned preconditioned system matrix corresponding to eigenvalue 1. For $A(\mathbb{P}^{(i)})^{-1}\tilde{Y}^{(i)} = X^{(i)}$, the right hand side $X^{(i)}$ spans an approximate invariant subspace of $A(\mathbb{P}^{(i)})^{-1}$, and the block-GMRES iteration counts needed for solving this preconditioned system do *not* increase with the outer iteration progress.

This idea of tuning is extended in [16] to an inexact IRA method for standard eigenvalue problems. Let m and k be the order of the Arnoldi decomposition, i.e., the number of columns in the Hessenberg matrix right before and after the implicit restart. Assume after the jth $(0 \le j \le m-k-1)$ Arnoldi step in the ith IRA cycle, an Arnoldi decomposition $\mathcal{A}U_{k+j}^{(i)} = U_{k+j}^{(i)}H_{k+j}^{(i)} + h_{k+j+1,k+j}^{(i)}u_{k+j+1}^{(i)}e_{k+j}^T$ is already computed, and $Ay = u_{k+j+1}^{(i)}$ needs to be solved

in the $(j+1)$th Arnoldi step. In [16], the tuned preconditioning matrix is defined as $\mathbb{P}_{k+j+1}^{(i)} = P + (A - P)XX^*$, where $X = U_{k+j+1}^{(i)}$ contains the Arnoldi vectors in the ith IRA cycle. It is shown that the inner iteration counts required to solve $A(\mathbb{P}_{k+j+1}^{(i)})^{-1}\tilde{y} = u_{k+j+1}^{(i)}$ are smaller than those needed to solve $AP^{-1}\tilde{y} = u_{k+j+1}^{(i)}$, because $A(\mathbb{P}_{k+j+1}^{(i)})^{-1}$ has better eigenvalue clustering than AP^{-1}. This "clustering" effect of tuning is quite different from the original motivation of this strategy studied in [13, 14, 26]. In particular, $u_{k+j+1}^{(i)}$ is generally not a very good approximate eigenvector of $A(\mathbb{P}_{k+j+1}^{(i)})^{-1}$.

In this section, we propose and study a new tuning strategy for solving the linear systems of equations that arise in inexact IRA for generalized non-Hermitian eigenvalue problems. To study the new strategy under ideal conditions, we assume in this section that the linear system in each Arnoldi step is solved accurately (to machine precision). We also show how tuning can be used in a new two-phase algorithm to solve the linear systems in each Arnoldi step. In addition, we discuss the use of subspace recycling with iterative solvers in the second phase of the two-phase algorithm.

3.1. The new tuning strategy. The motivation for the tuning strategy is similar to that discussed in [13, 14, 26]: to make the right-hand side of the linear system associated with the spectral transformation an approximate eigenvector of the preconditioned system matrix, so that the inner iteration counts can be greatly reduced. Suppose we are in the ith IRA cycle and already have $\mathcal{A}U_{k+j}^{(i)} = U_{k+j}^{(i)}H_{k+j}^{(i)} + h_{k+j+1,k+j}^{(i)}u_{k+j+1}^{(i)}e_{k+j}^T$. Computing $\mathcal{A}u_{k+j+1}^{(i)}$ entails solving $Ay = Bu_{k+j+1}^{(i)}$. Recall that for a given X with orthonormal columns, the tuned preconditioner $\mathbb{P} = P + (A - P)XX^*$ satisfies $\mathbb{P}X = AX$, i.e., $A\mathbb{P}^{-1}(AX) = AX$. Tuning requires that X be chosen so that the right-hand side $Bu_{k+j+1}^{(i)}$ of the current linear system approximately lies in the subspace spanned by AX, an invariant subspace of $A\mathbb{P}^{-1}$.

Consider the choice

$$(3.2) \qquad X = X_p^{(i,l)} = \left[\mathcal{A}U_m^{(i-l)}, \mathcal{A}U_{k+1:m}^{(i-l+1)}, ..., \mathcal{A}U_{k+1:m}^{(i-1)}, \mathcal{A}U_{k+1:k+j}^{(i)} \right],$$

where $U_{k+1:m}^{(r)}$ stands for the $(k+1)$th through the mth columns of $U_m^{(r)}$, and $p = m + (m-k)(l-1) + j$ is the number of vectors in $X_p^{(i,l)}$. We refer to $X_p^{(i,l)}$ as the set of "solution vectors," because its columns are solutions of the linear systems in previous Arnoldi steps. For example, the first vector in $X_p^{(i,l)}$ is $\mathcal{A}u_1^{(i-l)}$, the solution of $Ay = Bu_1^{(i-l)}$ in the first step of the $(i-l)$th IRA cycle. Note that this system may not literally be solved in practice due to the implicit restart.

Let $U_p^{(i,l)} = \left[U_m^{(i-l)}, U_{k+1:m}^{(i-l+1)}, ..., U_{k+1:k+j}^{(i)} \right]$ and $\mathcal{U}_p^{(i,l)} = \text{span}\{U_p^{(i,l)}\}$, $\mathcal{X}_p^{(i,l)} = \text{span}\{X_p^{(i,l)}\}$, $A\mathcal{X}_p^{(i,l)} = B\mathcal{U}_p^{(i,l)} = \text{span}\{BU_m^{(i-l)}, BU_{k+1:m}^{(i-l+1)}, ..., BU_{k+1:k+j}^{(i)}\}$. In the following derivation, we use the calligraphic letter to stand for the subspaces spanned by some set of *column vectors* denoted by the same letter in Roman fonts. For instance, $\mathcal{U}_{k+j}^{(i)} = \text{span}\{U_{k+j}^{(i)}\}$. To study the relation between $Bu_{k+j+1}^{(i)}$ and $A\mathcal{X}_p^{(i,l)}$, we begin with the following lemma, which shows that the range of $U_p^{(i,l)}$ is a Krylov subspace.

LEMMA 3.1. *Suppose IRA does not break down. Then $\mathcal{U}_p^{(i,l)} = \mathcal{K}_p(\mathcal{A}, u_1^{(i-l)})$.*

Proof. First, $\mathcal{U}_{m+1}^{(i-l)} = \mathcal{K}_{m+1}(\mathcal{A}, u_1^{(i-l)})$. Since $u_{k+1}^{(i-l+1)}$ is a linear combination of $u_{m+1}^{(i-l)}$ and $\tilde{u}_{k+1}^{(i-l)} \in \mathcal{U}_m^{(i-l)}$ (see (2.4)), $\text{span}\{U_m^{(i-l)}, u_{k+1}^{(i-l+1)}\} = \mathcal{K}_{m+1}(\mathcal{A}, u_1^{(i-l)})$ holds. As we have orthogonalized $\mathcal{A}u_{k+1}^{(i-l+1)}$ against $\mathcal{U}_{k+1}^{(i-l+1)} \subset \text{span}\{U_m^{(i-l)}, u_{k+1}^{(i-l+1)}\}$ (note that $\mathcal{U}_k^{(i-l+1)} \subset \mathcal{U}_m^{(i-l)}$; see (2.4)) to get $u_{k+2}^{(i-l+1)}$, $\text{span}\{U_m^{(i-l)}, u_{k+1}^{(i-l+1)}, u_{k+2}^{(i-l+1)}\} = \mathcal{K}_{m+2}(\mathcal{A}, u_1^{(i-l)})$ follows. Similar reasoning holds for all following Arnoldi vectors if IRA does not break down, and the theorem is established. ☐

The angle between a vector v and a subspace $\mathcal{U}$ (denoted as $\angle(v, \mathcal{U})$) is defined as the angle between v and the orthogonal projection of v onto $\mathcal{U}$. Obviously, $v \in \mathcal{U}$ if and only

if $\angle(v,\mathcal{U}) = 0$. Therefore, $Bu_{k+j+1}^{(i)}$ approximately lies in $A\mathcal{X}_p^{(i,l)} = B\mathcal{U}_p^{(i,l)}$ if and only if $\angle(Bu_{k+j+1}^{(i)}, B\mathcal{U}_p^{(i,l)})$ is small, and this small angle condition holds if $\varphi_p^{(i)} = \angle(u_{k+j+1}^{(i)}, \mathcal{U}_p^{(i,l)})$ is small. The following theorem suggests that, given the starting vector $\mathcal{A}u_1^{(i-l)}$ of $X_p^{(i,l)}$, $\varphi_p^{(i)}$ can be small enough for large p, because it decreases linearly with p when $p > m$.

THEOREM 3.2. *Let* $\varphi_{p-1}^{(i)} = \angle(u_{k+j}^{(i)}, \mathcal{U}_{p-1}^{(i,l)})$ $(1 \le j \le m-k)$, *so that* $u_{k+j}^{(i)} = u_{p-1}\cos\varphi_{p-1}^{(i)} + u_{p-1}^\perp \sin\varphi_{p-1}^{(i)}$, *where* $u_{p-1} \in \mathcal{U}_{p-1}^{(i,l)}$ *and* $u_{p-1}^\perp \perp \mathcal{U}_{p-1}^{(i,l)}$ *are unit vectors. Let the orthogonal projection of* $\|\mathcal{A}u_{p-1}^\perp\|^{-1}\mathcal{A}u_{k+j}^{(i)}$ *onto* $\mathcal{U}_p^{(i,l)}$ *be* $w_p\eta_p$, *where* $w_p \in \mathcal{U}_p^{(i,l)}$ *is a unit vector, and let* $\alpha_p = \angle(\mathcal{A}u_{p-1}^\perp, \mathcal{U}_p^{(i,l)})$, *and* $\beta_p = \angle(w_p, \mathcal{U}_{k+j}^{(i)})$. *Then*

$$\tan\varphi_p^{(i)} = \nu_p \sin\varphi_{p-1}^{(i)}, \quad \text{where } \nu_p = \frac{\sin\alpha_p}{\eta_p \sin\beta_p}. \quad (1 \le j \le m-k) \tag{3.3}$$

(Note that α_p, β_p, η_p and ν_p all depend on the IRA cycle number i. To simplify the notation, we omit the superscripts for these scalars.)

Proof. Let $\rho = \frac{\|\mathcal{A}u_{p-1}\|}{\|\mathcal{A}u_{p-1}^\perp\|}$. Since $\mathcal{U}_p^{(i,l)} = \mathcal{K}_p(\mathcal{A}, u_1^{(i-l)})$, we have

$$\mathcal{A}u_{k+j}^{(i)} = \mathcal{A}u_{p-1}\cos\varphi_{p-1}^{(i)} + \mathcal{A}u_{p-1}^\perp \sin\varphi_{p-1}^{(i)} \tag{3.4}$$
$$= \rho\|\mathcal{A}u_{p-1}^\perp\|w_{p1}\cos\varphi_{p-1}^{(i)} + \|\mathcal{A}u_{p-1}^\perp\|(w_{p2}\cos\alpha_p + w_p^\perp \sin\alpha_p)\sin\varphi_{p-1}^{(i)}$$
$$= \|\mathcal{A}u_{p-1}^\perp\|(w_{p1}\rho\cos\varphi_{p-1}^{(i)} + w_{p2}\cos\alpha_p \sin\varphi_{p-1}^{(i)} + w_p^\perp \sin\alpha_p \sin\varphi_{p-1}^{(i)})$$
$$= \|\mathcal{A}u_{p-1}^\perp\|(w_p\eta_p + w_p^\perp \sin\alpha_p \sin\varphi_{p-1}^{(i)}),$$

where $w_{p1}, w_{p2}, w_p \in \mathcal{U}_p^{(i,l)}$ and $w_p^\perp \perp \mathcal{U}_p^{(i,l)}$ are all unit vectors, and $w_p\eta_p = w_{p1}\rho\cos\varphi_{p-1}^{(i)} + w_{p2}\cos\alpha_p \sin\varphi_{p-1}^{(i)}$ is the orthogonal projection of $\frac{\mathcal{A}u_{k+j}^{(i+l)}}{\|\mathcal{A}u_{p-1}^\perp\|}$ onto $\mathcal{U}_p^{(i,l)}$. It follows immediately that $\tan\angle(\mathcal{A}u_{k+j}^{(i)}, \mathcal{U}_p^{(i,l)}) = \frac{\sin\alpha_p \sin\varphi_{p-1}^{(i)}}{\eta_p}$.

We then orthogonalize $\mathcal{A}u_{k+j}^{(i)}$ against $\mathcal{U}_{k+j}^{(i)} \subset \mathcal{U}_p^{(i,l)}$ to get $u_{k+j+1}^{(i)}$. Let $\mathcal{U}_{k+j}^{(i)\perp}$ be the orthogonal complement of $\mathcal{U}_{k+j}^{(i)}$ in $\mathcal{U}_p^{(i,l)}$. Then $w_p = w_{p3}\cos\beta_p + w_{p4}\sin\beta_p$, where $w_{p3} \in \mathcal{U}_{k+j}^{(i)}$ and $w_{p4} \in \mathcal{U}_{k+j}^{(i)\perp}$ are unit vectors, and $\beta_p = \angle(w_p, \mathcal{U}_{k+j}^{(i)})$. Orthogonalizing $\mathcal{A}u_{k+j}^{(i)}$ against $\mathcal{U}_{k+j}^{(i)}$ removes the w_{p3} component from w_p, so that $u_{k+j+1}^{(i)}$ equals $w = w_{p4}\eta_p \sin\beta_p + w_p^\perp \sin\alpha_p \sin\varphi_{p-1}^{(i)}$ up to a constant scaling factor. It follows that $\tan\angle(u_{k+j+1}^{(i)}, \mathcal{U}_p^{(i,l)}) = \frac{\sin\alpha_p \sin\varphi_{p-1}^{(i)}}{\eta_p \sin\beta_p}$, and (3.3) is established. □

Remark 3.1. In Theorem 3.2 we are interested in the nontrivial case where $l > 0$. If $l = 0$, then $p = k + j$, and $\mathcal{U}_p^{(i,0)} = \mathcal{U}_{k+j}^{(i)}$. Therefore $\beta_p = \angle(w_p, \mathcal{U}_{k+j}^{(i)}) = 0$ (because $w_p \in \mathcal{U}_p^{(i,0)}$ by definition), ν_p is infinity, and $\varphi_p^{(i)} = \pi/2$. This is consistent with the fact that Arnoldi vectors in the same IRA cycle are orthogonal.

Remark 3.2. We have assumed that exact shifts are used for the implicit restart. In this case $u_{k+1}^{(i)} = u_{m+1}^{(i-1)}$; see (2.4) and Proposition 2.1. Therefore Theorem 3.2 also holds for $j = 0$, with $u_{k+j}^{(i)}$ replaced by $u_m^{(i-1)}$.

It is obvious from (3.3) that $\varphi_p^{(i)}$ decreases linearly with p if ν_p remains a constant smaller than 1. In practice ν_p is not a constant, but we have solid empirical evidence that $\varphi_p^{(i)}$ does decrease with p linearly.

Theorem 3.2 shows that with the choice of $X_p^{(i,l)}$ in (3.2), $\angle(u_{k+j+1}^{(i)}, \mathcal{U}_p^{(i,l)})$ is small for large enough p. As a result, $\angle(Bu_{k+j+1}^{(i)}, B\mathcal{U}_p^{(i,l)}) = \angle(Bu_{k+j+1}^{(i)}, A\mathcal{X}_p^{(i,l)})$ is also small. In other words, $Bu_{k+j+1}^{(i)}$ is an approximate eigenvector of $A\mathbb{P}^{-1}$ (where the tuned preconditioner $\mathbb{P}$ is constructed using $X_p^{(i,l)}$), because it approximately lies in $A\mathcal{X}_p^{(i,l)}$, an invariant subspace

of $A\mathbb{P}^{-1}$. In the following subsection, we show how this observation can be used in a new two-phase algorithm for solving $Ay = Bu_{k+j+1}^{(i)}$.

3.2. A two-phase strategy to solve the linear systems in Arnoldi steps. With the new tuning discussed in subsection 3.1, we now propose a new two-phase algorithm for efficiently solving $Ay = Bu_{k+j+1}^{(i)}$ in Step 1 of Algorithm 1. This algorithm provides new insights into the use of tuning and its properties.

Algorithm 2: Two-phase strategy for solving $Ay = Bu_{k+j+1}^{(i)}$

---
1. Construct tuned preconditioner using (3.2) and apply a *single* step of preconditioned GMRES with tuning to get an approximate solution y_1.
2. For a given ϵ, choose either a fixed tolerance $\delta = \delta_f(\epsilon)$, or a relaxed tolerance $\delta = \delta_r(\epsilon)$ by some means. Solve the *correction equation*
   $Az = Bu_{k+j+1}^{(i)} - Ay_1$ with *any* appropriate preconditioned iterative solver
   to get an approximate correction z_q, such that the corrected iterate
   $y_{q+1} = y_1 + z_q$ satisfies $\frac{\|Bu_{k+j+1}^{(i)} - Ay_{q+1}\|}{\|Bu_{k+j+1}^{(i)}\|} \le \delta$, or equivalently, the correction
   z_q satisfies $\frac{\|(Bu_{k+j+1}^{(i)} - Ay_1) - Az_q\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|} \le \frac{\delta \|Bu_{k+j+1}^{(i)}\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|}$

In particular, tuning need *not* be used to solve the correction equation. Thus we can work with a fixed preconditioned system matrix for the correction equation in all Arnoldi steps.

Let $u_{k+j+1}^{(i)} = u_p c_p^{(i,l)} + u_p^\perp s_p^{(i,l)}$, where $u_p \in \mathcal{U}_p^{(i,l)}$ and $u_p^\perp \perp \mathcal{U}_p^{(i,l)}$ are unit vectors, $c_p^{(i,l)}$ and $s_p^{(i,l)}$ are the cosine and sine of $\angle(u_{k+j+1}^{(i)}, \mathcal{U}_p^{(i,l)})$. We have shown by Theorem 3.2 that $s_p^{(i,l)}$ can be small enough for large p. The analysis of Algorithm 2 is given in the following major theorem.

THEOREM 3.3. *Suppose Algorithm 2 is used to solve $Ay = Bu_{k+j+1}^{(i)}$. Then Phase I of Algorithm 2 gives $y_1 = \mathcal{A}u_p c_p^{(i,l)} + O(s_p^{(i,l)})$ (up to a constant scaling factor) and the corresponding relative residual norm $\frac{\|Bu_{k+j+1}^{(i)} - Ay_1\|}{\|Bu_{k+j+1}^{(i)}\|} = O(s_p^{(i,l)})$. Consequently, the stopping criterion of Algorithm 2 is satisfied if and only if the relative residual of the correction equation $\frac{\|(Bu_{k+j+1}^{(i)} - Ay_1) - Az_q\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|} \le \frac{\delta \|Bu_{k+j+1}^{(i)}\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|} = \frac{\delta}{O(s_p^{(i,l)})}$.*

Proof. It is shown in Section 3.1 that if the preconditioning matrix $\mathbb{P}_p^{(i,l)}$ is constructed using $X_p^{(i,l)}$, then $A(\mathbb{P}_p^{(i,l)})^{-1}(AX_p^{(i,l)}) = AX_p^{(i,l)}$. That is, $A\mathcal{X}_p^{(i,l)} = B\mathcal{U}_p^{(i,l)}$ is an invariant subspace of dimension p of $A(\mathbb{P}_p^{(i,l)})^{-1}$ corresponding to eigenvalue 1. It follows that for $u_p \in \mathcal{U}_p^{(i,l)}$, $(\mathbb{P}_p^{(i,l)})^{-1}(Bu_p) = A^{-1}(Bu_p)$.

The approximate solution to $A(\mathbb{P}_p^{(i,l)})^{-1}\tilde{y} = Bu_{k+j+1}^{(i)}$ after one step of GMRES iteration is $\tilde{y}_1 \in \text{span}\{Bu_{k+j+1}^{(i)}\}$, i.e., $\tilde{y}_1 = \gamma Bu_{k+j+1}^{(i)}$ with some scalar γ. Therefore

$$(3.5) \qquad y_1 = (\mathbb{P}_p^{(i,l)})^{-1}\tilde{y}_1 = (\mathbb{P}_p^{(i,l)})^{-1}(\gamma Bu_{k+j+1}^{(i)}) = \gamma(\mathbb{P}_p^{(i,l)})^{-1}(Bu_p c_p^{(i,l)} + Bu_p^\perp s_p^{(i,l)})$$
$$= \gamma\left(A^{-1}Bu_p c_p^{(i,l)} + (\mathbb{P}_p^{(i,l)})^{-1}Bu_p^\perp s_p^{(i,l)}\right) = \gamma(\mathcal{A}u_p c_p^{(i,l)} + \delta_p^{(i,l)}),$$

where $\|\delta_p^{(i,l)}\| = s_p^{(i,l)}\|(\mathbb{P}_p^{(i,l)})^{-1}Bu_p^\perp\| = O(s_p^{(i,l)})$.

Now consider the residual norm after one step of GMRES with a tuned preconditioner:

$$(3.6) \qquad \|Bu_{k+j+1}^{(i)} - Ay_1\| = \min_{\gamma} \|Bu_{k+j+1}^{(i)} - \gamma A(\mathbb{P}_p^{(i,l)})^{-1}(Bu_{k+j+1}^{(i)})\|$$

$$\leq \|Bu_{k+j+1}^{(i)} - A(\mathbb{P}_p^{(i,l)})^{-1}(Bu_{k+j+1}^{(i)})\|$$

$$= \|Bu_{k+j+1}^{(i)} - A(\mathbb{P}_p^{(i,l)})^{-1}(Bu_p c_p^{(i,l)} + Bu_p^\perp s_p^{(i,l)})\|$$

$$= \|Bu_p c_p^{(i,l)} + Bu_p^\perp s_p^{(i,l)} - A(A^{-1}Bu_p)c_p^{(i,l)} - A(\mathbb{P}_p^{(i,l)})^{-1}Bu_p^\perp s_p^{(i,l)})\|$$

$$= s_p^{(i,l)}\|(A(\mathbb{P}_p^{(i,l)})^{-1} - I)Bu_p^\perp\|.$$

Therefore the relative residual norm is $s_p^{(i,l)} \frac{\|(A(\mathbb{P}_p^{(i,l)})^{-1}-I)Bu_p^\perp\|}{\|Bu_{k+j+1}^{(i)}\|} = O(s_p^{(i,l)})$.

Finally, Phase II of Algorithm 2 requires that

$$(3.7) \qquad \frac{\|Bu_{k+j+1}^{(i)} - Ay_{q+1}\|}{\|Bu_{k+j+1}^{(i)}\|} = \frac{\|Bu_{k+j+1}^{(i)} - Ay_{q+1}\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|} \frac{\|Bu_{k+j+1}^{(i)} - Ay_1\|}{\|Bu_{k+j+1}^{(i)}\|} \leq \delta,$$

which is satisfied if and only if the relative residual of the correction equation

$$(3.8) \qquad \frac{\|(Bu_{k+j+1}^{(i)} - Ay_1) - Az_q\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|} = \frac{\|Bu_{k+j+1}^{(i)} - Ay_{q+1}\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|}$$

$$\leq \frac{\delta\|Bu_{k+j+1}^{(i)}\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|} = \frac{\delta}{O(s_p^{(i,l)})}.$$

The proof is thus concluded. □

Remark 3.3. The theorem shows that y_1 obtained in Phase I of Algorithm 2 equals $\mathcal{A}u_p c_p^{(i,l)}$ plus a small quantity proportional to $s_p^{(i,l)}$. As $u_p \in \mathcal{U}_p^{(i,l)}$, $\mathcal{A}u_p \in \mathcal{AU}_p^{(i,l)} = \mathcal{X}_p^{(i,l)}$; see (3.2). Recall that $X_p^{(i,l)}$ consists of the "solution vectors" of the linear systems in previous Arnoldi steps. Therefore, by constructing tuning as in Section 3.1 and applying one step of preconditioned GMRES with tuning to $Ay = Bu_{k+j+1}^{(i)}$, we get a good approximate solution y_1 which is roughly a linear combination of those solution vectors. The reason for the success of this approach is that $\angle(Bu_{k+j+1}^{(i)}, B\mathcal{U}_p^{(i,l)})$ is small, i.e., $Bu_{k+j+1}^{(i)}$ is roughly a linear combination of the right-hand sides of the previously solved systems. This perspective is quite different from the motivation of tuning in previous literature [13, 14, 16, 26].

Remark 3.4. The theorem shows that a good approximate solution y_1 can be computed inexpensively in Phase I by tuning so that $\frac{\|Bu_{k+j+1}^{(i)} - Ay_1\|}{\|Bu_{k+j+1}^{(i)}\|} = O(s_p^{(i,l)}) \ll 1$. In fact, a valid y_1 can also be obtained in other ways, in particular, by solving a least squares problem

$$(3.9) \qquad \min_{f} \|Bu_{k+j+1}^{(i)} - AX_p^{(i,l)}f\|,$$

which can be easily done using the QR decomposition of $AX_p^{(i,l)} = BU_p^{(i,l)}$ (recall the definition of $X_p^{(i,l)}$ in (3.2)). Given that $u_{k+j+1}^{(i)} = u_p c_p^{(i,l)} + u_p^\perp s_p^{(i,l)}$, where $u_p \in \mathcal{U}_p^{(i,l)}$ and $u_p^\perp \perp \mathcal{U}_p^{(i,l)}$, we have

$$(3.10) \qquad \min_{f \in \mathbb{C}^p} \|Bu_{k+j+1}^{(i)} - AX_p^{(i,l)}f\| = \min_{f \in \mathbb{C}^p} \|Bu_{k+j+1}^{(i)} - BU_p^{(i,l)}f\|$$

$$\leq \|B(u_p c_p^{(i,l)} + u_p^\perp s_p^{(i,l)}) - Bu_p c_p^{(i,l)}\| = s_p^{(i,l)}\|Bu_p^\perp\|.$$

Therefore, with $y_1 = X_p^{(i,l)}f$, we have $\frac{\|Bu_{k+j+1}^{(i)} - Ay_1\|}{\|Bu_{k+j+1}^{(i)}\|} = s_p^{(i,l)} \frac{\|Bu_p^\perp\|}{\|Bu_{k+j+1}^{(i)}\|} = O(s_p^{(i,l)})$. The Phase I computation is somewhat cheaper for the least squares approach than the one-step tuned

preconditioned GMRES, but the former method required slightly more iterations in Phase II for our test problems, and the total inner iteration counts are about the same for the two approaches. In the following, for the sake of brevity, we only study the two-phase strategy where tuning is applied in Phase I.

Remark 3.5. Due to the large reduction of the linear residual norm in Phase I, the stopping criterion in Algorithm 2, $\frac{\|Bu_{k+j+1}^{(i)} - Ay_{q+1}\|}{\|Bu_{k+j+1}^{(i)}\|} \leq \delta$, is satisfied if and only if the relative residual of the correction equation $\frac{\|(Bu_{k+j+1}^{(i)} - Ay_1) - Az_q\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|}$ is bounded by the much less stringent relative tolerance $\frac{\delta\|Bu_{k+j+1}^{(i)}\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|} = \frac{\delta}{O(s_p^{(i,l)})} \gg \delta$. This larger relative tolerance implies that the inner iterations required for solving the correction equation can be considerably smaller than those needed to solve the original equation directly.

3.3. Subspace recycling for the correction equation. Step 2 of Algorithm 2 can be improved using linear solvers with subspace recycling. This methodology has proved efficient for solving a long sequence of *slowly-changing* linear systems. When the iterative solution of one linear system is done, a small set of vectors from the current subspace for the candidate solutions is selected and "recycled," i.e., used for the solution of the next system in the sequence. Subspace recycling usually reduces the cost of solving subsequent linear systems, because the iterative solver does not have to build the candidate solution subspace from scratch. A popular solver of this type is the Generalized Conjugate Residual with implicit inner Orthogonalization and Deflated Restarting (GCRO-DR) [25] developed using ideas of special truncation [7] and restarting [22] for solving a single linear system.

Reference [25] makes a general assumption that the preconditioned system matrix changes from one linear system to the next, and thus the recycled subspace taken from the previous system must be transformed by matrix-vector products involving the current system matrix to fit into the solution of the current system. For the sequence of correction equations in Algorithm 2, fortunately, this transformation can be avoided, because the preconditioned system matrix without tuning is the same for the correction equation in all Arnoldi steps.

It is suggested in [25] that the harmonic Ritz vectors corresponding to smallest harmonic Ritz values can be chosen to span the recycled subspaces. These vectors are approximate eigenvectors of the preconditioned system matrix corresponding to smallest eigenvalues. If the harmonic Ritz vectors are good approximate eigenvectors, this strategy tends to reduce the duration of the initial latency of GMRES convergence typically observed when the system matrix has some eigenvalues of very small magnitude; see [11]. Our subspace recycling also includes dominant Ritz vectors, as suggested in [25]. In Section 5, we present experimental results to show that the set of dominant Ritz vectors is an effective choice for subspace recycling if the use of harmonic Ritz vectors fails to reduce the inner iteration counts.

4. A refined analysis of allowable errors in Arnoldi steps. It was observed empirically in [4] that for the unrestarted Arnoldi method, the matrix-vector products involving $\mathcal{A}$ must be computed with high accuracy in the initial Arnoldi steps, but the accuracy can be relaxed as the iteration proceeds. A similar observation was also given in [18] for an inexact Lanczos method. An analysis based on matrix perturbation theory in [27] shows that the allowable errors of the matrix-vector products need only be inversely proportional to the eigenvalue residual norm of the current desired approximate invariant subspace for the quality of the approximate invariant subspace generated by the inexact Arnoldi method to be good. This *relaxation strategy* is extended in [16] to the inexact IRA method, where a practical estimate of the allowable errors of the inner solves at each Arnoldi step is proposed. Ideally, accurately estimated allowable errors can help reduce the inner iteration counts to the best extent possible without compromising the performance of the eigenvalue solvers. In this section, we give a refined analysis and an alternative estimate of allowable errors of the inner solves.

Suppose the matrix-vector product involving $\mathcal{A} = A^{-1}B$ is applied inexactly for m Arnoldi steps, with an error $f_j = y - A^{-1}Bu_j$ $(1 \leq j \leq m)$ introduced in the linear solve of $Ay = Bu_j$.

Thus we have the following inexact Arnoldi decomposition:

$$(4.1) \qquad \mathcal{A}U_m + F_m = (\mathcal{A} + F_m U_m^*)U_m = U_m H_m + h_{m+1,m} u_{m+1} e_m^T,$$

where U_m spans a Krylov subspace of the perturbed matrix $\mathcal{A} + F_m U_m^*$. Let the Schur decomposition of H_m be

$$(4.2) \qquad H_m = W_m T_m W_m^*, \text{ with } W_m = \begin{bmatrix} W_m^{11} & W_m^{12} \\ W_m^{21} & W_m^{22} \end{bmatrix} \text{ and } T_m = \begin{bmatrix} T_m^{11} & T_m^{12} \\ 0 & T_m^{22} \end{bmatrix},$$

where $T_m^{11} \in \mathbb{C}^{k \times k}$, $T_m^{22} \in \mathbb{C}^{p \times p}$, and $\lambda(T_m^{11})$ are the wanted Ritz values and $\lambda(T_m^{22})$ are the unwanted ones. Then we use the Rayleigh-Ritz method (Section 4.1, Chapter 4 of [35]) to extract the desired approximate invariant subspace $U_m W_m^1$, where $W_m^1 = \begin{bmatrix} W_m^{11} \\ W_m^{21} \end{bmatrix}$ contains the wanted Ritz vectors. From (4.1) and (4.2), the corresponding eigenvalue residual is

$$(4.3) \qquad \begin{aligned} \mathcal{A}U_m W_m^1 - U_m H_m W_m^1 &= \mathcal{A}U_m W_m^1 - U_m W_m^1 T_m^{11} \\ &= h_{m+1,m} u_{m+1} e_m^T W_m^1 - F_m W_m^1, \end{aligned}$$

from which follows

$$(4.4) \qquad \|(\mathcal{A}U_m W_m^1 - U_m W_m^1 T_m^{11}) - h_{m+1,m} u_{m+1} e_m^T W_m^1\| = \|F_m W_m^1\|.$$

Here, as introduced in Section 2, $\mathcal{A}U_m W_m^1 - U_m W_m^1 T_m^{11}$ is the true eigenvalue residual, and $R_m = h_{m+1,m} u_{m+1} e_m^T W_m^1$ is the estimated residual (referred to as the "computed residual" in [27, 16]). The difference between the two residuals depends on $\|F_m W_m^1\|$. For the inexact Arnoldi method, we want to keep the quality of $U_m W_m^1$ under control in spite of the presence of the error matrix F_m. To achieve this goal, we need to control $\|F_m W_m^1\|$ so that the desired approximate invariant subspace $U_m W_m^1$ contained in U_m is not obviously contaminated by F_m, i.e., the true residual is still reasonably close to the estimated residual.

To see why the allowable errors at some Arnoldi steps can be relaxed, note that

$$(4.5) \qquad \|F_m W_m^1\| \leq \|F_k W_m^{11}\| + \|F_{k+1:m} W_m^{21}\| \leq \|F_k\| + \|F_{k+1:m}\|\|W_m^{21}\|.$$

Therefore, for a given k-step inexact Arnoldi decomposition with a small enough $\|F_k\|$, $\|F_{k+1:m}\|$ does not have to be very small as long as $\|W_m^{21}\|$, i.e., the magnitude of the last $m - k$ entries of the wanted Ritz vectors W_m^1 (see (4.2)), is small enough. The next theorem, which extends Theorem 3.2 of [16], shows that $\|W_m^{21}\|$ is proportional to the estimated residual at step k.

THEOREM 4.1. *Let $\mathcal{A}U_k + F_k = U_k H_k + h_{k+1,k} u_{k+1} e_k^T$ be a k-step inexact Arnoldi decomposition, where the Schur decomposition of H_k is $H_k = W_k T_k W_k^*$. Let $m - k$ additional inexact Arnoldi steps be performed, giving an m-step decomposition $\mathcal{A}U_m + F_m = U_m H_m + h_{m+1,m} u_{m+1} e_m^T$. Let $R_k = \mathcal{A}U_k W_k - U_k H_k W_k = h_{k+1,k} u_{k+1} e_k^T W_k$ be the estimated residual at Arnoldi step k. Given the Schur decomposition of H_m in (4.2), then*

$$(4.6) \qquad \frac{\|R_k\|}{\|R_k\| + \|\mathcal{S}_m\|} \leq \|W_m^{21}\| \leq \frac{\|R_k\|}{\text{sep}(T_k, T_m^{22})},$$

where $\mathcal{S}_m$ is the Sylvester operator $G \to \mathcal{S}_m(G) : T_m^{22} G - G T_k$, $\text{sep}(T_k, T_m^{22}) = \min_{\|G\|=1} \|\mathcal{S}_m(G)\|$, and $\|\mathcal{S}_m\| = \max_{\|G\|=1} \|\mathcal{S}_m(G)\|$.

Proof. We only need to prove the lower bound, as the upper bound is established in Theorem 3.2 of [16]. The estimated residual norm at step k is

$$(4.7) \qquad \|R_k\| = \|h_{k+1,k} u_{k+1} e_k^T W_k\| = h_{k+1,k} \|e_k^T W_k\| = h_{k+1,k}.$$

Consider the first block of the Schur decomposition of H_m (4.2):

$$(4.8) \quad \left\| H_m \begin{pmatrix} W_k \\ 0 \end{pmatrix} - \begin{pmatrix} W_k \\ 0 \end{pmatrix} T_k \right\|$$

$$= \left\| \begin{pmatrix} H_k & H_m^{12} \\ h_{k+1,k}e_1 e_k^T & H_m^{22} \end{pmatrix} \begin{pmatrix} W_k \\ 0 \end{pmatrix} - \begin{pmatrix} W_k \\ 0 \end{pmatrix} T_k \right\|$$

$$= \left\| \begin{pmatrix} H_k \\ h_{k+1,k}e_1 e_k^T \end{pmatrix} W_k - \begin{pmatrix} W_k \\ 0 \end{pmatrix} T_k \right\| = \left\| \begin{pmatrix} H_k W_k - W_k T_k \\ h_{k+1,k}e_1 e_k^T W_k \end{pmatrix} \right\|$$

$$= \left\| \begin{pmatrix} 0 \\ h_{k+1,k}e_1 e_k^T W_k \end{pmatrix} \right\| = h_{k+1,k}\|e_k^T W_k\| = h_{k+1,k},$$

so that the first line of (4.8) also equals the estimated residual norm. Since W_m is unitary, it follows that with the first expression in the last line of (4.8), we have

$$(4.9) \quad \|R_k\| = \left\| W_m^* \left(H_m \begin{pmatrix} W_k \\ 0 \end{pmatrix} - \begin{pmatrix} W_k \\ 0 \end{pmatrix} T_k \right) \right\|$$

$$= \left\| \begin{pmatrix} (W_m^{11})^* & (W_m^{21})^* \\ (W_m^{12})^* & (W_m^{22})^* \end{pmatrix} \begin{pmatrix} 0 \\ h_{k+1,k}e_1 e_k^T W_k \end{pmatrix} \right\| = \left\| \begin{pmatrix} h_{k+1,k}(W_m^{21})^* e_1 e_k^T W_k \\ h_{k+1,k}(W_m^{22})^* e_1 e_k^T W_k \end{pmatrix} \right\|.$$

On the other hand, using the Schur decomposition of H_m, we also have

$$(4.10) \quad \|R_k\| = \left\| W_m^* \left(H_m \begin{pmatrix} W_k \\ 0 \end{pmatrix} - \begin{pmatrix} W_k \\ 0 \end{pmatrix} T_k \right) \right\|$$

$$= \left\| T_m W_m^* \begin{pmatrix} W_k \\ 0 \end{pmatrix} - W_m^* \begin{pmatrix} W_k \\ 0 \end{pmatrix} T_k \right\|$$

$$= \left\| \begin{pmatrix} T_m^{11} & T_m^{12} \\ 0 & T_m^{22} \end{pmatrix} \begin{pmatrix} (W_m^{11})^* W_k \\ (W_m^{12})^* W_k \end{pmatrix} - \begin{pmatrix} (W_m^{11})^* W_k \\ (W_m^{12})^* W_k \end{pmatrix} T_k \right\|$$

$$= \left\| \begin{pmatrix} T_m^{11}(W_m^{11})^* W_k - (W_m^{11})^* W_k T_k + T_m^{12}(W_m^{12})^* W_k \\ T_m^{22}(W_m^{12})^* W_k - (W_m^{12})^* W_k T_k \end{pmatrix} \right\|.$$

Using the upper block from (4.9) and lower block from (4.10), we have

$$(4.11) \quad \|R_k\| = \left\| \begin{pmatrix} h_{k+1,k}(W_m^{21})^* e_1 e_k^T W_k \\ T_m^{22}(W_m^{12})^* W_k - (W_m^{12})^* W_k T_k \end{pmatrix} \right\|$$

$$\leq \|h_{k+1,k}(W_m^{21})^* e_1 e_k^T W_k\| + \|T_m^{22}(W_m^{12})^* W_k - (W_m^{12})^* W_k T_k\|$$

$$\leq h_{k+1,k}\|e_k^T W_k\|\|e_1^T W_m^{21}\| + \|\mathcal{S}_m\|\|(W_m^{12})^* W_k\|$$

$$= \|R_k\|\|e_1^T W_m^{21}\| + \|\mathcal{S}_m\|\|W_m^{12}\| \leq \|R_k\|\|W_m^{21}\| + \|\mathcal{S}_m\|\|W_m^{21}\|.$$

Note that in the last line of (4.11), $\|(W_m^{12})^* W_k\| = \|W_m^{12}\| = \|W_m^{21}\|$ (Theorem 2.6.1 in Golub and van Loan [17]). The lower bound in (4.6) is thus established. $\square$

As observed above, for a given k-step inexact Arnoldi decomposition with small $\|F_k\|$, the error matrix $\|F_{k+1:m}\|$ associated with the upcoming $m-k$ inexact Arnoldi steps must be controlled appropriately to make sure that $U_m W_m^1$ is not obviously contaminated after these steps. In particular, $\|f_{k+1}\|$ cannot be too big. The following theorem gives an upper bound of $\|f_{k+1}\|$.

THEOREM 4.2. *Given $\epsilon_1 > 0$, suppose we have a k-step inexact Arnoldi decomposition $\mathcal{A}U_k + F_k = U_k H_k + h_{k+1,k}u_{k+1}e_k^T$, where $\|F_k\| \leq \epsilon_1$. Let $\mathcal{S}_{k+1}$ be defined as in Theorem 4.1. Then for the next Arnoldi step,*

$$(4.12) \quad \|f_{k+1}\| \leq \left(1 + \frac{\|\mathcal{S}_{k+1}\|}{\|R_k\|} \right) (\epsilon_1 + \epsilon_2)$$

is a necessary condition to make $\|(\mathcal{A}U_{k+1}W_{k+1}^1 - U_{k+1}W_{k+1}^1 T_{k+1}^{11}) - R_{k+1}\| \leq \epsilon_2$

Proof. Let $m = k + 1$. We have the following estimate of the difference between the computed and true residual,

$$(4.13) \qquad \|(\mathcal{A}U_{k+1}W_{k+1}^1 - U_{k+1}W_{k+1}^1 T_{k+1}^{11}) - R_{k+1}\| = \|F_{k+1}W_{k+1}^1\|$$
$$= \|F_k W_{k+1}^{11} + f_{k+1}W_{k+1}^{21}\| \geq \|f_{k+1}W_{k+1}^{21}\| - \|F_k W_{k+1}^{11}\|$$
$$\geq \|f_{k+1}\|\|W_{k+1}^{21}\| - \|F_k\|\|W_{k+1}^{11}\| \geq \|f_{k+1}\|\frac{\|R_k\|}{\|R_k\| + \|\mathcal{S}_{k+1}\|} - \epsilon_1.$$

Note that $\|f_{k+1}W_{k+1}^{21}\| = \|f_{k+1}\|\|W_{k+1}^{21}\|$ because f_{k+1} and W_{k+1}^{21} are respectively a column vector and row vector, and $\|W_{k+1}^{11}\| \leq \|W_{k+1}^1\| = 1$. It follows immediately that (4.13) is bigger than ϵ_2 if $\|f_{k+1}\| > (1 + \frac{\|\mathcal{S}_{k+1}\|}{\|R_k\|})(\epsilon_1 + \epsilon_2)$. □

A practical choice would be $\epsilon_1 = \epsilon_2 \equiv \epsilon$. Using the upper bound of $\|W_m^{21}\|$ in Theorem 4.1, we can also show that $\|f_{k+1}\| \leq \frac{\text{sep}(T_m^{22}, T_k)}{\|R_k\|}\epsilon$ is *sufficient* to make $\|(\mathcal{A}U_{k+1}W_{k+1}^1 - U_{k+1}W_{k+1}^1 T_{k+1}^{11}) - R_{k+1}\| \leq 2\epsilon$.

However, the bounds of $\|f_{k+1}\|$ in the necessary and sufficient conditions might be severe overestimate and underestimate, respectively, of the actual allowable error in the $(k+1)$th Arnoldi step. In fact, $\text{sep}(T_{k+1}^{22}, T_k)$ and $\|\mathcal{S}_{k+1}\|$ are analogous to the smallest and the largest singular values of the Sylvester operator $\mathcal{S}_{k+1}$. The necessary condition is generally too weak, as an obviously smaller $\|f_{k+1}\|$ may still not suffice to keep the approximate invariant subspace from being contaminated. On the other hand, the sufficient condition might be overly conservative, giving excessively small tolerance for the linear system $Ay = Bu_{k+j+1}$ ($0 \leq j \leq m-k-1$) and leading to unnecessary extra inner iterations. To give a practical estimate of the allowable $\|F_{k+1:m}\|$, [16] substitutes $\min|\lambda(T_k) - \lambda(T_m^{22})|$ for $\text{sep}(T_m^{22}, T_k)$, which is difficult to estimate. Since $\min|\lambda(T_k) - \lambda(T_m^{22})| > \text{sep}(T_m^{22}, T_k)$ for non-normal $\mathcal{A}$, this substitution essentially gives a less conservative estimate.

A better estimate should be a trade-off between these two conditions. Theorem 3.2 of [16] uses $\|T_m^{22}(W_m^{12})^*W_k - (W_m^{12})^*W_k T_k\| \geq \text{sep}(T_m^{22}, T_k)\|(W_m^{12})^*W_k\|$, whereas Theorem 4.1 above applies $\|T_m^{22}(W_m^{12})^*W_k - (W_m^{12})^*W_k T_k\| \leq \|\mathcal{S}_m\|\|(W_m^{12})^*W_k\|$. Therefore, a more accurate estimate can be obtained by replacing the lower bound $\text{sep}(T_m^{22}, T_k)$ and upper bound $\|\mathcal{S}_m\|$ by

$$(4.14) \qquad \frac{\|T_m^{22}(W_m^{12})^*W_k - (W_m^{12})^*W_k T_k\|}{\|(W_m^{12})^*W_k\|} = \frac{\|T_m^{22}(W_m^{12})^* - (W_m^{12})^*H_k\|}{\|W_m^{12}\|},$$

which takes into account the actual effect of $\mathcal{S}_m$ on $(W_m^{12})^*W_k$. Here we use the fact that $H_k = W_k T_k W_k^*$ is a Schur decomposition.

The above strategy gives a theoretically more accurate estimate of $\|F_{k+1:m}\|$. However, like the estimate $\min|\lambda(T_k) - \lambda(T_m^{22})|$ in [16], it depends on the Schur decomposition of H_m, which is not available at step k. The practical (heuristic) solution is to use the decomposition of H_m from the previous IRA cycle and H_k of the current cycle. Specifically, suppose at the beginning of the ith IRA cycle, we have $\mathcal{A}U_k^{(i)} + F_k^{(i)} = U_k^{(i)}H_k^{(i)} + h_{k+1,k}^{(i)}u_{k+1}^{(i)}e_k^T$. Then we define

$$(4.15) \qquad \sigma_{est}^{(i)} \equiv \frac{\|T_m^{22(i-1)}(W_m^{12(i-1)})^* - (W_m^{12(i-1)})^*H_k^{(i)}\|}{\|W_m^{12(i-1)}\|},$$

which is very easy to compute. Note that $H_k^{(i)} = \tilde{H}_k^{(i-1)}$ if the exact shift strategy is used; see (2.4). Substituting $\sigma_{est}^{(i)}$ for $\text{sep}(T_m^{22(i)}, T_k^{(i)})$ in the relaxation strategy (3.13) in [16], we have the following heuristic estimate of the allowable errors:

$$(4.16) \qquad \|f_j^{(i)}\| \leq \frac{\epsilon}{2k} \qquad\qquad (i = 1, 1 \leq j \leq m), \quad \text{and}$$

$$\|f_{k+j+1}^{(i)}\| \leq \frac{\epsilon}{2(m-k)}\frac{\sigma_{est}^{(i)}}{\|R_k^{(i)}\|} \qquad (i > 1, 0 \leq j \leq m-k-1).$$

Remark. To the best of our knowledge, given a k-step inexact Arnoldi decomposition with small $\|F_k\|$, none of the existing *practical* (computable) estimates of allowable $\|F_{k+1:m}\|$ can theoretically guarantee that the desired approximate invariant subspace $U_m W_m^1$ will not be contaminated after $m-k$ inexact Arnoldi steps. The estimate in [27] for the unrestarted Arnoldi method involves the distance between the desired Ritz value extracted from H_k and the rest of the spectrum of H_k, assuming that the computed eigenvalue residual at step k is already small enough, which might not be the case (see Section 3.1 of [27] for the explicit formula); reference [16] uses $\min |\lambda(T_k^{(i)}) - \lambda(T_m^{22(i-1)})|$ in place of $\mathrm{sep}(T_k^{(i)}, T_m^{22(i)})$, which is replaced by $\sigma_{est}^{(i)}$ in our new estimate. We will compare the new estimate with that in [16] in Section 5.

Finally, we point out that $\|F_m\|$ should be properly scaled. In fact, as $\mathcal{A}U_m + F_m = U_m H_m + h_{m+1,m} u_{m+1} e_m^T$, the *relative* quantity $\frac{\|F_m\|}{\|\mathcal{A}U_m\|}$ should be used to measure the magnitude of errors, especially if $\|\mathcal{A}U_m\|$ is *not* moderate. Specifically, at the $(k+j+1)$th Arnoldi step, the linear system $Ay = Bu_{k+j+1}$ needs to be solved inexactly. The relative error $\frac{\|f_{k+j+1}\|}{\|\mathcal{A}u_{k+j+1}\|} = \frac{\|y - A^{-1}Bu_{k+j+1}\|}{\|A^{-1}Bu_{k+j+1}\|}$ is not available as we do not have $A^{-1}Bu_{k+j+1}$. A reasonable and convenient substitute is the relative residual norm of this linear system $\frac{\|Ay - Bu_{k+j+1}\|}{\|Bu_{k+j+1}\|}$. For our inexact IRA method, we require this quantity to be bounded above by the new estimate in (4.16).

5. Numerical Experiments. We present and discuss the results of numerical experiments in this section, showing the effectiveness of the new tuning strategy, subspace recycling and the new relaxation strategy. The following issues are addressed:

1. We show that the tuning strategy constructed using solution vectors obtained from previous Arnoldi steps works as Theorem 3.3 describes: a single GMRES step with this tuning applied to $Ay = Bu_{k+j+1}^{(i)}$ gives a good approximate solution y_1 for which the residual norm $\frac{\|Bu_{k+j+1}^{(i)} - Ay_1\|}{\|Bu_{k+j+1}^{(i)}\|} = O(s_p^{(i,l)}) \ll 1$, and therefore the correction equation can be solved with a less stringent relative tolerance $\frac{\delta\|Bu_{k+j+1}^{(i)}\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|} \gg \delta$, no matter if δ is a fixed or relaxed relative tolerance for $Ay = Bu_{k+j+1}^{(i)}$. The new tuning strategy is compared with the original tuning strategy in [16].

2. We compare inexact IRA methods with non-relaxed (fixed) tolerances $\delta_f = \frac{\epsilon}{2k}$ where ϵ and k are given in Table 5.1 (this δ_f is used in [16] for inexact IRA with a fixed tolerance for the inner solve) and relaxed tolerances δ_r given by either the original estimate in [16] or the new estimate in (4.16). The accuracy of the two estimates is discussed based on the numerical results.

3. We show that further reduction of inner iteration counts can be achieved at little cost by proper subspace recycling.

We first explain the stopping criterion for the inexact IRA method. Suppose at the beginning of the ith IRA cycle, we have $\mathcal{A}U_k^{(i)} + F_k^{(i)} = U_k^{(i)} H_k^{(i)} + h_{k+1,k}^{(i)} u_{k+1}^{(i)} e_k^T$. Let $(\theta_j^{(i)}, v_j^{(i)})$ $(1 \le j \le k)$ be a Ritz pair, i.e., an eigenpair of $H_k^{(i)}$. Post-multiplying the above equation by $v_j^{(i)}$, we have

$$(5.1) \qquad \mathcal{A}(U_k^{(i)} v_j^{(i)}) - \theta_j^{(i)}(U_k^{(i)} v_j^{(i)}) - (h_{k+1,k}^{(i)} v_{kj}^{(i)}) u_{k+1}^{(i)} = -F_k^{(i)} v_j^{(i)},$$

where $v_{kj}^{(i)}$ is the k-th (last) entry of $v_j^{(i)}$. Here $U_k^{(i)} v_j^{(i)}$ is an approximate eigenvector of $\mathcal{A}$, $\mathcal{A}(U_k^{(i)} v_j^{(i)}) - \theta_j^{(i)}(U_k^{(i)} v_j^{(i)})$ is the true eigenvalue residual, and $(h_{k+1,k}^{(i)} v_{kj}^{(i)}) u_{k+1}^{(i)}$ is the estimated residual. As the magnitude of errors has been kept under control to guarantee that the true residual is close enough to the estimated one, we check if

$$(5.2) \qquad \left| \frac{h_{k+1,k}^{(i)} v_{kj}^{(i)}}{\theta_j^{(i)}} \right| \approx \frac{\|\mathcal{A}(U_k^{(i)} v_j^{(i)}) - \theta_j^{(i)}(U_k^{(i)} v_j^{(i)})\|}{|\theta_j^{(i)}|}$$

is smaller than some prescribed tolerance ϵ_{eig}. Using estimated residuals avoids the overhead of the Rayleigh-Ritz procedure.

Another issue is that k does not have to be equal to the number of desired eigenpairs k_w. One can choose a slightly bigger k for the IRA method, and only test $|(\theta_j^{(i)})^{-1} h_{k+1,k}^{(i)} v_{kj}^{(i)}|$ in (5.2) for $1 \le j \le k_w$. Our experience is that for fixed $m-k$ (the number of Arnoldi steps in each *restarted* IRA cycle), more often than not, this choice of k reduces the number of IRA cycles. We speculate that this strategy makes the unwanted Ritz values more separated from the desired eigenvalues; therefore it is less likely for the filter polynomial to damp the desired eigenvector components during the restart.

Four test problems are used in our numerical experiments. The first is SHERMAN5 from MatrixMarket [23], a real matrix of order 3312 arising from oil reservoir modeling. We use the shift-invert operator $\mathcal{A} = A^{-1}$ (with $B = I$) to detect some eigenvalues closest to zero. The inner solve is done with ILU preconditioning with drop tolerance 0.008 given by MATLAB's `ilu` function. This example is used in [16] to show the effectiveness of the tuning and the relaxation strategy therein.

The second problem UTM1700A/B, also from MatrixMarket, is a real matrix pencil of order 1700 arising from a tokamak model in plasma physics. We use Cayley transformation to detect that the leftmost eigenvalues are $\lambda_{1,2} = -0.032735 \pm 0.3347i$ and $\lambda_3 = 0.032428$. Here $\Im(\lambda_{1,2})$ is 10 times bigger than λ_3, and there are some real eigenvalues to the right of λ_3 with magnitude smaller than $\Im(\lambda_{1,2})$. An ILU preconditioner with drop tolerance 0.001 is used for the inner iteration.

Problems 3 and 4 arise from the linear stability analysis of a model of two-dimensional incompressible fluid flow over a backward facing step, constructed using the IFISS software package [8, 9]. The domain is $[-1, L] \times [-1, 1]$ with $[-1, 0] \times [-1, 0]$ cut out, where $L = 15$ in Problem 3 and $L = 23$ in Problem 4; the Reynolds numbers are 600 and 1200 respectively. Let u and v be the horizontal and vertical component of the velocity, p the pressure, and ν the viscosity. The boundary conditions are as follows:

(5.3)
$$u = 4y(1-y),\ v = 0 \text{ (parabolic inflow)} \qquad \text{on } x = -1,\ y \in [0,1];$$
$$\nu \frac{\partial u}{\partial x} - p = 0,\ \frac{\partial v}{\partial y} = 0 \text{ (natural outflow)} \qquad \text{on } x = L,\ y \in [-1,1];$$
$$u = v = 0 \text{ (no-slip)} \qquad\qquad \text{on all other boundaries.}$$

We use a biquadratic/bilinear (Q_2-Q_1) finite element discretization with element width $\frac{1}{16}$ (grid parameter 6 in the IFISS code). The two problems are of order 72867 and 110371 respectively. We use the least squares commutator preconditioner [10] for the inner solves. For both problems, we try shift-invert (subproblem (a)) and Cayley transformation (subproblem (b)) to detect a small number of critical eigenvalues.

For completeness, the parameters used in the solution of each test problem are given in Table 5.1. These parameters are chosen to deliver approximate eigenpairs of adequate accuracies and show representative behavior of each solution strategy.

1. k_w, k, m – we use the IRA method to compute k_w eigenpairs; m and k are the order of the Arnoldi decomposition before and after the implicit restart.
2. $\sigma, \sigma_1, \sigma_2$ – the shifts of $\mathcal{A} = (A - \sigma B)^{-1} B$ and $\mathcal{A} = (A - \sigma_1 B)^{-1}(A - \sigma_2 B)$
3. τ – we stop the IRA method if the estimated residual in (5.2) is smaller than τ for all k_w desired approximate eigenpairs
4. ϵ – the small quantity used in (4.16) to estimate the allowable tolerances for the linear systems
5. l_1, l_2 – l_1 harmonic Ritz vectors corresponding to harmonic Ritz values of smallest magnitude and l_2 dominant Ritz vectors are used for subspace recycling

Figure 5.1 plots the relative tolerances δ for the original systems $Ay = Bu_{k+j+1}^{(i)}$ (solid lines) and the derived relative tolerances $\frac{\delta \|Bu_{k+j+1}^{(i)}\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|}$ for the correction equations $Az =$

	k_w	p	k	σ (σ_1)	σ_2	τ	ϵ	l_1	l_2
Prob 1	8	8	12	0	–	2×10^{-14}	2×10^{-11}	10	10
Prob 2	3	4	9	−0.0325	0.125	2×10^{-9}	2×10^{-10}	10	10
Prob 3(a)	5	7	13	0	–	5×10^{-10}	5×10^{-9}	0	30
Prob 3(b)					−0.46				
Prob 4(a)					–			0	45
Prob 4(b)					−0.24				

TABLE 5.1
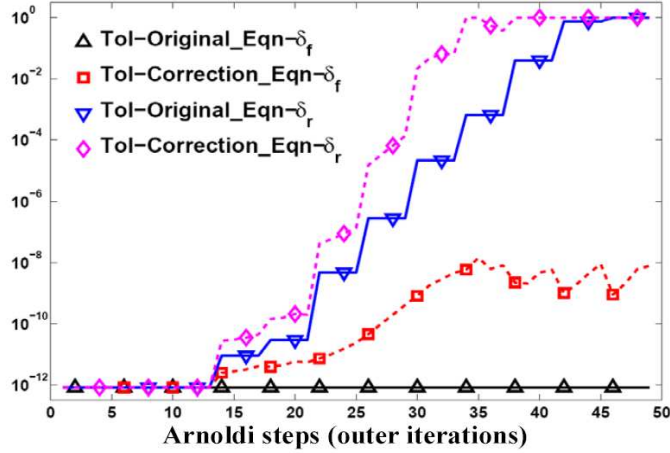
Parameters used to solve the test problems

FIG. 5.1. *Problem 1: relative tolerances for the original systems and correction equations*

$Bu_{k+j+1}^{(i)} - Ay_1$ (dashed lines) against the Arnoldi steps for Problem 1. The curves are as follows:

- $\triangle$ Tol-Original_Eqn-δ_f and $\square$ Tol-Correction_Eqn-δ_f – The fixed relative tolerance $\delta_f = \frac{\epsilon}{2k} \approx 10^{-12}$ for the original system $Ay = Bu_{k+j+1}^{(i)}$ and the derived relative tolerance $\frac{\delta_f \|Bu_{k+j+1}^{(i)}\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|}$ for the correction equation $Az = Bu_{k+j+1}^{(i)} - Ay_1$.
- $\triangledown$ Tol-Original_Eqn-δ_r and $\diamondsuit$ Tol-Correction_Eqn-δ_r – The relaxed relative tolerances δ_r estimated by (4.16) for $Ay = Bu_{k+j+1}^{(i)}$ and the derived relative tolerance $\frac{\delta_r \|Bu_{k+j+1}^{(i)}\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|}$ for $Az = Bu_{k+j+1}^{(i)} - Ay_1$.

Figure 5.1 corroborates the property of the two-phase algorithm described in Theorem 3.3. Specifically, by applying one step of GMRES with the new tuning to the original system, we get a good approximate solution y_1 for which the relative residual norm $\frac{\|Bu_{k+j+1}^{(i)} - Ay_1\|}{\|Bu_{k+j+1}^{(i)}\|} = O(s_p^{(i,l)}) \ll 1$, and therefore the derived relative tolerance of the correction equation $\frac{\delta \|Bu_{k+j+1}^{(i)}\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|} \gg \delta$ (see Remark 3.5).

The reduction of inner iterations by the two-phase algorithm (with the new tuning used in Phase I) can be seen from Figure 5.2, where the inner iteration counts required by three different strategies for solving $Ay = Bu_{k+j+1}^{(i)}$ are plotted against the Arnoldi steps:

- "No Tuning" (dotted line) – Solve the original systems $Ay = Bu_{k+j+1}^{(i)}$ by preconditioned GMRES to the fixed tolerance $\delta_f = \frac{\epsilon}{2k}$ without any enhancements.
- "Original Tuning" ($\triangle$, solid line) – Solve $Ay = Bu_{k+j+1}^{(i)}$ by GMRES with the original version of tuning in [16] to the fixed tolerance δ_f (note that the original tuning needs to
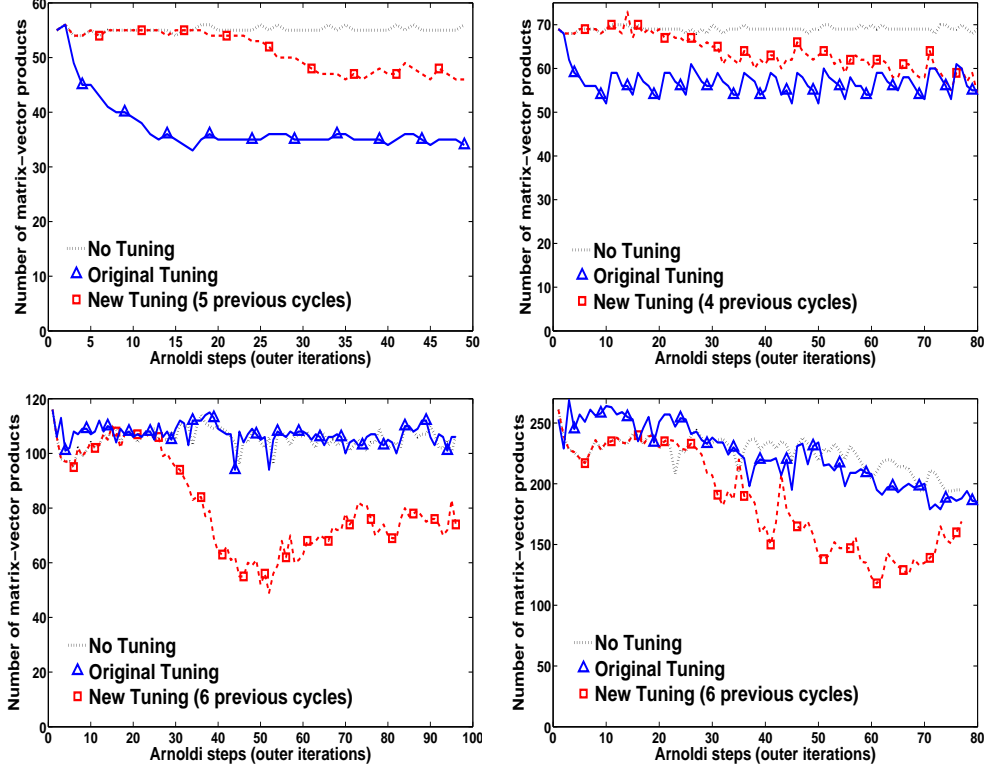
FIG. 5.2. *Performance of different strategies with fixed tolerances of inner solves for Problems 1, 2, 3(a) and 4(a)*

be applied at each GMRES step so that the eigenvalues of the preconditioned system matrix can be clustered and hence inner iteration counts can be reduced; this effect cannot be realized by trying the two-phase approach with this tuning strategy).

- "New Tuning (5 previous cycles)" ($\Box$, dashed line) – Solve $Ay = Bu_{k+j+1}^{(i)}$ by the two-phase algorithm to the fixed tolerance δ_f; in the first phase, the new tuning is constructed using solution vectors from the current and 5 previous IRA cycles.

Clearly, compared to the "No Tuning" strategy, the two-phase algorithm ($\Box$, with the new tuning used in Phase I) requires fewer inner iterations due to the larger relative tolerances for the correction equations.

We now compare the effectiveness of the two-phase algorithm (the new tuning is used in Phase I only for a single GMRES step) and the original tuning strategy that is applied at each GMRES step. In Figure 5.2, we see that for Problems 1 and 2, the relative reduction of inner iterations obtained by the use of the two-phase algorithm is only 15% at most; for Problems 3(a) and 4(a), however, the relative reduction can be as large as $30\% - 40\%$. On the other hand, the original tuning strategy works well for the first two problems, but fails to obviously reduce the inner iterations for the latter two.

The reason can be seen from Figure 5.3, which plots the relative residual norms of $Ay = Bu_{k+j+1}^{(i)}$ or $Az = Bu_{k+j+1}^{(i)} - Ay_1$ against inner iterations at a single Arnoldi step (step 34):

- Relres-CrtEqn (Solid line) – relative residual norms $\frac{\|(Bu_{k+j+1}^{(i)} - Ay_1) - Az_q\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|}$ of the correction equation $Az = Bu_{k+j+1}^{(i)} - Ay_1$ solved by preconditioned GMRES without tuning in Phase II of the two-phase strategy.

- Relres-OrigEqn (Dashed line) – residual norms $\frac{\|Bu_{k+j+1}^{(i)} - Ay_q\|}{\|Bu_{k+j+1}^{(i)}\|}$ of the original system $Ay = Bu_{k+j+1}^{(i)}$ solved by GMRES with the original version of tuning (two-phase strat-
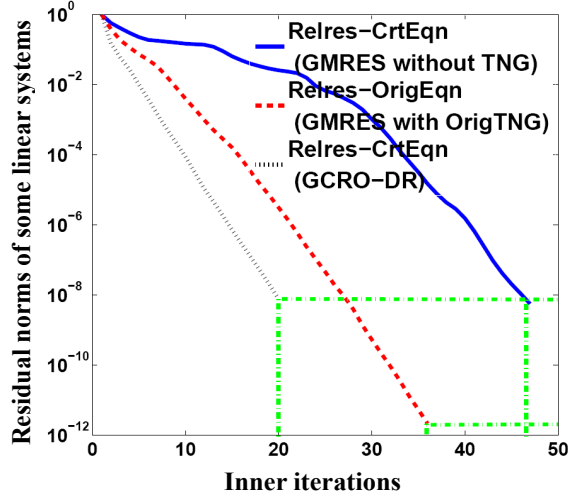
FIG. 5.3. *Problem 1: residual norms of $Ay = Bu_j^{(i)}$ or $Az = Bu_j^{(i)} - Ay_1$ at a single Arnoldi step, for three solution strategies*

egy cannot be used here).

- Relres-CrtEqn (Dash-dot line) – residual norms $\frac{\|(Bu_{k+j+1}^{(i)} - Ay_1) - Az_q\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|}$ of the correction equation solved by GCRO-DR (using subspace recycling) in Phase II of the two-phase strategy.

Here a fixed tolerance $\delta_f = \frac{\epsilon}{2k}$ is used for $Ay = Bu_{k+j+1}^{(i)}$. We include the behavior of the linear solver with subspace recycling (GCRO-DR), for reasons explained below. In this Arnoldi step, we see from Figure 5.1 that the relative tolerances for $Ay = Bu_{k+j+1}^{(i)}$ and $Az = Bu_{k+j+1}^{(i)} - Ay_1$ are $\delta_f \approx 10^{-12}$ and $\frac{\delta_f \|Bu_{k+j+1}^{(i)}\|}{\|Bu_{k+j+1}^{(i)} - Ay_1\|} \approx 10^{-8}$ respectively. Therefore, as shown in Figure 5.3, 47 GMRES steps without tuning are needed to solve the correction equation (48 GMRES steps in total for solving the original system, including the one step in Phase I), 36 GMRES steps with the original tuning are needed for solving the original system, and only 20 GCRO-DR stepsare required for the correction equation (21 inner steps in total for solving the original system). The asymptotic convergence rate of the three solves are roughly the same.

It is clear from Figure 5.3 that there are two types of strategies to reduce the inner iteration counts: one is to reduce the length of the latencies observed in the initial inner iteration steps, and the other is to use larger tolerances for the inner solves. The first type of strategy includes the original tuning and linear solvers with subspace recycling. Specifically, let P be an existing untuned preconditioner, and $\mathbb{P}$ be the tuned version defined in [16]. It is shown that the preconditioned operator $A\mathbb{P}^{-1}$ tends to have better eigenvalue clustering than AP^{-1}, especially if P is not strong. For linear solvers with subspace recycling, the recycled subspaces are spanned by approximate eigenvectors corresponding to smallest and/or largest eigenvalues of AP^{-1}. Both approaches essentially eliminate the eigenvalues of smallest and/or largest magnitude of the preconditioned system matrix; these are usually the source of the initial latencies exhibited during the inner iterations.

However, the original tuning is *not* always effective for this purpose. As Figure 5.2 shows, for Problems 3 and 4, the solution strategy with the original tuning ($\triangle$) requires almost as many inner iterations as with the untuned preconditioner (dotted line), while the new tuning ($\square$) reduces the inner iteration counts considerably. The reason is that the linear solves for these two problems are performed with the least square commutator preconditioner [10, 11], for which the preconditioned system matrix AP^{-1} has most eigenvalues clustered around 1 and

only a small number of outliers [11]. For this strong preconditioner, in our experience, it is hard for the original tuning method to further cluster the eigenvalues and reduce the initial latencies of GMRES iterations. On the other hand, as shown in Section 3.2, the improved performance of the new tuning method comes from the fact that the right-hand side of the system being solved is nearly an eigenvector of the preconditioned operator.
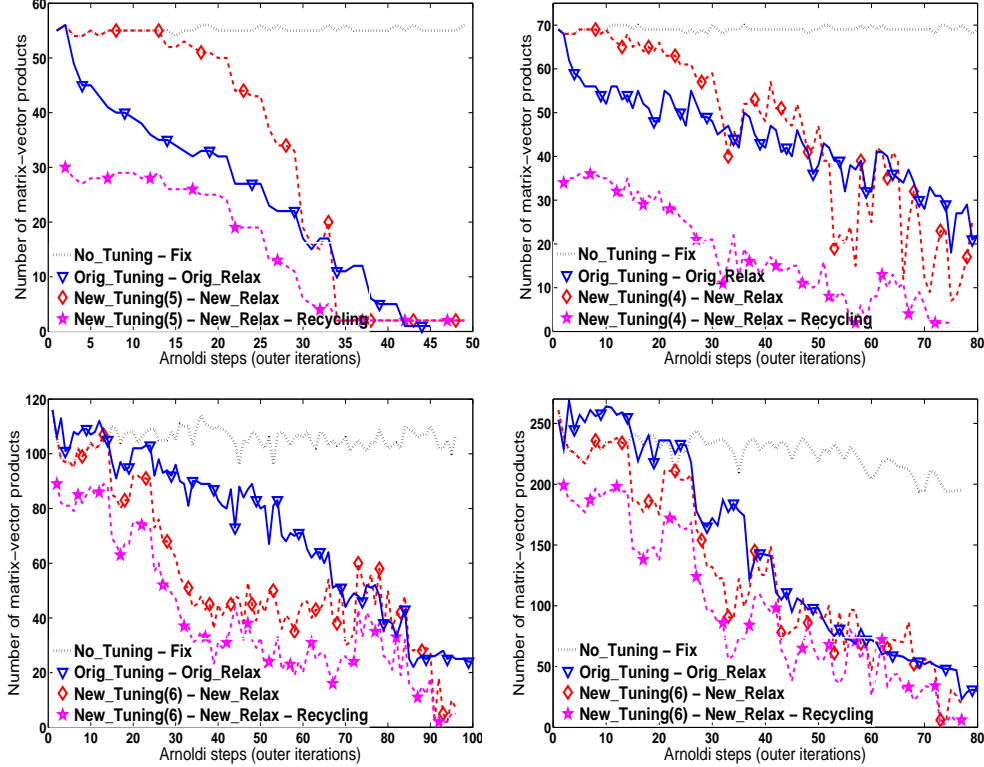

FIG. 5.4. *Performance of different strategies with relaxed tolerances of inner solves for Problems 1, 2, 3(a) and 4(a)*

The second type of strategy includes the two-phase algorithm and the relaxation strategy. The motivation of the two-phase algorithm is to generate a good approximate solution y_1 for $Ay = Bu_{k+j+1}^{(i)}$ by one GMRES step with the new tuning, so that the tolerance for the correction equation can be much larger than that of the original system. Additional larger tolerances can be obtained by the use of the relaxation strategy: as Section 4 shows, the allowable tolerances for $Ay = Bu_{k+j+1}^{(i)}$ are inversely proportional to the current eigenvalue residual norm. Therefore as the IRA method proceeds and converges to the desired invariant subspace, the relaxed tolerances keep increasing. Figure 5.4 shows the inner iteration counts required by four strategies with relaxed tolerances for solving $Ay = Bu_{k+j+1}^{(i)}$:

- "No Tuning-Fix" (dotted line) – Solve the original systems $Ay = Bu_{k+j+1}^{(i)}$ with preconditioned GMRES to the *fixed* tolerance $\delta_f = \frac{\epsilon}{2k}$. This performance of this strategy is already given in Figure 5.2; it is shown again to illustrate the performance improvement obtained by the following advanced strategies.
- "Orig_Tuning-Orig_Relax" ($\triangledown$, solid line) – Solve $Ay = Bu_{k+j+1}^{(i)}$ by GMRES with the original tuning to the *relaxed* tolerances δ_r given by the original estimate.
- "New_Tuning(5)-New_Relax" ($\diamondsuit$, dashed line) – Solve $Ay = Bu_{k+j+1}^{(i)}$ by the two-phase strategy to the new estimated tolerances δ_r in (4.16); tuning is constructed using solution vectors from the current and 5 previous IRA cycles.
- "New_Tuning(5)-New_Relax-Recycling" ($\bigstar$, dashed line) – Solve $Ay = Bu_{k+j+1}^{(i)}$ by by the two-phase strategy to the new estimated tolerances δ_r; in addition, subspace

recycling is used to solve the correction equations.

It is clear from Figure 5.4 that the relaxed tolerances help gradually reduce the inner iteration counts to very small numbers (curves with $\triangledown$, $\diamondsuit$, and $\bigstar$).

Figure 5.4 also shows the effectiveness of subspace recycling. For Problems 1 and 2, the use of this technique reduces the inner iteration counts by 40%–50% in initial Arnoldi steps (compare curves with $\diamondsuit$ to those with $\bigstar$). For Problems 3 and 4, where the original tuning does not perform well (see Figure 5.2), subspace recycling still decreases the inner iteration counts by numbers commensurate to the dimensions of recycled subspaces. As we discussed, subspace recycling achieves this improvement because it helps reduce the initial latencies of inner iterations. In addition, as observed in Section 3.3, the recycled subspaces obtained from one correction equation can be applied directly to the solution of the next equation, because the preconditioned system matrix is identical for the correction equations in all Arnoldi steps. This makes subspace recycling very cheap to use.

	No Tuning	New Tuning	New Tuning New Relaxation	**New Tuning New Relaxation Subspace Recycling**	New Tuning Original Relaxation Subspace Recycling	Original Tuning	**Original Tuning Original Relaxation**
Prob 1	2687	2430	1560	787	806	1842	1184
Prob 2	5524	5090	3631	1401	1469	4549	3494
Prob 3(a)	10114	7780	5334	3930	4163	10263	7619
Prob 3(b)	9966	8129	5775	4641	4786	9889	7193
Prob 4(a)	17321	14294	9934	7584	7897	17299	11365
Prob 4(b)	21186	17099	12570	9446	9635	21072	14424

TABLE 5.2

Inner iteration counts for different solution strategy for each problem

Table 5.2 summarizes the total inner iteration counts needed for each strategy for solving $Ay = B_{k+j+1}^{(i)}$ arising in inexact IRA. Here, "New Tuning + New Relaxation + Subspace Recycling" and "Original Tuning + Original Relaxation" are the most efficient strategies in this paper and [16] respectively. Clearly, the best approach is to combine the two-phase algorithm (with the new tuning), relaxation strategy and subspace recycling.

Finally, we discuss the two approaches assessing the allowable errors of inner solves in the Arnoldi steps. For all problems, we found that solution strategies with the new estimated allowable tolerances (4.16) require slightly smaller numbers of inner iterations than are needed for strategies with the original estimated tolerances. Table 5.2 shows that the new estimated tolerances help decrease the inner iteration counts by about 2%–5% (compare the "New Tuning + New Relaxation + Subspace Recycling" with "New Tuning + Original Relaxation + Subspace Recycling") when used with the two-phase strategy and subspace recycling. In fact, the new estimated tolerances tend to be a small multiple (say, 2 to 10) of the original estimated ones in most IRA cycles for all test problems.

Some heuristic remarks can be made for the two estimations. First, the substitution of $\min |\lambda(T_k) - \lambda(T_m^{22})|$ for $\operatorname{sep}(T_m^{22}, T_k)$ in the original estimation seems reasonable, in the sense that the former is usually not obviously larger than the latter. In fact, in the setting of eigenvalue computation, we expect two basic properties to hold: (1) the desired Ritz vectors generated by the Rayleigh-Ritz procedure is not far from the best approximation available in subspace from which the Ritz vectors are extracted, and (2) small eigenvalue residual of the desired approximate invariant subspace implies good eigenvector approximation. However, by analogy to the results in [36] and Chapter 2 of [19], both properties may not be true if $\operatorname{sep}(T_m^{22}, T_k)$ is considerably smaller than $\|T_m^{12}\|$ or $\min |\lambda(T_k) - \lambda(T_m^{22})|$ in our context. In the usual situations when the two properties hold, $\min |\lambda(T_k) - \lambda(T_m^{22})|$ is expected to be not much larger than $\operatorname{sep}(T_m^{22}, T_k)$. Second, numerical evidence help to understand why the new estimate tends to be slightly larger than the original estimate. In fact, note that $\min |\lambda(T_k) - \lambda(T_m^{22})|$ and $\max |\lambda(T_k) - \lambda(T_m^{22})|$ are the smallest and largest eigenvalue of the Sylvester operator $\mathcal{S}_m$ $(G \rightarrow \mathcal{S}_m(G) : T_m^{22}G - GT_k)$; see [35], page 17. For the test problems with spectral

transformation, it was consistently found that the largest eigenvalue of $\mathcal{S}_m$ is only about $10-100$ times larger than the smallest eigenvalue of $\mathcal{S}_m$, as long as the shift is not too close to an eigenvalue of the matrix pair (A, B). As the quantity in (4.14) used in the new estimation is always between the two extreme eigenvalues of $\mathcal{S}_m$ in practice, it is not surprising that this quantity tends to be a small multiple of $\min |\lambda(T_k) - \lambda(T_m^{22})|$. In conclusion, the original estimated allowable tolerance seems reasonably accurate for the test problems.

6. Conclusions. We have studied an inexact implicitly restarted Arnoldi (IRA) method for solving generalized eigenvalue problems with shift-invert and Cayley transformations, with focus on a few strategies that help reduce the inner iteration counts. We present a new tuning strategy using the solution vectors from the current and previous IRA cycles, and discuss a two-phase algorithm involving a correction equation for which the tolerance can be considerably bigger than that for the original system. In addition, subspace recycling can be used easily for the correction equation to further reduce the inner iteration counts. We analyze the allowable errors of matrix-vector products performed in Arnoldi steps and propose an alternative estimate of relaxed tolerances for the original linear systems. Numerical experiments show that the combined use of these strategies lead to significant speedup of inner iterations.

REFERENCES

[1] J. BERNS-MÜLLER AND A. SPENCE, *Inexact inverse iteration and GMRES*, Tech Report maths0507, University of Bath, Bath, UK, 2005

[2] J. BERNS-MÜLLER AND A. SPENCE, *Inexact inverse iteration with variable shift for nonsymmetric generalized eigenvalue problems*, SIAM Journal on Matrix Analysis and Applications, Vol. 28, No. 4 (2006), pp. 1069–1082.

[3] J. BERNS-MÜLLER, I. G. GRAHAM, AND A. SPENCE, *Inexact inverse iteration for symmetric matrices*, Linear Algebra and its Applications, Vol. 416, No. 2–3 (2006), pp. 389–413.

[4] A. BOURAS AND V. FRAYSSÉ, *A relaxation strategy for the Arnoldi method in eigenproblems*. Technical Report TR/PA/00/16, CERFACS, Toulouse, France, 2000.

[5] A. BOURAS AND V. FRAYSSÉ, *A relaxation strategy for inner-outer linear solvers in domain decomposition methods*. Technical Report TR/PA/00/16, CERFACS, Toulouse, France, 2000.

[6] A. BOURAS AND V. FRAYSSÉ, *Inexact matrix-vector products in Krylov methods for solving linear systems: a relaxation strategy*, SIAM Journal on Matrix Analysis and Applications, Vol. 26, No. 3 (2005), pp. 660–678.

[7] E. DE STURLER, *Truncation strategies for optimal Krylov subspace methods*, SIAM Journal on Numerical Analysis, Vol. 36, No. 3, (1999), pp. 864–889.

[8] H. C. ELMAN, A. R. RAMAGE, D. J. SILVESTER, AND A. J. WATHEN, *Incompressible Flow Iterative Solution Software Package*, http://www.cs.umd.edu/ elman/ifiss.html.

[9] H. C. ELMAN, A. R. RAMAGE, D. J. SILVESTER, *IFISS: A Matlab toolbox for modelling incompressible flow*, ACM Transactions on Mathematical Software, Vol. 33, No. 2, (2007), Article 14, 18 pages.

[10] H. C. ELMAN, D. J. SILVESTER AND A. J. WATHEN, *Finite Elements and Fast Iterative Solvers*, Oxford University Press, New York, 2005.

[11] H. C. ELMAN, D. J. SILVESTER AND A. J. WATHEN, *Performance and analysis of saddle point preconditioners for the discrete steady-state Navier-Stokes equations*, Numerische Mathematik, Vol. 90, No. 4, (2002), pp. 665–688.

[12] M. FREITAG AND A. SPENCE, *Convergence theory for inexact inverse iteration applied to the generalised nonsymmetric eigenproblem*, Electronic Transactions on Numerical Analysis, Vol. 28 (2007), pp. 40–64.

[13] M. FREITAG AND A. SPENCE, *Convergence rates for inexact inverse iteration with application to preconditioned iterative solves*, BIT Numerical Mathematics, Vol. 47, No.1 (2007), pp. 27–44.

[14] M. FREITAG AND A. SPENCE, *A tuned preconditioner for inexact inverse iteration applied to Hermitian eigenvalue problems*, IMA Journal on Numerical Analysis, Vol. 28, No. 3 (2007), pp. 522–551.

[15] M. FREITAG AND A. SPENCE, *Rayleigh quotient iteration and simplified Jacobi-Davidson method with preconditioned iterative solves*, Linear Algebra and its Applications, Vol. 428, No. 8–9 (2008), pp. 2049–2060.

[16] M. FREITAG AND A. SPENCE, *Shift-invert Arnoldi's method with preconditioned iterative solves*, SIAM Journal on Matrix Analysis and Applications, Vol. 31, No. 3 (2009), pp. 942–969.

[17] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations, 3rd Edition*, The Johns Hopkins University Press, Baltimore, 1996.

[18] G. H. GOLUB, Z. ZHANG, AND H. ZHA, *Large sparse symmetric eigenvalue problems with homogeneous linear constraints: the Lanczos process with inner-outer iterations*, Linear Algebra and its Applications, Vol. 309, No. 1-3 (2000), pp. 289–306.

[19] C.R. LEE, *Residual Arnoldi Methods: Theory, Package and Experiments*, Ph.D thesis, Department of Computer Science, University of Maryland, 2007.

[20] R. B. LEHOUCQ, D. C. SORENSEN, AND C. YANG, *ARPACK Users' Guide: Solution of Large Scale Eigenvalue Problems by Implicitly Restarted Arnoldi Methods.* SIAM, Philadelphia, 1998.

[21] K. MEERBERGEN, A. SPENCE AND D. ROOSE, *Shift-invert and Cayley transforms for detection of rightmost eigenvalues of nonsymmetric matrices*, BIT Numerical Mathematics, Vol. 34, No. 3, (1994), pp. 409–423.

[22] R. B. MORGAN, *GMRES with deflated restarting*, SIAM Journal on Scientific Computing, Vol. 24, No. 1, (2002), pp. 20–37.

[23] *Matrix Market*, http://math.nist.gov/MatrixMarket/

[24] R. B. MORGAN, *Computing interior eigenvalues of large matrices*, Linear Algebra Appl., 154–156 (1991), pp. 289–309.

[25] M. L. PARKS, E. DE STURLER, G. MACKEY, D. D. JOHNSON AND S. MAITI, *Recycling Krylov subspaces for sequences of linear systems*, SIAM Journal on Scientific Computing, Vol. 28, No. 5, (2006), pp. 1651–1674.

[26] M. ROBBÉ, M. SADKANE AND A. SPENCE, *Inexact inverse subspace iteration with preconditioning applied to non-Hermitian eigenvalue problems*, SIAM Journal on Matrix Analysis and Applications, Vol. 31, No. 1 (2009), pp. 92-113.

[27] V. SIMONCINI, *Variable accuracy of matrix-vector products in projection methods for eigencomputation*, SIAM Journal on Numerical Analysis, Vol. 43, No. 3 (2005), pp. 1155–1174.

[28] V. SIMONCINI AND L. ELDÉN, *Inexact Rayleigh quotient-type methods for eigenvalue computations*, BIT Numerical Mathematics, Vol. 42, No.1 (2002), pp.159–182.

[29] V. SIMONCINI AND D. B. SZYLD, *Theory of inexact Krylov subspace methods and applications to scientific computing*, SIAM Journal on Scientific Computing, Vol. 25, No. 2 (2003), pp. 454–477.

[30] V. SIMONCINI AND D. B. SZYLD, *Relaxed Krylov subspace approximation*, Proceedings in Applied Mathematics and Mechanics, Vol. 5, No. 1 (2005), pp. 797–800.

[31] V. SIMONCINI AND D. B. SZYLD, *Recent computational developments in Krylov subspace methods for linear systems*, Numerical Linear Algebra with Applications, Vol. 14, No. 1 (2007), pp. 1–59.

[32] G. L. G. SLEIJPEN AND J. VAN DEN ESHOF, *On the use of harmonic Ritz pairs in approximating internal eigenpairs*, Linear Algebra Appl., 358 (2003), pp. 115–137.

[33] G. L. G. SLEIJPEN, J. VAN DEN ESHOF AND M. B. VAN GIJZEN, *Relaxation strategies for nested Krylov methods*, Journal of Computational and Applied Mathematics, Volume 177, No. 2 (2005), pp. 347–365.

[34] D. C. SORENSEN, *Implicit application of polynomial filters in a k-step Arnoldi method*, SIAM Journal on Matrix Analysis and Applications, Vol. 13, No. 1 (1992), pp. 357–385.

[35] G. W. STEWART, *Matrix Algorithms Volumn II: Eigensystems*, SIAM, Philadelphia, 2001.

[36] G. W. STEWART, *A generalization of Saad's theorem on Rayleigh-Ritz approximations*, Linear Algebra and its Applications, Vol. 327, No. 1–3 (2001), pp. 115–119.

[37] G. W. STEWART, *An Unreliable Convergence Criterion for Arnoldi's Method*, Technical Report, CMSC TR-4938, University of Maryland, 2009.

[38] J. VAN DEN ESHOF AND G. L. G. SLEIJPEN, *Inexact Krylov subspace methods for linear systems*, SIAM Journal on Matrix Analysis and Applications, Vol. 26, No. 1 (2005), pp. 125–153.

[39] F. XUE AND H. ELMAN, *Fast inexact subspace iteration for generalized eigenvalue problems with spectral transformation*, submitted.