KRYLOV RECYCLING FOR SEQUENCES OF SHIFTED SYSTEMS*

MEGHAN J. O'CONNELL† AND MISHA E. KILMER†

Abstract. Solving many large scale sequences of shifted linear systems with multiple right-hand sides is the main computational bottleneck of many problems, such as nonlinear inverse problems. Specifically, this issue arises in the optimization problem found in diffuse optical tomography (DOT). An inner-outer Krylov recycling method was developed in [8] for the non-shifted case to reduce this computational cost. In this paper, we extend this approach to the shifted case for a single right-hand side. Both real shifts as well as complex identity shifts are investigated. We show the value of our approach with two examples from DOT, however, our method has the potential to be useful for other applications as well.

1. Introduction. The need to solve large scale sequences of shifted linear systems with multiple right-hand sides arises in many important applications such as solving nonlinear inverse problems. The expense of solving these systems can be the computational bottleneck of the larger problem in which they are involved. Many approaches have been developed to address the computational cost associated with solving sequences of shifted systems. Some methods use Lanczos recurrences or Arnoldi iterations, see e.g., [5, 7, 9, 11, 13]. Reduced order modeling also reduces this computational cost [3].

Imagine solving an optimization problem where the function evaluation requires the solution of a shifted system of right-hand sides over the course of the optimization. These system solves become the dominant computational cost. We were first motivated to develop the method presented in this paper by solving the optimization problem that arises in the nonlinear inverse problem in the context of diffuse optical tomography (DOT) [4, 6]. The goal of solving an inverse problem is to find an image of an unknown quantity. In order to find this image when the forward problem is nonlinear in the unknowns describing the desired quantity, we must solve an optimization problem, which in turn requires solving the forward problem multiple times. The forward model valuation, which requires the solution of a sequence of shifted systems, links the measured data to the unknown image we seek. These forward model solves are the main computational bottleneck of these image reconstruction problems. In this paper, we extend the 2-level Krylov recycling method found in [8] to systems with shifts in order to reduce this computational cost. At present, we only consider a single right-hand side.

We are looking to solve shifted systems of the form

$$\left(\mathbf{A}^{(k)} + \gamma_\ell \mathbf{E}\right) \mathbf{x}^{(k,\ell)} = \mathbf{b}, \tag{1.1}$$

for symmetric $\mathbf{A}^{(k)}$ and $\mathbf{E}$ and several values of k and ℓ via recycling. The method we describe works for more general applications, but we will demonstrate on two examples from DOT.

This paper is organized as follows. In Section 2, we provide an introduction to recycling for a single system and give an overview of inner-outer recycling. In Section 3, we present our algorithm for inner-outer recycling for shifted systems with a single right-hand side and give an analysis. In Section 4, we give numerical results for two different DOT problems. Conclusions and future work can be found in Section 5.

*This material is based upon work supported by the National Science Foundation under Grants No. NSF DMS 1217156 and 1217161 and NIH R01-CA154774.

†Department of Mathematics, Tufts University, Medford, MA 02115.

2. Background. We will begin by providing a brief discussion of recycling for a single system and a single right-hand side as explained in [7]. We will consider the linear system $\mathbf{Ax} = \mathbf{b}$, with symmetric $\mathbf{A} \in \mathbb{R}^{N\times N}$ and $\mathbf{b} \in \mathbb{R}^N$. Next, we explain inner-outer recycling as in [8] for the same system. More on recycling can be found in the literature, see [2, 7, 10, 15].

2.1. General Krylov Recycling. Let $\tilde{\mathbf{U}} \in \mathbb{R}^{N\times n_c}$ be given as the space over which we wish, initially, to look for solutions. Next, compute $\mathbf{A}\tilde{\mathbf{U}} = \tilde{\mathbf{K}}$. Set $\mathbf{K}$ to be the Q factor in the skinny QR factorization of $\tilde{\mathbf{K}}$ and $\mathbf{U} = \tilde{\mathbf{U}}\mathbf{R}^{-1}$, where $\mathbf{R}$ is the R factor. Now $\mathbf{K}^T\mathbf{K} = \mathbf{I}$. Next, we assume that the solution of $\mathbf{Ax} = \mathbf{b}$ is in Range($\mathbf{U}$) = Range($\tilde{\mathbf{U}}$) and find the approximate solution

$$\mathbf{x}_0 = \mathbf{U}\mathbf{K}^T\mathbf{b}, \tag{2.1}$$

giving an initial residual of $\mathbf{r} = \mathbf{b} - \mathbf{K}\mathbf{K}^T\mathbf{b}$. If this solution is not sufficient, then we expand $\mathbf{U}$ using a Lanczos recurrence with $(\mathbf{I}-\mathbf{K}\mathbf{K}^T)\mathbf{A}$ and $\mathbf{v}_1 = (\mathbf{I}-\mathbf{K}\mathbf{K}^T)\mathbf{b}/\|(\mathbf{I}-\mathbf{K}\mathbf{K}^T)\mathbf{b}\|_2$ to generate

$$\begin{aligned}(\mathbf{I}-\mathbf{K}\mathbf{K}^T)\mathbf{A}\mathbf{V}_m &= \mathbf{V}_{m+1}\underline{\mathrm{T}}_m \Leftrightarrow \\ \mathbf{A}\mathbf{V}_m &= \mathbf{K}\mathbf{K}^T\mathbf{A}\mathbf{V}_m + \mathbf{V}_{m+1}\underline{\mathrm{T}}_m\end{aligned} \tag{2.2}$$

where $\underline{\mathrm{T}}_m$ is an $m+1\times m$ tridiagonal matrix. [1] We then compute the approximate solution in Range($[\mathbf{V}_m\ \mathbf{U}]$) by looking for the solution that minimizes $\|\mathbf{b}-\mathbf{A}(\mathbf{V}_m\mathbf{y}+\mathbf{U}\mathbf{z})\|_2$, as follows:

$$\begin{aligned}&\min_{\mathbf{y},\mathbf{z}} \left\| \mathbf{b} - \mathbf{A}[\mathbf{U}\ \mathbf{V}_m] \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix} \right\|_2 \\ = &\min_{\mathbf{y},\mathbf{z}} \left\| \begin{bmatrix} \mathbf{K}^T\mathbf{b} \\ \xi\mathbf{e}_1 \end{bmatrix} - \begin{bmatrix} \mathbf{I} & \mathbf{K}^T\mathbf{A}\mathbf{V}_m \\ 0 & \underline{\mathrm{T}}_m \end{bmatrix} \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix} \right\|_2,\end{aligned} \tag{2.3}$$

where $\mathbf{e}_1$ denotes the first Cartesian basis vector in $\mathbb{R}^{m+1}$ and $\xi = \|(\mathbf{I}-\mathbf{K}\mathbf{K}^T)\mathbf{b}\|_2$. The solution can then be found by finding the solution to the projected problem

$$\min_{\mathbf{y}} \|\underline{\mathrm{T}}_m\mathbf{y} - \xi\mathbf{e}_1\|_2,$$

then computing the $\mathbf{z}$ that satisfies (2.3), and finally, with $\mathbf{y}_m := \mathbf{V}_m\mathbf{y}$, setting $\mathbf{x} = \mathbf{y}_m + \mathbf{U}\mathbf{z}$. Since $\mathbf{V}_m\mathbf{y}$ is computed via short term recurrences (MINRES), we do not have to explicitly store $\mathbf{V}_m$.

2.2. Inner-Outer Krylov Recycling. As was shown in [8], we can adapt recycling both to solve systems efficiently and also to create a projection matrix $\mathbf{V}$, which we can use to project our systems to smaller ones via Galerkin Projection. In the DOT application, where we need to solve for multiple right-hand sides $\mathbf{b}_j$ for each $\mathbf{A}^{(k)}$, this means we replace solves with (1.1) by solves with $(\mathbf{V}^T(\mathbf{A}^{(k)}+\gamma_\ell\mathbf{E})\mathbf{V})\tilde{\mathbf{x}} = \tilde{\mathbf{b}}_j$ and set $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$. The method in [8] involves maintaining two recycle spaces. Let $\mathbf{U}$ be a matrix whose columns span the outer-most recycle space. As described in [8], $\mathbf{U}$ contains information from all systems and all right-hand sides and will ultimately become the projection matrix. If the solution in this space is "good enough" as measured by

[1] The matrix recurrence in (2.2) is unchanged if $(\mathbf{I}-\mathbf{K}\mathbf{K}^T)\mathbf{A}$ is replaced by $(\mathbf{I}-\mathbf{K}\mathbf{K}^T)\mathbf{A}(\mathbf{I}-\mathbf{K}\mathbf{K}^T)$ since $\mathbf{A}$ is symmetric.

relative residual norm, the system is not further treated. Otherwise, a correction to the solution is computed using recycling with a smaller recycle space (smaller matrix $\mathbf{U}_j$). The details are as follows for solving the linear system $\mathbf{Ax} = \mathbf{b}$.

Let $\tilde{\mathbf{U}}, \mathbf{U}, \mathbf{K}$ as defined above. In addition, we let $\tilde{\mathbf{U}}_j \subset \tilde{\mathbf{U}} \in \mathbb{R}^{N \times n_j}$ be given, and then define $\mathbf{U}_j$ from $\tilde{\mathbf{U}}_j$ analogous to how $\tilde{\mathbf{U}}$ is computed to $\mathbf{U}$. That is, $\mathbf{AU}_j = \mathbf{K}_j$ and $\mathbf{K}_j^T\mathbf{K}_j = \mathbf{I}$. Note Range($\mathbf{U}_j$) $\subset$ Range($\mathbf{U}$), which ensures Range($\mathbf{K}_j$) $\subset$ Range($\mathbf{K}$). For more details about the selection for the DOT application, see [8].

If we decompose $\mathbf{b}_j$ using the orthogonal projector $\mathbf{KK}^T$, we get

$$\begin{aligned} \mathbf{Ax} &= (\mathbf{I} - \mathbf{KK}^T)\mathbf{b}_j + \mathbf{KK}^T\mathbf{b}_j \\ \mathbf{Ax} - \mathbf{KK}^T\mathbf{b} &= (\mathbf{I} - \mathbf{KK}^T)\mathbf{b}_j \\ \mathbf{A}\underbrace{(\mathbf{x} - \mathbf{UK}^T\mathbf{b})}_{\mathbf{g}} &= \mathbf{r}_j. \end{aligned} \tag{2.4}$$

Since $\mathbf{g}$ is the incremental change from the initial guess of $\mathbf{UK}^T\mathbf{b}_j$, found in (2.1), it would make sense to solve (2.4) in order to construct a projection basis without redundant information. However, let us assume $\mathbf{U}$ has too many columns to use effectively as a recycle space due to the need to orthogonalize against $\mathbf{K}$ at each iteration. Then if the initial residual, $\mathbf{r}_j$, is not small enough, we solve

$$\min_{\mathbf{y},\mathbf{z}} \left\| \mathbf{r}_j - \mathbf{A}[\mathbf{U}_j, \mathbf{V}_m] \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix} \right\|,$$

where we use a Lanczos recurrence with $(\mathbf{I} - \mathbf{K}_j\mathbf{K}_j^T)\mathbf{A}$ and the same $\mathbf{v}_1 = (\mathbf{I} - \mathbf{KK}^T)\mathbf{b}/\|(\mathbf{I} - \mathbf{KK}^T)\mathbf{b}\|_2$ as before to generate $\mathbf{V}_m$ satisfying

$$\begin{aligned} (\mathbf{I} - \mathbf{K}_j\mathbf{K}_j^T)\mathbf{AV}_m = \mathbf{V}_{m+1}\underline{\mathrm{T}}_m \Leftrightarrow \\ \mathbf{AV}_m = \mathbf{K}_j\mathbf{K}_j^T\mathbf{AV}_m + \mathbf{V}_{m+1}\underline{\mathrm{T}}_m. \end{aligned} \tag{2.5}$$

Since $\mathbf{r}_j$ is already orthogonal to Range($\mathbf{K}$), and Range($\mathbf{K}_j$) $\subset$ Range($\mathbf{K}$), we can show that $\mathbf{y}, \mathbf{z}$ can be found by solving

$$\min_{\mathbf{y},\mathbf{z}} \left\| \begin{bmatrix} 0 \\ \xi\mathbf{e}_1 \end{bmatrix} - \begin{bmatrix} \mathbf{I} & \mathbf{K}_j^T\mathbf{AV}_m \\ 0 & \underline{\mathrm{T}}_m \end{bmatrix} \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix} \right\|_2.$$

Therefore, using $\mathbf{y}_m := \mathbf{V}_m\mathbf{y}$, we have $\mathbf{g} = \mathbf{y}_m + \mathbf{U}_j\mathbf{z}$, where $\mathbf{z} = -\mathbf{K}_j^T\mathbf{AV}_m\mathbf{y}$ and we can get back to the original solution, $\mathbf{x} = \mathbf{y}_m - \mathbf{U}_j\mathbf{K}_j^T\mathbf{Ay}_m + \mathbf{UK}^T\mathbf{b}_j$. Since $\mathbf{y}_m$ contains the information not already in Range($\mathbf{U}$), we update both $\mathbf{U}$ and $\mathbf{U}_j$ by appending only that part of the solution.

3. Inner-Outer Recycling for Shifted Systems. We will now give our new extension of inner-outer recycling for shifted systems with a single right-hand side. The extension to shifted systems needs to address the following: how to recover the solution to the shifted problem and what information from the shifted solution to include in the recycle spaces. We consider solving

$$\left(\mathbf{A}^{(k)} + \gamma_\ell\mathbf{E}\right)\mathbf{x}^{(k,\ell)} = \mathbf{b}, \tag{3.1}$$

with a variant of inner-outer recycling, as we now describe.

3.1. The Method. Without loss of generality, we will assume that the first value of γ_ℓ we wish to solve for is 0, and we need to solve (3.1) for L values of γ_ℓ, between $\gamma_1 = 0$ and γ_L. This is due to the fact we can always create a "zero shift" by letting $\widehat{\mathbf{A}^{(k)}} = (\mathbf{A}^{(k)} + \gamma_1\mathbf{E})$ for the smallest shift, γ_1, we want to solve, and then we write the remaining shifts relative to γ_1: $\gamma_\ell = \gamma_1 + \widehat{\gamma_\ell}$, and solve the remaining systems by rewriting the matrices as $(\widehat{\mathbf{A}^{(k)}} + \widehat{\gamma_\ell}\mathbf{E})$.

As before, assume $\mathbf{U} \in \mathbb{R}^{N\times n_c}$ is generated from the recycle space $\tilde{\mathbf{U}}$ such that $\mathbf{A}^{(k)}\mathbf{U} = \mathbf{K}$ and $\mathbf{K}^T\mathbf{K} = \mathbf{I}$. Select $\tilde{\mathbf{U}}_\ell \in \mathbb{R}^{N\times n_\ell}$ from among the columns of $\tilde{\mathbf{U}}$. Generate $\mathbf{U}_\ell$ from $\tilde{\mathbf{U}}_\ell$, such that $(\mathbf{A}^{(k)} + \gamma_\ell\mathbf{E})\mathbf{U}_\ell = \mathbf{K}_\ell$ and $\mathbf{K}_\ell^T\mathbf{K}_\ell = \mathbf{I}$. For the shifted system, $\mathbf{U}_\ell$ will be a shift specific recycle space instead of a right-hand side specific recycle space as in subsection 2.2.

First, we estimate that our solution is in Range($\mathbf{U}$) (i.e. it is $\mathbf{Uq}$ for some $\mathbf{q}$),

$$\begin{aligned}\mathbf{r}^{(k,\ell)} &= \mathbf{b} - \left(\mathbf{A}^{(k)} + \gamma_\ell\mathbf{E}\right)\mathbf{Uq}\\ &= \mathbf{b} - \left(\mathbf{Kq} + \gamma_\ell\mathbf{EUq}\right).\end{aligned}$$

Using Petrov-Galerkin projection, we put $\mathbf{K}^T\mathbf{r}^{(k,\ell)} = 0$, and find the estimate

$$\mathbf{x}^{(k,\ell)} \approx \mathbf{U}\left(\mathbf{I} + \gamma_\ell\mathbf{K}^T\mathbf{EU}\right)^{-1}\mathbf{K}^T\mathbf{b}. \tag{3.2}$$

When $\gamma_\ell = 0$, this gives the same initial solution as (2.1)

As in (2.4), we would like to solve for the incremental change from this initial guess, if the initial residual $\mathbf{r}^{(k,\ell)}$ is not small enough. Using $\mathbf{x}^{(k,\ell)} = \mathbf{g} + \mathbf{x}_i$, where $\mathbf{x}_i$ is the initial guess found on the right of (3.2), the goal is to recover $\mathbf{g}$. Therefore, the problem we want to solve is

$$\begin{aligned}\left(\mathbf{A}^{(k)} + \gamma_\ell\mathbf{E}\right)\mathbf{g} &= \mathbf{b} - \left(\mathbf{A}^{(k)} + \gamma_\ell\mathbf{E}\right)\mathbf{x}_i \qquad (3.3)\\ &= (\mathbf{I} - \mathbf{KK}^T)\mathbf{b} - (\mathbf{I} - \mathbf{KK}^T)(\mathbf{EU})(\frac{1}{\gamma_\ell}\mathbf{I} + \mathbf{K}^T\mathbf{EU})^{-1}\mathbf{K}^T\mathbf{b}\end{aligned}$$

The last line comes from the fact that $\mathbf{K}^T\mathbf{r}^{(k,\ell)} = 0$ and is of value in terms of illustrating the relationship to the residual for the $\gamma_\ell = 0$ system.

If the initial residual is not small enough, we solve

$$\min_{\mathbf{g}\in\mathcal{S}} \|\mathbf{r}^{(k,\ell)} - \left(\mathbf{A}^{(k)} + \gamma_\ell\mathbf{E}\right)\mathbf{g}\|_2,$$

over an appropriate $\mathcal{S}$. We want $\mathcal{S}$ to contain Range($\mathbf{U}_\ell$) and to generate the space with which to augment $\mathbf{U}_\ell$, we use the Krylov subspace generated by $(\mathbf{I} - \mathbf{K}_\ell\mathbf{K}_\ell^T)\left(\mathbf{A}^{(k)} + \gamma_\ell\mathbf{E}\right)$ and

$$\mathbf{v}_1 = \left(\left(\mathbf{I} - \mathbf{K}_\ell\mathbf{K}_\ell^T\right)\mathbf{r}^{(k,\ell)}\right) / \| \left(\mathbf{I} - \mathbf{K}_\ell\mathbf{K}_\ell^T\right)\mathbf{r}^{(k,\ell)}\|_2.$$

We call the reader's attention to the following important fact, which makes this inner-outer recycling approach different than the MRHS case in [8]: $\mathbf{K}$ is in Range($\mathbf{A}^{(k)}$), while $\mathbf{K}_\ell$ is in Range($\mathbf{A}^{(k)} + \gamma_\ell\mathbf{E}$). We obtain

$$\begin{aligned}(\mathbf{I} - \mathbf{K}_\ell\mathbf{K}_\ell^T)\left(\mathbf{A}^{(k)} + \gamma_\ell\mathbf{E}\right)\mathbf{V}_m &= \mathbf{V}_{m+1}\underline{\mathrm{T}}_m \Leftrightarrow\\ \left(\mathbf{A}^{(k)} + \gamma_\ell\mathbf{E}\right)\mathbf{V}_m &= \mathbf{K}_\ell\mathbf{K}_\ell^T\left(\mathbf{A}^{(k)} + \gamma_\ell\mathbf{E}\right)\mathbf{V}_m + \mathbf{V}_{m+1}\underline{\mathrm{T}}_m.\end{aligned}$$

We can then find $\mathbf{y}, \mathbf{z}$ by solving

$$\min_{\mathbf{y},\mathbf{z}} \left\| \begin{bmatrix} \mathbf{K}_\ell^T \mathbf{r}^{(k,\ell)} \\ \xi \mathbf{e}_1 \end{bmatrix} - \begin{bmatrix} \mathbf{I} & \mathbf{K}_\ell^T \left(\mathbf{A}^{(k)} + \gamma_\ell \mathbf{E}\right) \mathbf{V}_m \\ 0 & \underline{\mathrm{T}}_m \end{bmatrix} \begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix} \right\|_2, \tag{3.4}$$

where $\xi = \| \left(\mathbf{I} - \mathbf{K}_\ell \mathbf{K}_\ell^T\right) \mathbf{r}^{(k,\ell)} \|_2$. Therefore, $\mathbf{g} = \mathbf{y}_m + \mathbf{U}_\ell \mathbf{z}$, where $\mathbf{y}_m = \mathbf{V}_m \mathbf{y}$, $\mathbf{z} = \mathbf{K}_\ell^T \mathbf{r}^{(k,\ell)} - \mathbf{K}_\ell^T \left(\mathbf{A}^{(k)} + \gamma_\ell \mathbf{E}\right) \mathbf{V}_m \mathbf{y}$ and we estimate the solution as

$$\mathbf{x}^{(k,\ell)} = \mathbf{y}_m + \mathbf{U}_\ell \mathbf{K}_\ell^T \mathbf{r}^{(k,\ell)} - \mathbf{U}_\ell \mathbf{K}_\ell^T \left(\mathbf{A}^{(k)} + \gamma_\ell \mathbf{E}\right) \mathbf{y}_m + \mathbf{x}_i. \tag{3.5}$$

3.2. Identifying and Updating Recycle Spaces. Now that we have found the solution to the shifted problem via recycling, we will discuss how to construct our recycle spaces. In [8], the initial recycle spaces included an invariant subspace consisting of 10 eigenvectors corresponding to the smallest eigenvalues and solutions to the initial system. This was because the corresponding subspace was found to be nearly invariant across all $\mathbf{A}^{(k)}$. In addition, $\mathbf{U}$ was seeded with solutions to the initial system for all right-hand sides, while $\mathbf{U}_j$ contained only the approximate invariant subspace and the solution to the j^{th} right-hand side. We adopt an analogous strategy here, but $\mathbf{U}$ will have solutions to the initial system for all shifts and $\mathbf{U}_\ell$ will have only the initial solution to the system corresponding to the shift γ_ℓ.

If the initial residual for shift ℓ is such that we need to perform the inner recycling, we append information from $\mathbf{y}_m$ to both $\mathbf{U}$ and $\mathbf{U}_\ell$. This ensures that $\mathbf{U}$ contains information pertinent to the entire set of shifted systems, while $\mathbf{U}_\ell$ is kept small. This is important for keeping computational costs down.

Note that appending a column to $\mathbf{U}$ means that $\mathbf{K}$ must also be increased by one column. We must (a) compute $\mathbf{A}^{(k)}\mathbf{y}_m$ and (b) orthogonalize the result against the previous columns of $\mathbf{K}$ to get our new $\mathbf{K}$. We then (c) update the initial residual estimate *for any shifted systems we haven't yet solved* to reflect the projection onto the space as increased in dimension by 1. In this way, new information about the systems that are "close by" in terms of neighboring γ_ℓ is used to improve the current solution space.

The overhead involved in appending information to $\mathbf{U}_\ell$ is the cost of computing $(\mathbf{A}^{(k)} + \gamma_\ell \mathbf{E})\mathbf{y}_m$ (but note that $\mathbf{A}^{(k)}\mathbf{y}_m$ has already been computed) and then orthogonalizing the result against the previous columns of $\mathbf{K}_\ell$. Note again that we only append columns when there's a system for which the inner Krylov recycling became necessary to reduce the residual.

The next subsection outlines our algorithm (but note that the sequential updating approaches noted here have been replaced by more expensive calls to qr factorizations to keep the algorithm looking more tidy and make the general idea easier to follow.)

3.3. The Algorithm. Algorithm 1 describes our inner-outer recycling process for shifted systems with one right-hand side.

3.4. Complex Identity Shift. We will now consider the special case of a complex identity shift, that is, $\mathbf{E} = \mathbf{I}$ and the shift is $i\gamma_\ell$. Therefore, the problem is

$$\left(\mathbf{A}^{(k)} + i\gamma_\ell \mathbf{I}\right) \mathbf{x}^{(k,\ell)} = \mathbf{b}. \tag{3.6}$$

The algorithm provided in the previous subsection will work for this special case, but we need to be careful about how we construct and update $\mathbf{U}$ and $\mathbf{U}_\ell$. We will initialize and update $\mathbf{U}$ and $\mathbf{U}_\ell$ just as in Algorithm 1, but if we are dealing with a complex

Algorithm 1: Krylov Recycling for Shifted Systems

$\mathbf{U}_0 \Leftarrow 10$ eigenvectors of $\mathbf{A}^{(0)}$
$\mathbf{X}^{(0)} \Leftarrow$ solutions to $\left(\mathbf{A}^{(0)} + \gamma_\ell \mathbf{E}\right)\mathbf{X}^{(0)} = \mathbf{b}$ for all γ_ℓ
$\mathbf{U} \Leftarrow$ basis for Range$([\mathbf{U}_0, \mathbf{X}^{(0)}])$
$\mathbf{U}_\ell \Leftarrow [\mathbf{U}_0, \mathbf{X}^{(0)}(:,\ell)]$
for $k = 1 : K$ **do**
  for $\ell = 1 : L$ **do**
    % Check if $\mathbf{U}$ is a good enough space
    $\tilde{\mathbf{K}} = \mathbf{A}^{(k)}\mathbf{U}$
    $[\mathbf{K}, \mathbf{R}] = qr(\tilde{\mathbf{K}}, 0)$
    $\mathbf{U} = \mathbf{U}/\mathbf{R}$
    $\mathbf{r}^{(k,\ell)} = \mathbf{b} - (\mathbf{K} + \gamma_\ell \mathbf{E}\mathbf{U})\left(\mathbf{I} + \gamma_\ell \mathbf{K}^T\mathbf{E}\mathbf{U}\right)^{-1}\mathbf{K}^T\mathbf{b}$
    if $\frac{||\mathbf{r}^{(k,\ell)}||}{||\mathbf{b}||} > tol$ **then**
      % MINRES recycling using $\mathbf{U}_\ell$
      $\tilde{\mathbf{K}}_\ell = \left(\mathbf{A}^{(k)} + \gamma_\ell \mathbf{E}\right)\mathbf{U}_\ell$
      $[\mathbf{K}_\ell, \mathbf{R}] = qr(\tilde{\mathbf{K}}_\ell, 0)$
      $\mathbf{U}_\ell = \mathbf{U}_\ell/\mathbf{R}$
      Solve (3.4)
      Find $\mathbf{x}^{(k,\ell)}$ by solving (3.5)
      $\mathbf{r}^{(k,\ell)} = (\mathbf{I} - \mathbf{K}_\ell\mathbf{K}_\ell^T)(\mathbf{r}^{(k,\ell)} - \left(\mathbf{A}^{(k)} + \gamma_\ell \mathbf{E}\right)\mathbf{y}_m)$
      $\mathbf{U} \Leftarrow [\mathbf{U}, \mathbf{y}_m]$
      $\mathbf{U}_\ell \Leftarrow [\mathbf{U}_\ell, \mathbf{y}_m]$
    end
  end
end

shift, we only append the imaginary component of $\mathbf{x}^{(0,\ell)}$ and $\mathbf{y}_m$. This means that $\mathbf{U}$ and $\mathbf{K}$ will remain real. Let $\tilde{\mathbf{U}}_\ell = \tilde{\mathbf{U}}(:, i_1 : i_2)$, and $[\mathbf{K}_\ell, \mathbf{R}] = (\mathbf{A}^{(k)} + i\gamma_\ell \mathbf{I})\tilde{\mathbf{U}}_\ell$, so $\mathbf{U}_\ell = \tilde{\mathbf{U}}_\ell \mathbf{R}^{-1}$. It is clear that $\mathbf{K}_\ell$, $\mathbf{U}_\ell$, and $\mathbf{V}_m$ will be complex. In the case when many shifts are used, it may be necessary to add information from the real component of some $\mathbf{x}^{(0,\ell)}$ to the initial $\mathbf{U}$. This is because as you move farther away from the "zero shift" the solutions can change significantly and only adding the complex component is not sufficient. We also note that letting $\mathbf{E} = \mathbf{I}$ provides some savings in (3.3) and (3.4) since we no longer need to find $\mathbf{E}\mathbf{U}$ or $\mathbf{E}\mathbf{y}_m$.

3.5. Algorithm Analysis. We will now discuss the computational costs associated with Algorithm 1. We include an invariant subspace in $\mathbf{U}$ and $\mathbf{U}_\ell$ formed from eigenvectors corresponding to the smallest eigenvalues from our initial non-shifted system. While Krylov solvers can usually take advantage of the shift invariant property [14], our method of recycling cannot take advantage of this property since the presence of the shift changes the space where $\mathbf{U}_\ell$ gets mapped. Even though the invariant subspace may not deflate the spectrum for the shifted systems, it will deflate part of the spectrum for the non-shifted problem. Additionally, the right-hand side of (3.3) for the non-shifted and shifted problem is made small across spectral components in the $\mathbf{K}$ direction. This is the large $\mathbf{K}$ and not the smaller, shift specific $\mathbf{K}_\ell$.

There is a cost associated with the QR factorizations done to compute both $\mathbf{K}$

and $\mathbf{K}_\ell$. However, the initial work can be done up front and saved for many inverse problems if the same $\mathbf{E}$ is used. Once inside the loop over shifts, we are only adding one column at a time to both $\mathbf{U}$ and $\mathbf{U}_j$. As was shown in [8], we can compute $\mathbf{K}$ and $\mathbf{K}_\ell$ such that we do not need to do a full re-orthogonalization every time they are computed. An efficient alternative for computing the initial $\mathbf{K}_\ell$ is given in section 3.2 of [12], but there is a potential trade off in accuracy.

For problems with many shifts, the overhead cost associated with our method is offset by the decreasing number of iterations. The overhead cost in the initialization can be computed off-line for the DOT problem, since $\mathbf{A}^{(0)}$ is the same across many images. Therefore, the eigenvectors can be precomputed and reused for many problems. If the same shifts are used across problems as well, the basis for the Range($[\mathbf{U}_0, \mathbf{X}^{(0)}]$) can also be precomputed. The overhead costs for the main loop of the algorithm include, updating $\mathbf{K}$ and finding the initial residual and solution. If we have to do recycling, we have to update $\mathbf{K}_\ell$, solve (3.4), and update the solution and residual. We argue that small number of MINRES iterations required offsets these costs.

4. Numerical Results. In this section, we will look at how we can use the inner-outer recycling for shifted systems in the DOT setting for two different shifts. In DOT, we are looking to recover images of optical absorption in human tissue. We model the photon flux/fluence with a time-domain diffusion model. The inverse problem involves using the data captured at the detectors by illuminating the tissue with signal sources to determine the absorption, $\mu(\mathbf{x})$. In order to reduce our search space, we will adopt the PaLs approach, as in [1], to parametrize the absorption field, $\mu(\cdot)$. This means that we assume that we can express $\mu(\cdot)$ in terms of a finite set of parameters, $\mathbf{p} = [p_1, \ldots, p_{n_p}]^T$, giving us $\mu(\cdot) = \mu(\cdot, \mathbf{p})$. The optimization problem we ultimately want to solve is,

$$\min_{\mathbf{p}\in\mathbb{R}^\ell} \|\mathcal{M}(\mathbf{p}) - \mathbb{D}\|_2, \tag{4.1}$$

where $\mathcal{M}(\mathbf{p})$ is a vector of estimated observations for all sources and all frequencies and $\mathbb{D}$ is the corresponding vector from the acquired data. All of the experiments were run using a laptop with a 3.20 GHz processor and 16.0 GB RAM using MATLAB R2015b.

4.1. Experiment 1. For the DOT problem, we discretize the diffusion equation in the frequency domain. A single function evaluation amounts to solving $\left(\mathbf{A}^{(k)} + i\gamma_\ell \mathbf{I}\right)\mathbf{X}^{(k,\ell)} = \mathbf{B}$ for all ℓ at which we have collected data, then computing $\mathbf{C}^T\mathbf{X}^{(k,\ell)}$, where $\mathbf{C}$ represents the detectors and $\mathbf{B}$ the sources. $\mathbf{A}^{(k)}$ represents the discretized diffusion operator with varying absorption, so the k-th system corresponds to evaluating the diffusion equation for an absorption coefficient that depends on the k-th estimate of parameters, $\mathbf{p}^{(k)}$. In 4.1, $\mathcal{M}(\mathbf{p})$ is formed by stacking the vectorized form of $\mathbf{C}^T\mathbf{X}^{(k,\ell)}$ for all frequencies. Though this application requires we solve each system for both multiple shifts and multiple right-hand sides, we will compute the solution for each across all shifts, independently, using our new approach. Extension of our method to handle multiple right-hand sides is the subject of on-going research.

In this numerical experiment we consider a 201×201 mesh, and discretize with finite differences as in [6], which gives us 40401 degrees of freedom for the forward problem. We adopt the technique in [8] to identify $\mathbf{A}^{(k)}$ with a SPD matrix, which now has 39999 degrees of freedom. We will use one right-hand side from the DOT problem, which will be a multiple of a column of the identity matrix. The shift in this application is, $\gamma_\ell = \frac{2\pi 10^6 \omega_\ell}{\nu}$, where ω_ℓ is the frequency and ν is the speed of light

in the medium. We will solve the problem for 21 frequencies, $0 : 10 : 200$ MHz, and 10 systems. As was stated above, since we are using many shifts we will add the real component, as well as the imaginary component, of the initial solution for $\ell = 6$, 11, and 16 to the initial $\mathbf{U}$. Table 4.1 shows the number of (unpreconditioned) MINRES iterations for a sample of systems and shifts for this experiment with and without recycling. It also shows the initial relative residuals and the number of columns of $\mathbf{U}$ and $\mathbf{U}_\ell$. We used a tolerance of 10^{-7}. Figure 4.1 shows the number of MINRES iterations for each shift and all systems. It is clear that the iterations generally decrease from one shift to the next, and system to system, using our approach.

		Our Approach				MINRES
System	Shift	Initial Relative Residual	Cols $\mathbf{U}$	Cols $\mathbf{U}_\ell$	Its	Its
1	1	6.997652e-05	34	11	138	463
	5	1.488217e-07	38	11	11	463
	13	1.249818e-07	46	11	6	460
	21	1.703870e-07	54	11	12	455
3	1	3.110010e-05	76	13	117	491
	5	1.273549e-07	80	13	8	490
	13	1.388263e-07	88	13	7	487
	21	1.185384e-07	96	13	4	480
5	1	1.903012e-05	118	15	92	480
	5	1.228727e-07	122	15	5	479
	13	1.555352e-07	130	15	8	476
	21	1.192817e-07	138	15	4	471
7	1	4.306398e-06	160	17	49	481
	5	1.130465e-07	163	17	3	481
	13	1.089243e-07	171	17	2	478
	21	1.210036e-07	179	17	4	473
9	1	2.146463e-06	200	19	29	481
	5	1.063270e-07	201	19	1	481
	13	1.020278e-07	208	19	1	478
	21	1.236055e-07	216	19	4	473

Table 4.1: Comparison of the inner-outer approach for shifted systems as described in Algorithm 1 vs. MINRES for Experiment 1.

4.2. Experiment 2. In Experiment 2, we will be using the same $\mathbf{A}^{(k)}$ matrices and the same right-hand side as in Experiment 1. $\mathbf{E}$ will be a real valued, diagonal matrix representing a random perturbation of the background of the image of the absorption coefficient. The shifts will be 0, .01, .02, .03, .04. We might encounter this scenario if we were to try to gather statistical or sensitivity information during the optimization process. Table 4.2 shows the number of (unpreconditioned) MINRES iterations for Experiment 2 with and without recycling. We used a tolerance of 10^{-7}. Once again, the iterations generally decrease from one shift to the next, and system to system, using our approach.

5. Conclusions and Future Work. We developed an inner-outer Krylov recycling approach for solving sequences of shifted systems with a single right-hand side. Two numerical examples show the success of our approach in the DOT setting.

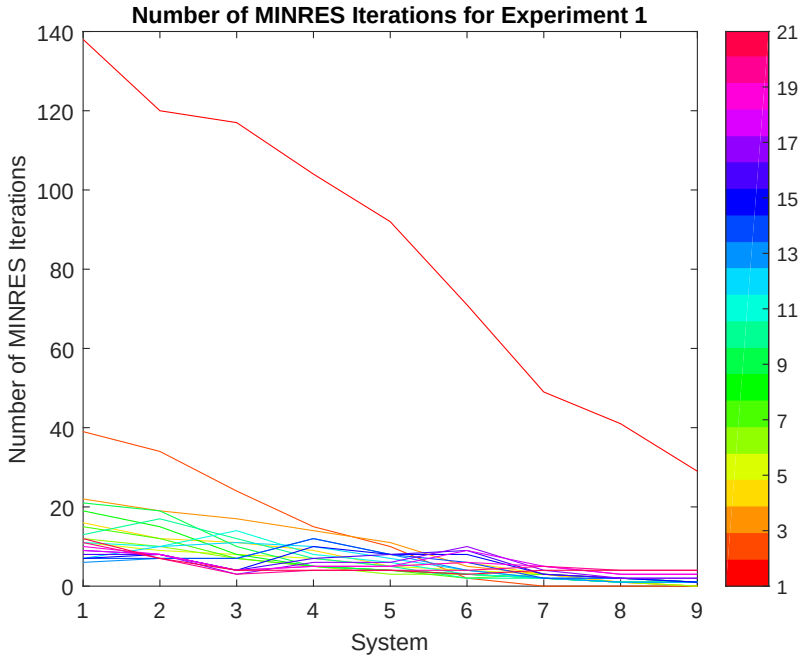

Fig. 4.1: The number of MINRES iterations for Experiment 1 for all shifts. The iteration counts for the "zero shift" are found in the red curve above the rest. Shifts 2 and 3 are the next two largest curves for the earlier systems, while shift 21 is the next largest by system 9.

We are currently investigating different ways to construct $\mathbf{U}$ and $\mathbf{U}_\ell$. For the complex identity shift, we are looking at better selection strategies for adding the real component from the initial solution to $\mathbf{U}$. We are also considering how we could use the same $\mathbf{K}_\ell$ across all shifts. In addition, we are looking at ways to identify when we might need to purge or refresh the information in our recycle spaces.

In this paper, only one right-hand side was considered, but we would like to extend the method to multiple right-hand sides. Since our recycle space, $\mathbf{U}$, could get very large for multiple right-hand sides, we are considering ways to bin information for shifts that are "close". In the DOT problem, we would also want to use the recycle space constructed from inner-outer recycling for shifted systems with multiple right-hand sides as a projection basis for reduced order modeling.

REFERENCES

[1] A. Aghasi, M. Kilmer, and E. Miller. Parametric level set methods for inverse problems. *SIAM J. Imaging Sci.*, 4(2):618–650, 2011.

[2] K. Ahuja, E. de Sturler, S. Gugercin, and E. Chang. Recycling BiCG with an application to model reduction. *SIAM J. Imaging Sci.*, 34(4):A1925–A1949, 2012.

[3] A. Antoulas, C. Beattie, and S. Gugercin. Interpolatory model reduction of large-scale dynamical systems. In J. Mohammadpour and K. Grigoriadis, editors, *Efficient Modeling and Control of Large-Scale Systems*. Springer-Verlag, 2010.

		Our Approach				MINRES
System	Shift	Initial Relative Residual	Cols $\mathbf{U}$	Cols $\mathbf{U}_\ell$	Its	Its
1	1	7.451061e-05	15	11	140	463
	2	8.910013e-07	16	11	45	458
	3	1.816563e-07	17	11	17	452
	4	1.429132e-07	18	11	14	447
	5	1.104739e-07	19	11	3	443
2	1	4.933852e-05	20	12	128	474
	2	6.396233e-07	21	12	42	469
	3	1.951564e-07	22	12	19	464
	4	1.477872e-07	23	12	16	460
	5	1.176433e-07	24	12	8	455
3	1	3.723381e-05	25	13	128	491
	2	4.602222e-07	26	13	36	484
	3	2.086688e-07	27	13	25	477
	4	1.347959e-07	28	13	9	471
	5	1.173205e-07	29	13	6	465
4	1	2.747551e-05	30	14	115	485
	2	3.213505e-07	31	14	22	479
	3	1.722536e-07	32	14	17	473
	4	1.285942e-07	33	14	8	468
	5	1.207813e-07	34	14	8	463

Table 4.2: Comparison of the inner-outer approach for shifted systems as described in Algorithm 1 vs. MINRES for Experiment 2.

[4] S. R. Arridge. Optical tomography in medical imaging. *Inverse Problems*, 15(2):R41–R93, 1999.

[5] D. Darnell, R. Morgan, and W. Wilcox. Deflated GMRES for systems with multiple shifts and multiple right-hand sides. *Linear Algebra and its Applications*, 429(10):2415–2434, 2008.

[6] E. de Sturler, S. Gugercin, M. Kilmer, S. Chaturantabut, C. Beattie, and M. O'Connell. Nonlinear parametric inversion using interpolatory model reduction. *SIAM J. Sci. Comput.*, 37(3):B495–B517, 2015.

[7] M. Kilmer and E. de Sturler. Recycling subspace information for diffuse optical tomography. *SIAM J. Sci. Comput.*, 27(6):2140–2166, 2006.

[8] M. Kilmer, M. O'Connell, E. de Sturler, S. Gugercin, and C. Beattie. Computing reduced order models via inner-outer Krylov recycling in diffuse optical tomography. *TBS*.

[9] K. Meerbergen. The solution of parameterized symmetric linear systems. *SIAM J. Matrix Anal. & Appl.*, 24(4):1038–1059, 2003.

[10] M. Parks, E. de Sturler, G. Mackey, D. Johnson, and S. Maiti. Recycling Krylov subspaces for sequences of linear systems. *SIAM J. Sci. Comput.*, 28(5):1651–1674, 2006.

[11] A. Saibaba, T. Bakhos, and P. Kitanidis. A flexible Krylov solver for shifted systems with application to oscillatory hydraulic tomography. *SIAM J. Sci. Comput.*, 35(6):A3001–A3023, 2013.

[12] A. Saibaba, M. Kilmer, E. Miller, and S. Fantini. Fast algorithms for hyperspectral diffuse optical tomography. *SIAM J. Sci. Comput.*, 37(5):B712–B743, 2015.

[13] V. Simoncini. Restarted full orthogonalization method for shifted linear systems. *BIT Numerical Mathematics*, 42(2):459–466, 2003.

[14] V. Simoncini and D. Szyld. Recent computational developments in Krylov subspace methods for linear systems. *Numer. Linear Algebra Appl.*, 14:1–59, 2007.

[15] S. Wang, E. de Sturler, and G. Paulino. Large-scale topology optimization using preconditioned Krylov subspace methods with recycling. *Int. J. Numer. Meth. Engng.*, 69(12):2441–2468, 2007.