BIT manuscript No.
(will be inserted by the editor)

Tikhonov regularization via flexible Arnoldi reduction

Lothar Reichel · Xuebo Yu

Received: date / Accepted: date

Abstract Flexible GMRES, introduced by Saad, is a generalization of the standard GMRES method for the solution of large linear systems of equations. It is based on the flexible Arnoldi process for reducing a large square matrix to a small matrix. We describe how the flexible Arnoldi process can be applied to implement one-parameter and multi-parameter Tikhonov regularization of linear discrete ill-posed problems. The method proposed is well suited for large-scale problems. Moreover, computed examples show that our method can give approximate solutions of higher accuracy than available direct methods for small-scale problems.

Keywords ill-posed problem · Tikhonov regularization · Arnoldi process · flexible GMRES

Mathematics Subject Classification (2010) 65R30, 65R32, 65F10, 65F22

1 Introduction

We consider the problem of computing an approximate solution of linear systems of equations

$$A\boldsymbol{x} \approx \boldsymbol{b} \tag{1.1}$$

with a large matrix $A \in \mathbb{R}^{n\times n}$ whose singular values gradually decay to zero (without a significant gap). Hence, A is severely ill-conditioned and may be

Lothar Reichel
Department of Mathematical Sciences
Kent State University, Kent, OH 44242, USA
E-mail: reichel@math.kent.edu

Xuebo Yu
Department of Mathematical Sciences
Kent State University, Kent, OH 44242, USA
E-mail: xyu@math.kent.edu

singular. Linear systems of equations with a matrix of this kind are commonly referred to as linear discrete ill-posed problems. They stem, for instance, from the discretization of linear ill-posed problems, such as Fredholm integral equations of the first kind with a smooth kernel. The right-hand side $\boldsymbol{b} \in \mathbb{R}^n$ of linear discrete ill-posed problems that arise in applications in science and engineering represents available data and typically is contaminated by an error $\boldsymbol{e} \in \mathbb{R}^n$ that may stem from measurement inaccuracies, discretization error, and electronic noise in the device used (such as in computerized tomography). Let $\hat{\boldsymbol{b}} \in \mathbb{R}^n$ denote the (unknown) noise-free vector associated with $\boldsymbol{b}$, i.e.,

$$\boldsymbol{b} = \hat{\boldsymbol{b}} + \boldsymbol{e}, \tag{1.2}$$

and assume that the unavailable linear system of equations with the noise-free right-hand side,

$$A\boldsymbol{x} = \hat{\boldsymbol{b}}, \tag{1.3}$$

is consistent. Let $A^\dagger$ denote the Moore–Penrose pseudoinverse of A. We would like to determine an approximation of $\hat{\boldsymbol{x}} = A^\dagger \hat{\boldsymbol{b}}$ by computing a suitable approximate solution of (1.1).

Straightforward solution of (1.1) generally does not yield a meaningful approximation of $\hat{\boldsymbol{x}}$ due to the ill-conditioning of A and the error $\boldsymbol{e}$ in $\boldsymbol{b}$. Therefore, one often first replaces (1.1) by a minimization problem that is less sensitive to the error $\boldsymbol{e}$, and then solves the latter problem. This replacement is known as regularization.

One of the most popular regularization methods is due to Tikhonov. This method replaces (1.1) by a penalized least-squares problem of the form

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \mu \|B\boldsymbol{x}\|^2 \right\}, \tag{1.4}$$

where $B \in \mathbb{R}^{n\times n}$ is referred to as the regularization matrix and $\mu > 0$ as the regularization parameter. Throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm or the associated induced matrix norm. When B is the identity matrix, the Tikhonov minimization problem (1.4) is said to be in *standard form*, otherwise it is said to be in *general form*. We assume that B is chosen so that

$$\mathcal{N}(A) \cap \mathcal{N}(B) = \{\boldsymbol{0}\}, \tag{1.5}$$

where $\mathcal{N}(M)$ denotes the null space of the matrix M. Then (1.4) has the unique solution

$$\boldsymbol{x}_\mu = (A^*A + \mu B^*B)^{-1} A^*\boldsymbol{b}$$

for any $\mu > 0$. Here M^* denotes the adjoint of the matrix M. The value of μ determines how sensitive $\boldsymbol{x}_\mu$ is to the error $\boldsymbol{e}$ and how close $\boldsymbol{x}_\mu$ is to $\hat{\boldsymbol{x}}$; see, e.g., Engl et al. [13] and Hansen [17] for discussions on Tikhonov regularization.

We would like to compute the solution $\boldsymbol{x}_\mu$ of (1.4), or an approximation thereof, that is close to $\hat{\boldsymbol{x}}$. This requires the determination of a suitable value of the regularization parameter $\mu > 0$. The latter generally requires the solution of (1.4) for several values of μ. For instance, the determination of μ by the

discrepancy principle, the L-curve criterion, or generalized cross validation all require that (1.4) be solved for several values of $\mu > 0$; see [13,17,25,32] for detailed discussions on these and other methods for computing a suitable value of μ. We note that the repeated solution of (1.4) for different μ-values can be expensive when the matrices A and B are large.

When A and B are small to moderately sized matrices, the computational effort required for solving (1.4) for several μ-values can be reduced by first computing the generalized singular value decomposition (GSVD) of the matrix pair $\{A, B\}$. The GSVD determines the decompositions

$$A = \widetilde{U}\widetilde{\Sigma}\widetilde{X}, \qquad B = \widetilde{V}\widetilde{M}\widetilde{X}, \tag{1.6}$$

where $\widetilde{U}, \widetilde{V} \in \mathbb{R}^{n\times n}$ are orthogonal, $\widetilde{\Sigma}, \widetilde{M} \in \mathbb{R}^{n\times n}$ are diagonal with nonnegative entries, and $\widetilde{X} \in \mathbb{R}^{n\times n}$ is nonsingular; see, e.g., [23] for discussions on this decomposition. When the GSVD of the pair $\{A, B\}$ is available, the computations of the solution $\boldsymbol{x}_\mu$ for a particular value of $\mu > 0$ requires only $\mathcal{O}(n^2)$ arithmetic floating point operations (flops). However, the computation of the decomposition (1.6) is expensive; $\mathcal{O}(n^3)$ flops are needed. Therefore, the evaluation of the GSVD of $\{A, B\}$ is not attractive when the matrices A and B are large.

We are interested in developing a new efficient method for the approximate solution of large-scale Tikhonov minimization problems (1.4) and describe how the flexible Arnoldi process can be applied to first reduce the matrix A in (1.4) to small size. The matrix B is restricted to the same solution subspace as A, and this restriction can be represented by a small matrix. Then we solve the reduced Tikhonov minimization problem determined by these small matrices by a direct method, e.g., by computing the GSVD of the pair of reduced matrices. We remark that our reduction method may be attractive to use also for small to medium sized problems, because it may yield more accurate approximations of $\hat{\boldsymbol{x}}$ than the GSVD. This is illustrated in Section 5.

Application of $\ell \ll n$ steps of the flexible Arnoldi process to A with a (not necessarily orthonormal) user-chosen solution subspace basis $\{\boldsymbol{v}_i\}_{i=1}^{\ell}$ in $\mathbb{R}^n$ and a user-chosen unit vector $\boldsymbol{u}_1 \in \mathbb{R}^n$ yields a decomposition of the form

$$AV_\ell = U_{\ell+1}H_{\ell+1,\ell}, \tag{1.7}$$

where $V_\ell = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_\ell] \in \mathbb{R}^{n\times\ell}$, the matrix $U_{\ell+1} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{\ell+1}] \in \mathbb{R}^{n\times(\ell+1)}$ has orthonormal columns, and $H_{\ell+1,\ell} \in \mathbb{R}^{(\ell+1)\times\ell}$ is an upper Hessenberg matrix. Note that the decomposition (1.7) represents the QR factorization

$$[\boldsymbol{u}_1, AV_\ell] = U_{\ell+1}[\boldsymbol{e}_1, H_{\ell+1,\ell}],$$

where $\boldsymbol{e}_1 = [1, 0, \ldots, 0]^* \in \mathbb{R}^{\ell+1}$ denotes the first axis vector. We assume that ℓ is small enough so that no breakdown occurs. This is the generic situation. The decomposition (1.7) is the basis for the flexible GMRES (FGMRES) iterative method introduced by Saad [35] for the solution of large linear systems of equations with a nonsingular matrix A. The range of V_ℓ, denoted by $\mathcal{R}(V_\ell)$, is the solution subspace for FGMRES. Saad's interest in FGMRES stems from

the fact that the method allows the application of a sequence of different preconditioners while building up the solution subspace. An application of FGMRES to the solution of large linear discrete ill-posed problems has recently been described by Morikuni et al. [29].

Throughout this paper, we let the first column of the matrix $U_{\ell+1}$ in (1.7) be defined by

$$\boldsymbol{u}_1 = \boldsymbol{b}/\|\boldsymbol{b}\|. \tag{1.8}$$

Our reduction method is determined by the decomposition (1.7) of A and the QR factorization

$$BV_\ell = W_\ell R_\ell, \tag{1.9}$$

where the matrix $W_\ell \in \mathbb{R}^{n\times\ell}$ has orthonormal columns and $R_\ell \in \mathbb{R}^{\ell\times\ell}$ is upper triangular. Restricting the search subspace in (1.4) to $\mathcal{R}(V_\ell)$, substituting the decompositions (1.7) and (1.9) into (1.4), and using (1.8), yields the reduced Tikhonov minimization problem

$$\min_{\boldsymbol{y}\in\mathbb{R}^\ell} \left\{ \|H_{\ell+1,\ell}\boldsymbol{y} - \boldsymbol{e}_1\|\boldsymbol{b}\|\,\|^2 + \mu\|R_\ell\boldsymbol{y}\|^2 \right\}. \tag{1.10}$$

The solution $\boldsymbol{y}_{\ell,\mu}$ of (1.10) gives the approximate solution

$$\boldsymbol{x}_{\ell,\mu} = V_\ell \boldsymbol{y}_{\ell,\mu} \tag{1.11}$$

of (1.4) and (1.1). We will discuss the solution of the reduced problem (1.10) in Section 3.

The matrices $U_{\ell+1}$ and V_ℓ in (1.7) generated by Algorithm 2.1 below have orthonormal columns. Therefore, the condition number $\kappa(H_{\ell+1,\ell}) = \|H_{\ell+1,\ell}\|\|H_{\ell+1,\ell}^\dagger\|$ is an increasing function of ℓ and is bounded above by $\kappa(A)$. In particular, $\kappa(H_{\ell+1,\ell})$ can be quite large already for small or modest values of ℓ. It is therefore meaningful to regularize the reduced problem (1.10).

The reduction method of this paper is inspired by the generalized Arnoldi process described by Li and Ye [26] for reducing a pair of square matrices to smaller matrices. Their method differs from our in several respects. For instance, in the method in [26] only the first vector $\boldsymbol{v}_1$ of the solution subspace can be prescribed, and the determination of a solution subspace of dimension ℓ requires the computation of an orthonormal basis of dimension $2\ell+1$. Therefore, the decomposition of the present paper requires somewhat less computational effort to determine a solution subspace of the same dimension. Moreover, when the solution subspace is of dimension ℓ, the reduced matrices obtained from A and B by the generalized Arnoldi process [26] are of larger sizes than the matrices $H_{\ell+1,\ell}$ and R_ℓ in (1.7) and (1.9), respectively. Because of these differences with the generalized Arnoldi process, we feel that it is worthwhile to develop and investigate properties of the method of the present paper. Application of the generalized Arnoldi process to Tikhonov regularization is described in [33]; see also Section 5.

Both the reduction methods by Li and Ye [26] and the one of the present paper require the matrices A and B to be square. While several of the regularization matrices commonly used in the literature are rectangular, see, e.g.,

Hansen [17] and Kilmer et al. [24], the development of square regularization matrices has recently received considerable attention and a variety of square regularization matrices are available; see [3,8,9,34].

Several approaches to solving (1.4) when the matrices A and B are large have been described in the literature. Eldén [12] discusses how the problem (1.4) can be transformed to standard form with the aid of the A-weighted generalized inverse of B. The standard form problem then can be solved by applying a standard Krylov subspace iterative method, such as partial Golub–Kahan bidiagonalization. This approach requires B to have a special structure to be feasible. For instance, when B is banded with a small bandwidth or an orthogonal projector, the matrix-vector products with the transformed matrix can be evaluated efficiently; see, e.g., [12,28]. However, this is not the case when B is a general square matrix without particular structure. Kilmer et al. [24] describe an inner-outer iterative method for the solution of (1.4) with a general matrix $B \in \mathbb{R}^{m\times n}$, where m is allowed to be different from n. The inner iterations implement a scheme by Zha [38] for determining approximations of generalized singular vectors associated with the largest generalized singular values of the matrix pair $\{A, B\}$. This method produces nice results but can be expensive when many inner iterations are required. Hochstenbach et al. [22] present a reduction method that allows the matrices A and B to be rectangular, but requires matrix-vector product evaluations not only with A and B, but also with their transposes. The method of this paper, as described by Algorithm 2.1 below, does not use A^* and B^*. This allows application to problems for which the matrices A or B are not explicitly formed, but only matrix-vector products are evaluated with these matrices. This situation arises, for instance, in certain nonlinear problems where A represents the Jacobian of a nonlinear operator; see, e.g., [6] for a discussion. The method in [21] determines a solution subspace independent of B.

A careful comparison of the available methods is outside the scope of the present paper. The choice of method should depend on the problem to be solved and on whether matrix-vector product evaluations with A^* and B^* can be computed. The computed examples of Section 5 show that it can be important to determine a solution subspace that depends not only on A but also on B. Moreover, some problems benefit from the use of A^* when building the solution subspace. Algorithm 2.1 of Section 2 can be modified to accommodate this. This is illustrated in Example 5.5.

Throughout this paper, we denote the Krylov subspace of dimension at most ℓ determined by the matrix $M \in \mathbb{R}^{n\times n}$ and vector $\boldsymbol{c} \in \mathbb{R}^n$ by

$$\mathcal{K}_\ell(M, \boldsymbol{c}) = \text{span}\{\boldsymbol{c}, M\boldsymbol{c}, \ldots, M^{\ell-1}\boldsymbol{c}\}.$$

Generically, the dimension of this space is ℓ.

This paper is organized as follows. Our reduction method for the matrix pair $\{A, B\}$ is described in Section 2. Its application to the solution of the Tikhonov minimization problem (1.4) is discussed in Section 3. An extension of the reduction method of Section 2 to matrix tuplets is considered in Section 4, where also an application to multi-parameter Tikhonov regularization is

outlined. A few numerical examples are presented in Section 5, and Section 6 contains concluding remarks.

2 A reduction method for matrix pairs

This section discusses our choice of the matrix V_ℓ in (1.7) and describes an algorithm for computing the decompositions (1.7) and (1.9) for increasing values of ℓ. Recall that the initial column of the matrix $U_{\ell+1}$ in (1.7) is given by (1.8).

The columns of the matrix $V_\ell = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_\ell]$ span the solution subspace; cf. (1.11). They can be chosen in many ways. For instance, letting $\boldsymbol{v}_1 = \boldsymbol{u}_1 = \boldsymbol{b}/\|\boldsymbol{b}\|$, we may generate the remaining columns of the matrices $U_{\ell+1}$ and V_ℓ in (1.7) by the standard Arnoldi process. Then

$$\mathcal{R}(V_\ell) = \mathcal{K}_\ell(A, \boldsymbol{b}) \tag{2.1}$$

and the columns of V_ℓ agree with the first ℓ columns of $U_{\ell+1}$. In particular, the columns of V_ℓ are orthonormal. We assume for notational simplicity that the generic situation $\dim(\mathcal{K}_\ell(A, \boldsymbol{b})) = \ell$ holds.

Alternatively, we may let $\boldsymbol{v}_1 = A\boldsymbol{b}/\|A\boldsymbol{b}\|$ and determine the remaining columns of V_ℓ with the range restricted Arnoldi method [30, Algorithm 2.2]. Then the columns of V_ℓ are orthonormal with

$$\mathcal{R}(V_\ell) = \mathcal{K}_\ell(A, A\boldsymbol{b}). \tag{2.2}$$

Solution subspaces that are determined by matrix-vector product evaluation with both A and A^* are advocated in [21], where the columns of V_ℓ are computed by partial Golub–Kahan bidiagonalization of A, i.e.,

$$\mathcal{R}(V_\ell) = \mathcal{K}_\ell(A^*A, A^*\boldsymbol{b}). \tag{2.3}$$

The approaches to generate the solution subspace $\mathcal{R}(V_\ell)$ in (2.1), (2.2), and (2.3) do not use the matrix B to determine the columns of $V_\ell = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_\ell]$.

The present paper describes the construction of generalized Krylov subspaces and their application as solution subspaces. These spaces contain information about both the matrices A and B, which is gathered by evaluating matrix-vector products. In iteration j with Algorithm 2.1 below, one computes matrix-vector products $A\boldsymbol{v}_j$ and $B\boldsymbol{v}_j$, where $\boldsymbol{v}_j$ is the jth column of V_ℓ. The vector $A\boldsymbol{v}_j$ is orthogonalized against the columns of $U_j = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_j]$ and normalized to yield the last column $\boldsymbol{u}_{j+1}$ of the matrix U_{j+1} and, similarly, $B\boldsymbol{v}_j$ is orthogonalized against the columns of $W_{j-1} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{j-1}]$ and normalized to give the last column $\boldsymbol{w}_j$ of the matrix W_j. Thus,

$$\boldsymbol{u}_{j+1} = \frac{(I - U_jU_j^*)A\boldsymbol{v}_j}{\|(I - U_jU_j^*)A\boldsymbol{v}_j\|}, \qquad \boldsymbol{w}_j = \frac{(I - W_{j-1}W_{j-1}^*)B\boldsymbol{v}_j}{\|(I - W_{j-1}W_{j-1}^*)B\boldsymbol{v}_j\|}.$$

The solution subspace is then expanded by the last column of U_{j+1} or W_j that has not already been used to expand the solution subspace. The parameter

$\rho > 0$ in the algorithm specifies the ratio of the number of columns from $U_{\ell+1}$ and W_ℓ that are used to expand the solution subspace. Thus, the value $\rho = 1$ indicates that every other vector used to expand the solution subspace is a column of the matrix $U_{\ell+1}$. When $\rho = 5$ the solution subspace is expanded by five columns from $U_{\ell+1}$ for every column from W_ℓ; see Proposition 2.2 for details. We have found that for certain Tikhonov minimization problems (1.4), it is beneficial to choose $\rho \neq 1$. The counters $N(\boldsymbol{u})$ and $N(\boldsymbol{w})$ in Algorithm 2.1 keep track of how many columns from the matrices $U_{\ell+1}$ and W_ℓ, respectively, have been included in the solution subspace at step j of the algorithm. We discuss properties of Algorithm 2.1 below.

Algorithm 2.1 Flexible Arnoldi reduction of a matrix pair $\{A, B\}$.

1. *Input: matrices $A, B \in \mathbb{R}^{n\times n}$, vector $\boldsymbol{b} \in \mathbb{R}^n$, ratio $\rho > 0$, and number of steps ℓ*
2. $h_{1,1} := \|\boldsymbol{b}\|$; $\boldsymbol{u}_1 := \boldsymbol{b}/h_{1,1}$; $\boldsymbol{v}_1 := \boldsymbol{u}_1$;
3. $N(\boldsymbol{u}) := 1$; $N(\boldsymbol{w}) := 1$
4. **for** $j = 1, 2, \ldots, \ell$ **do**
5. $\quad \widehat{\boldsymbol{u}} := A\boldsymbol{v}_j$ — new $\boldsymbol{u}$-vector
6. $\quad$ **for** $i = 1, 2, \ldots, j$ **do**
7. $\quad\quad h_{i,j} := \boldsymbol{u}_i^* \widehat{\boldsymbol{u}}$; $\widehat{\boldsymbol{u}} := \widehat{\boldsymbol{u}} - \boldsymbol{u}_i h_{i,j}$
8. $\quad$ **end for**
9. $\quad h_{j+1,j} := \|\widehat{\boldsymbol{u}}\|$
10. $\quad \boldsymbol{u}_{j+1} := \widehat{\boldsymbol{u}}/h_{j+1,j}$ — *if $h_{j+1,j} = 0$: see text*
11. $\quad \widehat{\boldsymbol{w}} := B\boldsymbol{v}_j$ — new $\boldsymbol{w}$-vector
12. $\quad$ **for** $i = 1, 2, \ldots, j-1$ **do**
13. $\quad\quad r_{i,j} := \boldsymbol{w}_i^* \widehat{\boldsymbol{w}}$; $\widehat{\boldsymbol{w}} := \widehat{\boldsymbol{w}} - \boldsymbol{w}_i r_{i,j}$
14. $\quad$ **end for**
15. $\quad r_{j,j} := \|\widehat{\boldsymbol{w}}\|$
16. $\quad \boldsymbol{w}_j := \widehat{\boldsymbol{w}}/r_{j,j}$ — *if $r_{j,j} = 0$: see text*
17. $\quad$ **if** $N(\boldsymbol{w})/N(\boldsymbol{u}) > 1/\rho$
18. $\quad\quad \boldsymbol{v} := \boldsymbol{u}_{N(\boldsymbol{u})+1}$; $N(\boldsymbol{u}) := N(\boldsymbol{u}) + 1$
19. $\quad$ **else**
20. $\quad\quad \boldsymbol{v} := \boldsymbol{w}_{N(\boldsymbol{w})}$; $N(\boldsymbol{w}) := N(\boldsymbol{w}) + 1$
21. $\quad$ **end**
22. $\quad$ **for** $i = 1, 2, \ldots, j$ **do**
23. $\quad\quad \boldsymbol{v} := \boldsymbol{v} - (\boldsymbol{v}_i^* \boldsymbol{v})\boldsymbol{v}_i$;
24. $\quad$ **end for**
25. $\quad \alpha_j := \|\boldsymbol{v}\|$;
26. $\quad \boldsymbol{v}_{j+1} := \boldsymbol{v}/\alpha_j$; — new $\boldsymbol{v}$-vector
27. **end for**

Algorithm 2.1 determines the orthonormal columns $\boldsymbol{u}_j$, $\boldsymbol{v}_j$, and $\boldsymbol{w}_j$ of the matrices $U_{\ell+1}$, V_ℓ, and W_ℓ, respectively, in the decompositions (1.7) and (1.9). In addition, the algorithm computes the last column of $V_{\ell+1}$. The elements $h_{i,j}$ and $r_{i,j}$ determined by Algorithm 2.1 are the nontrivial entries of the upper Hessenberg matrix $H_{\ell+1,\ell}$ in (1.7) and of the upper triangular matrix R_ℓ in (1.9), respectively. Lines 5-10 of the algorithm describe one step of the flexible Arnoldi process introduced by Saad [35]. This process is said to break down at step j if $h_{j+1,j}$ vanishes in line 9. The computations can be continued after breakdown by letting $\boldsymbol{u}_{j+1}$ in line 10 be an arbitrary unit vector that is orthogonal to the already generated vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_j$. Here and throughout this paper, we say that breakdown occurs when the vanishing of a quantity makes it necessary to take special measures in order to be able

to continue with the computations. In an actual implementation of Algorithm 2.1, special measures have to be taken already when pertinent quantities, such as $h_{j+1,j}$ in line 9, are of small enough magnitude. Since the occurrence of breakdown is very rare, we will not dwell on the implementation details further.

Lines 12-16 update the QR factorization of the matrix BV_{j-1} to yield a QR factorization of BV_j, where $V_k = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k]$ for $k \in \{j, j+1\}$. Daniel et al. [7] provide detailed discussions on updating and downdating of QR factorizations. The updating process is said to break down if $r_{j,j}$ vanishes in line 15. The computations can be continued after breakdown by letting $\boldsymbol{w}_j$ in line 16 be an arbitrary unit vector that is orthogonal to the already generated vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{j-1}$. We remark that the occurrence of breakdown is rare unless the vectors $\boldsymbol{v}_j$ are chosen to be in $\mathcal{N}(B)$.

Lines 22-26 orthogonalize the new vector for the solution subspace against the already available orthonormal solution subspace basis. Breakdown occurs when $\alpha_j = 0$ in line 25. In this situation, the vector $\boldsymbol{v}_{j+1}$ in line 26 can be chosen to be a unit vector that is orthogonal to the already available orthonormal vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_j$.

Proposition 2.1 *Let $A, B \in \mathbb{R}^{n\times n}$. Then application of j steps of Algorithm 2.1 to the matrix pair $\{A, B\}$ with breakdowns handled as described above gives a flexible Arnoldi decomposition (1.7) and a QR factorization (1.9) for any $1 \le j \le n$.*

The following results illustrate the role of the parameter ρ in Algorithm 2.1.

Proposition 2.2 *Assume that no breakdown occurs in Algorithm 2.1. When ρ is sufficiently large, the algorithm yields after $\ell-1$ steps the solution subspace (2.1). If instead $\rho > 0$ is sufficiently small, we have*

$$\mathcal{R}(V_\ell) = \mathcal{K}_\ell(B, \boldsymbol{b}).$$

The solution subspace above and (2.1) are standard Krylov subspaces. For most values of $0 < \rho < \infty$ and ℓ sufficiently large, Algorithm 2.1 determines a solution subspace that is a generalized Krylov subspace. In particular,

$$\begin{aligned}
\rho = 1: &\quad \mathcal{R}(V_\ell) = \text{span}\{\boldsymbol{b}, B\boldsymbol{b}, A\boldsymbol{b}, B^2\boldsymbol{b}, AB\boldsymbol{b}, BA\boldsymbol{b}, A^2\boldsymbol{b}, B^3\boldsymbol{b}, \ldots\},\\
\rho = 5: &\quad \mathcal{R}(V_\ell) = \text{span}\{\boldsymbol{b}, A\boldsymbol{b}, A^2\boldsymbol{b}, A^3\boldsymbol{b}, A^4\boldsymbol{b}, B\boldsymbol{b}, A^5\boldsymbol{b}, AB\boldsymbol{b}, A^6\boldsymbol{b}, \ldots\},\\
\rho = \tfrac{1}{5}: &\quad \mathcal{R}(V_\ell) = \text{span}\{\boldsymbol{b}, B\boldsymbol{b}, B^2\boldsymbol{b}, B^3\boldsymbol{b}, B^4\boldsymbol{b}, B^5\boldsymbol{b}, A\boldsymbol{b}, B^6\boldsymbol{b}, BA\boldsymbol{b}, \ldots\},
\end{aligned}$$

where ℓ is the number of elements included in the span counted from left to right.

Proof The proposition can be shown by direct computations. □

Line 2 of Algorithm 2.1 sets $\boldsymbol{v}_1 := \boldsymbol{u}_1$. Many other choices of $\boldsymbol{v}_1$ are possible and may be attractive. For instance, $\boldsymbol{v}_1 = A\boldsymbol{u}_1/\|A\boldsymbol{u}_1\|$ and ρ very large yield the solution subspace (2.2). When B has an explicitly known nontrivial null space of dimension i, it may be advantageous to choose the first i vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_i$ of Algorithm 2.1 to be an orthonormal basis for $\mathcal{N}(B)$. This choice requires a simple modification of the algorithm and is illustrated in Example 5.2..

3 Tikhonov regularization

Restricting the solution subspace in (1.4) to $\mathcal{R}(V_\ell)$, substituting the decompositions (1.7) and (1.9) into (1.4), and letting $\boldsymbol{x} = V_\ell \boldsymbol{y}$, yields the small minimization problem (1.10). This problem can be solved in a variety of ways for several values of $\mu > 0$. One approach, already mentioned in Section 1, is to determine the solution $\boldsymbol{y}_{\ell,\mu}$ of (1.10) with the aid of the GSVD of the matrix pair $\{H_{\ell+1,\ell}, R_\ell\}$. An approximate solution of (1.4) is then furnished by (1.11). It may be faster to compute the solution $\boldsymbol{y}_{\ell,\mu}$ of (1.10) by using the decomposition described in [10]. Finally, when the upper triangular matrix R_ℓ in (1.10) is not very ill-conditioned, i.e., when its condition number $\kappa(R_\ell)$ is not very large, we can transform the Tikhonov minimization problem (1.10) to standard form

$$\min_{\boldsymbol{z}\in\mathbb{R}^\ell} \left\{ \|\widetilde{H}_{\ell+1,\ell}\boldsymbol{z} - \|\boldsymbol{b}\|\boldsymbol{e}_1\|^2 + \mu \|\boldsymbol{z}\|^2 \right\}, \qquad \boldsymbol{z} = R_\ell \boldsymbol{y}, \tag{3.1}$$

where

$$\widetilde{H}_{\ell+1,\ell} = H_{\ell+1,\ell} R_\ell^{-1}. \tag{3.2}$$

The minimization problem (3.1) can be solved in several ways, for instance, by first determining the entries of $\widetilde{H}_{\ell+1,\ell}$ by solving the system of equations $\widetilde{H}_{\ell+1,\ell} R_\ell = H_{\ell+1,\ell}$ column by column, and then computing the singular value decomposition (SVD) of the matrix (3.2). Eldén [11] describes how, instead of computing the SVD, one can reduce the computational effort by only transforming $\widetilde{H}_{\ell+1,\ell}$ to bidiagonal form.

In the numerical examples of Section 5, we assume a bound ε for the norm of the error $\boldsymbol{e}$ in $\boldsymbol{b}$ to be available. This assumption is reasonable in some applications. If no bound for $\|\boldsymbol{e}\|$ is known, then various methods can be applied to determine an estimate; see, e.g., [19,20,32]. Knowledge of the bound ε allows us to determine the regularization parameter μ by the discrepancy principle. Thus, we seek to compute a value $\mu > 0$ so that the solution $\boldsymbol{z}_{\ell,\mu}$ of (3.1) satisfies

$$\|\widetilde{H}_{\ell+1,\ell}\boldsymbol{z}_{\ell,\mu} - \|\boldsymbol{b}\|\boldsymbol{e}_1\| = \eta\varepsilon, \tag{3.3}$$

where $\eta \geq 1$ is a user-specified constant independent of ε. Denote this value of μ by μ_ℓ and the associated solution of (3.1) by $\boldsymbol{z}_{\ell,\mu_\ell}$. The following result shows that generally it is possible to determine μ_ℓ and $\boldsymbol{z}_{\ell,\mu_\ell}$ with the desired properties.

Proposition 3.1 *Let the matrix $\widetilde{H}_{\ell+1,\ell}$ be defined by (3.2) and assume that $\widetilde{H}^*_{\ell+1,\ell}\boldsymbol{e}_1 \neq \boldsymbol{0}$. Let $P_{\mathcal{N}(\widetilde{H}^*_{\ell+1,\ell})}$ denote the orthogonal projector onto $\mathcal{N}(\widetilde{H}^*_{\ell+1,\ell})$. Then the function*

$$\varphi(\nu) = \|\widetilde{H}_{\ell+1,\ell}\boldsymbol{z}_{\ell,1/\nu} - \|\boldsymbol{b}\|\,\boldsymbol{e}_1\|^2$$

is strictly decreasing, convex, and

$$\varphi(0) = \|\boldsymbol{b}\|^2, \qquad \lim_{\nu\to\infty} \varphi(\nu) = \|P_{\mathcal{N}(\widetilde{H}^*_{\ell+1,\ell})}\boldsymbol{e}_1\|^2 \|\boldsymbol{b}\|^2.$$

Proof The results can be shown by using the representation

$$\varphi(\nu) = \|\boldsymbol{b}\|^2 \boldsymbol{e}_1^*(\nu\widetilde{H}_{\ell+1,\ell}\widetilde{H}_{\ell+1,\ell}^* + I)^{-2}\boldsymbol{e}_1, \qquad \nu \geq 0$$

and the SVD of $\widetilde{H}_{\ell+1,\ell}$. An analogous result is established in [5, Theorem 2.1], where a detailed proof is provided. □

Typically,

$$\|P_{\mathcal{N}(\widetilde{H}_{\ell+1,\ell}^*)}\boldsymbol{e}_1\|\|\boldsymbol{b}\| < \eta\varepsilon < \|\boldsymbol{b}\|. \tag{3.4}$$

Then, by Proposition 3.1, the equation

$$\varphi(\nu) = \eta^2\varepsilon^2 \tag{3.5}$$

has a unique solution ν_ℓ, and μ_ℓ defined above is given by $\mu_\ell = 1/\nu_\ell$. The solution $\boldsymbol{y}_{\ell,\mu_\ell}$ of (1.10) yields the approximate solution $\boldsymbol{x}_{\ell,\mu_\ell} = V_\ell \boldsymbol{y}_{\ell,\mu_\ell}$ of (1.4) and (1.1). It follows that $\boldsymbol{x}_{\ell,\mu_\ell}$ satisfies

$$\|A\boldsymbol{x}_{\ell,\mu_\ell} - \boldsymbol{b}\| = \eta\varepsilon.$$

Many zero-finders can be used to solve (3.5). The monotonic decrease and convexity of $\varphi(\nu)$ secure that Newton's method with an initial iterate smaller than ν_ℓ, such as $\nu_\ell = 0$, converges monotonically and quadratically.

Let $\boldsymbol{x}_{\ell,\mu_\ell}$ solve (1.4) with $\mu = \mu_\ell$. Our computed examples indicate that it is important that $\|B\boldsymbol{x}_{\ell,\mu_\ell}\|$ be fairly small. We next show that this quantity is a decreasing function of ℓ.

Proposition 3.2 *Assume that (3.4) holds and let ℓ be large enough so that the equation*

$$\|A\boldsymbol{x} - \boldsymbol{b}\| = \eta\varepsilon \tag{3.6}$$

has the solutions $\boldsymbol{x}_{p,\mu_p} \in \mathcal{R}(V_p)$ for $p \in \{\ell, \ell+1\}$. Then

$$\|B\boldsymbol{x}_{\ell+1,\mu_{\ell+1}}\| \leq \|B\boldsymbol{x}_{\ell,\mu_\ell}\|. \tag{3.7}$$

Proof We have $\boldsymbol{x}_{p,\mu_p} = V_p\boldsymbol{y}_{p,\mu_p}$ for some $\boldsymbol{y}_{p,\mu_p} \in \mathbb{R}^p$ and $p \in \{\ell, \ell+1\}$. Then

$$\begin{aligned}
&\|AV_{\ell+1}\boldsymbol{y}_{\ell+1,\mu_{\ell+1}} - \boldsymbol{b}\|^2 + \mu_{\ell+1}\|BV_{\ell+1}\boldsymbol{y}_{\ell+1,\mu_{\ell+1}}\|^2 \\
&\quad = \min_{\boldsymbol{y}\in\mathbb{R}^{\ell+1}} \left\{ \|AV_{\ell+1}\boldsymbol{y} - \boldsymbol{b}\|^2 + \mu_{\ell+1}\|BV_{\ell+1}\boldsymbol{y}\|^2 \right\} \\
&\quad \leq \min_{\boldsymbol{y}\in\mathbb{R}^{\ell}} \left\{ \|AV_{\ell}\boldsymbol{y} - \boldsymbol{b}\|^2 + \mu_{\ell+1}\|BV_{\ell}\boldsymbol{y}\|^2 \right\} \\
&\quad \leq \|AV_\ell\boldsymbol{y}_{\ell,\mu_\ell} - \boldsymbol{b}\|^2 + \mu_{\ell+1}\|BV_\ell\boldsymbol{y}_{\ell,\mu_\ell}\|^2.
\end{aligned} \tag{3.8}$$

It follows from (3.6) that

$$\|AV_{\ell+1}\boldsymbol{y}_{\ell+1,\mu_{\ell+1}} - \boldsymbol{b}\|^2 = \|AV_\ell\boldsymbol{y}_{\ell,\mu_\ell} - \boldsymbol{b}\|^2 = \eta^2\varepsilon^2.$$

This property together with (3.8) shows (3.7). □

We have observed that the norm $\|B\boldsymbol{x}_{\ell,\mu_\ell}\|$ typically decreases faster during the initial steps of Algorithm 2.1 than during later steps. A possible stopping criterion for the algorithm therefore is to terminate the computations when $\|B\boldsymbol{x}_{\ell,\mu_\ell}\|$ stops decreasing rapidly. We illustrate this stopping criterion in Section 5.

4 Reduction of matrix tuplets and Tikhonov regularization

The reduction method of Section 2 can be generalized to a reduction method for matrix q-tuplets $\{A, B^{(1)}, B^{(2)}, \ldots, B^{(q-1)}\}$, where $A, B^{(i)} \in \mathbb{R}^{n\times n}$ for $1 \leq i < q$. An extension of Algorithm 2.1 that determines the decompositions

$$AV_\ell = U_{\ell+1} H_{\ell+1,\ell}, \tag{4.1}$$

$$B^{(i)} V_\ell = W_\ell^{(i)} R_\ell^{(i)}, \qquad i = 1, 2, \ldots, q-1, \tag{4.2}$$

can be developed, where the matrices $U_{\ell+1} \in \mathbb{R}^{n\times(\ell+1)}$ and $V_\ell, W_\ell^{(i)} \in \mathbb{R}^{n\times\ell}$ have orthonormal columns. The first column of $U_{\ell+1}$ is a specified unit vector such as (1.8). The columns of V_ℓ can be determined in a similar fashion as in Algorithm 2.1 or be chosen to be arbitrary linearly independent vectors. We note that the extension of Algorithm 2.1 requires user-chosen parameters ρ_i, $1 \leq i < q$, that are analogues of the parameter ρ of Algorithm 2.1. The matrix $H_{\ell+1,\ell} \in \mathbb{R}^{(\ell+1)\times\ell}$ is of upper Hessenberg form and the matrices $R_\ell^{(i)} \in \mathbb{R}^{\ell\times\ell}$ are upper triangular. The relation (4.1) is a flexible Arnoldi decomposition, and the relations (4.2) are QR factorizations. An analogue of Propositions 2.1 can be shown.

We turn to an application of the decompositions (4.1) and (4.2) to multi-parameter Tikhonov regularization

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \Big\{ \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \sum_{i=1}^{q-1} \mu_i \|B^{(i)}\boldsymbol{x}\|^2 \Big\}, \tag{4.3}$$

where the $B^{(i)}$ are regularization matrices and the scalars μ_i are regularization parameters. The parameter q is assumed to be larger than or equal to three. We require $\mu_i \geq 0$ for all i and $\sum_{i=1}^{q-1} \mu_i > 0$. If the regularization matrices $B^{(i)}$ satisfy

$$\mathcal{N}(A) \cap \mathcal{N}(B^{(i)}) = \{\boldsymbol{0}\}, \qquad 1 \leq i < q,$$

then (4.3) has a unique solution. This is a straightforward generalization of (1.5). If we instead require all μ_i to be positive, then it suffices that the regularization matrices $B^{(i)}$ satisfy

$$\mathcal{N}(A) \cap \mathcal{N}(B^{(1)}) \cap \cdots \cap \mathcal{N}(B^{(q-1)}) = \{\boldsymbol{0}\}$$

in order for (4.3) to have a unique solution.

An insightful discussion on multi-parameter Tikhonov regularization for small to moderately sized problems (4.3) is provided by Brezinski et al. [4]. More recent discussions are given by Gazzola and Novati [14,15], and Lu and Pereverzyev [27].

Letting $\boldsymbol{x} = V_\ell \boldsymbol{y}$ and applying the decompositions (4.1) and (4.2), we proceed similarly as in the derivation of (1.10) and obtain that the minimization

problem (4.3) restricted to the space $\mathcal{R}(V_\ell)$ is equivalent to the reduced minimization problem

$$\min_{\boldsymbol{y}\in\mathbb{R}^\ell}\Big\{\,\|H_{\ell+1,\ell}\,\boldsymbol{y}-\|\boldsymbol{b}\|\,\boldsymbol{e}_1\|^2+\sum_{i=1}^{q-1}\mu_i\,\|R_\ell^{(i)}\boldsymbol{y}\|^2\Big\}.$$

Methods for determining suitable regularization parameters for this minimization problem are described in [4,14,15,27].

5 Numerical examples

We present a few examples that illustrate the performance of one-parameter Tikhonov regularization based on Algorithm 2.1 for various values of the parameter ρ. We refer to this method as the flexible Arnoldi–Tikhonov regularization (FATR) method. Our implementation first evaluates one matrix-vector product with the matrix A before matrix-vector products with B are computed, even when $\rho < 1$ in Algorithm 2.1. The error vector $\boldsymbol{e}$ has normally distributed pseudorandom entries with mean zero in all examples. The vector is scaled to correspond to a chosen error level

$$\delta = \|\boldsymbol{e}\|/\|\hat{\boldsymbol{b}}\|,$$

where $\hat{\boldsymbol{b}}$ denotes the error-free right-hand side vector in (1.3). Thus, $\varepsilon = \delta\|\hat{\boldsymbol{b}}\|$ in (3.3). The regularization parameter μ is in all example determined with the aid of the discrepancy principle with $\eta = 1$ in (3.3). The first two examples are test problems from Regularization Tools [18], while the last three examples show applications to the restoration of images that have been contaminated by blur and noise. The regularization matrix in the last two examples is such that its A-weighted pseudoinverse is unattractive to use in computations.

We compare the quality of the computed approximations of the desired solution $\hat{\boldsymbol{x}}$ obtained with the FATR method of the present paper with that achieved with the generalized Krylov subspace method described in [33]. The latter method is referred to as generalized Arnoldi–Tikhonov regularization (GATR) in the examples below. Similarly as in [33], we determine regularization parameter for GATR with the discrepancy principle and use the user-chosen parameter $\eta = 1$. This parameter is analogous to η in (3.3).

Algorithm 2.1 does not apply reorthogonalization when determining the columns of the matrices $U_{\ell+1}$, V_ℓ, and W_ℓ. Due to finite-precision arithmetic, the computed columns therefore might not be orthogonal. We found that this does not affect the computed solutions significantly. Specifically, when reorthogonalization is employed, the computed approximate solutions are of the same quality as the approximate solutions presented in this section.

5.1 Integral equations in one space-dimension

The first two examples are small enough to allow the computation of the GSVD of the matrix pair $\{A, B\}$. We therefore are able to compare the quality of the computed approximations of $\hat{\boldsymbol{x}}$ achieved with FATR with those determined by GSVD. The regularization matrix in these examples is the tridiagonal square matrix

$$B = \begin{bmatrix} 0 & 0 & \dots & & 0 \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & \dots & & 0 & 0 \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

It is a scaled finite difference approximation of the second derivative operator in one space-dimension with vanishing first and last rows. Its properties are discussed in [9,34]. The relative error is computed according to $\|\boldsymbol{x}_{\ell,\mu_\ell} - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\|$, where $\hat{\boldsymbol{x}}$ is the minimal-norm solution of (1.3). The regularization parameter μ_ℓ is determined by the discrepancy principle, i.e., so that (3.3) holds. All computations were carried out in MATLAB with about 15 significant decimal digits.

Example 5.1. The Fredholm integral equation of the first kind,

$$\int_0^{\pi/2} \kappa(\sigma,\tau)x(\sigma)d\sigma = b(\tau), \qquad 0 \le \tau \le \pi, \tag{5.1}$$

with $\kappa(\sigma,\tau) = \exp(\sigma\cos(\tau))$, $b(\tau) = 2\sinh(\tau)/\tau$, and solution $x(\tau) = \sin(\tau)$ is discussed by Baart [1]. We use the MATLAB function baart from [18] to discretize (5.1) by a Galerkin method with 500 orthonormal box functions as test and trial functions. The function baart produces a nonsymmetric matrix $A \in \mathbb{R}^{500\times 500}$ and a scaled discrete approximation $\hat{\boldsymbol{x}} \in \mathbb{R}^{500}$ of $x(\tau)$, with which we compute the error-free right-hand side $\hat{\boldsymbol{b}} := A\hat{\boldsymbol{x}}$. The error vector $\boldsymbol{e} \in \mathbb{R}^{500}$ corresponds to the error level $\delta = 1\cdot 10^{-3}$. The right-hand side $\boldsymbol{b}$ in the system (1.1) is obtained from (1.2).

Method	ρ	ℓ	$\|\boldsymbol{x}_{\ell,\mu_\ell} - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\|$	$\|B\boldsymbol{x}_{\ell,\mu_\ell}\|$
FATR	∞	7	$2.76\cdot 10^{-2}$	$4.84\cdot 10^{-5}$
FATR	1	11	$1.20\cdot 10^{-2}$	$4.88\cdot 10^{-5}$
FATR	1/5	31	$8.18\cdot 10^{-3}$	$4.93\cdot 10^{-5}$
GSVD			$3.19\cdot 10^{-2}$	

Table 5.1 Example 5.1. Relative errors and $\|B\boldsymbol{x}_{\ell,\mu_\ell}\|$ in computed approximate solutions for noise level 10^{-3}. The parameter ℓ denotes the number of steps with FATR.

Figure 5.1(a) displays the approximate solution $\boldsymbol{x}_{31,\mu_{31}}$ (red solid curve) computed by the FATR method with $\rho = 1/5$ and the approximate solution

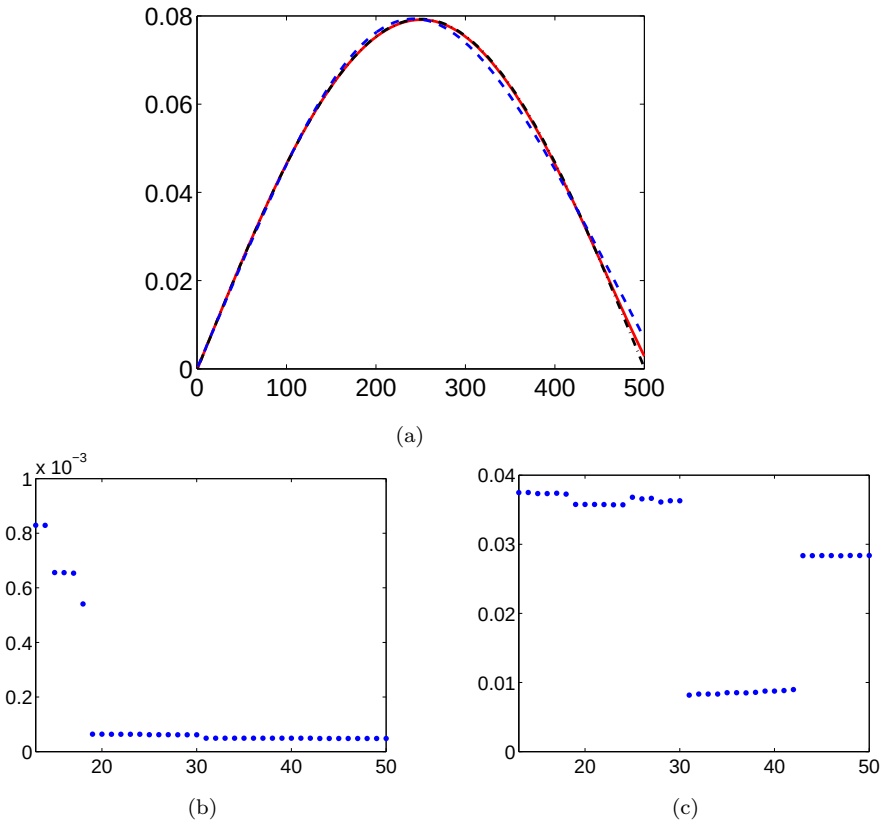


Fig. 5.1 Example 5.1. (a) Approximate solution $\boldsymbol{x}_{31,\mu_{31}}$ determined by FATR with $\rho = 1/5$ for the test problem baart with error level 10^{-3} (red solid curve), approximate solution determined with GSVD (blue dashed curve), and desired solution $\hat{\boldsymbol{x}}$ (black dash-dotted curve). (b) $\|B\boldsymbol{x}_{\ell,\mu_\ell}\|$ as a function of ℓ for the test problem baart with $\boldsymbol{x}_{\ell,\mu_\ell}$ computed by FATR with $\rho = 1/5$ for $\ell = 13, 14, ..., 50$. (c) $\|\boldsymbol{x}_{\ell,\mu_\ell} - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\|$ as a function of ℓ

$\boldsymbol{x}_{\mathrm{GSVD}}$ obtained with GSVD (blue dashed curve). The value $\rho = 1/5$ gives the most accurate approximation of the desired solution $\hat{\boldsymbol{x}}$ among the ρ-values considered; see Table 5.1. Figure 5.1(a) also shows $\hat{\boldsymbol{x}}$ (black dash-dotted curve). FATR can be seen to deliver a more accurate approximation of $\hat{\boldsymbol{x}}$ than GSVD. The reason for this is that the approximate solution $\boldsymbol{x}_{31,\mu_{31}}$ is restricted to a fairly low-dimensional subspace, while $\boldsymbol{x}_{\mathrm{GSVD}}$, the solution computed by GSVD, is not.

Figure 5.1(b) shows the values of $\|B\boldsymbol{x}_{\ell,\mu_\ell}\|$ as a function of the number of steps ℓ for $\rho = 1/5$. The values of $\|B\boldsymbol{x}_{\ell,\mu_\ell}\|$ for $\ell < 13$ are very large and are therefore not included in the graph. The norm $\|B\boldsymbol{x}_{\ell,\mu_\ell}\|$ does not change much with ℓ for $\ell \geq 31$. Generally, this indicates that an accurate approximation of $\hat{\boldsymbol{x}}$ has been found. This is the case for the present example. Figure 5.1(c) shows

the relative errors $\|\boldsymbol{x}_{\ell,\mu_\ell} - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\|$ as a function of ℓ. We observe a drop in the relative error at $\ell = 31$ and a jump at $\ell = 43$. The jump stems from the fact that the solution $\boldsymbol{x}_{\ell,\mu_\ell}$ has converged to the solution $\boldsymbol{x}_{\mathrm{GSVD}}$ obtained with GSVD when $\ell \geq 43$. For many linear discrete ill-posed problems, FATR determines more accurate approximations of $\hat{\boldsymbol{x}}$ than $\boldsymbol{x}_{\mathrm{GSVD}}$ before converging to the latter. We remark that when ρ is changed, Figures 5.1(b) and (c) also change, but they still display jumps analogous to the ones in the displayed figures.

Table 5.1 shows numerical results. The choice of the ratio ρ is seen to be important for the accuracy. The notation $\rho = \infty$ stands for a value of ρ large enough so that the solution subspace is a standard Krylov subspace determined by the matrix A. This choice of ρ yields a worse approximations of $\hat{\boldsymbol{x}}$ than smaller values of ρ. We recall that $\rho = 1$ and $\rho = 1/5$ give solution subspaces that are generalized Krylov subspaces; cf. Proposition 2.2. The number of steps, ℓ, of FATR is chosen to yield approximate solutions $\boldsymbol{x}_{\ell,\mu_\ell}$ of highest accuracy for a fixed ρ.

This example shows that FATR can give better approximations of $\hat{\boldsymbol{x}}$ than Tikhonov regularization based on the GSVD of $\{A, B\}$. Moreover, the error $\|\boldsymbol{x}_{\ell,\mu_\ell} - \hat{\boldsymbol{x}}\|$ generally does not change much with ℓ for ℓ sufficiently large, but not too large. It is, therefore, not crucial for the good performance of FATR that the number of iterations ℓ that yields the smallest error be known. □

Example 5.2. Consider the Fredholm integral equation of the first kind

$$\int_0^1 k(s,t)x(t)dt = \exp(s) + (1-e)s - 1, \qquad 0 \leq s \leq 1, \tag{5.2}$$

where

$$k(s,t) = \begin{cases} s(t-1), & s < t, \\ t(s-1), & s \geq t. \end{cases}$$

We discretize the integral equation by a Galerkin method with orthonormal box functions as test and trial functions using the MATLAB function deriv2 from [18]. This yields the matrix $A \in \mathbb{R}^{500\times 500}$. The function deriv2 also produces a scaled discrete approximation $\hat{\boldsymbol{x}} \in \mathbb{R}^{500}$ of the solution $x(t) = \exp(t)$ of (5.2). The error vector $\boldsymbol{e} \in \mathbb{R}^{500}$ corresponds to the error level 10^{-3}.

Method	ρ	ℓ	$\|\boldsymbol{x}_{\ell,\mu_\ell} - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\|$	$\|B\boldsymbol{x}_{\ell,\mu_\ell}\|$
FATR, no augm.	∞	11	$1.62 \cdot 10^{-1}$	$3.79 \cdot 10^{-3}$
FATR, no augm.	1	9	$1.78 \cdot 10^{-3}$	$5.88 \cdot 10^{-6}$
FATR, no augm.	1/5	19	$4.95 \cdot 10^{-4}$	$7.15 \cdot 10^{-6}$
FATR, augm. by (5.3)	∞	3	$5.69 \cdot 10^{-4}$	$3.62 \cdot 10^{-4}$
GATR, augm. by (5.3)		20	$5.08 \cdot 10^{-3}$	
GSVD			$2.70 \cdot 10^{-3}$	

Table 5.2 Example 5.2. Relative errors and $\|B\boldsymbol{x}_{\ell,\mu_\ell}\|$ in computed approximate solutions for the test problem deriv2 with noise level 10^{-3}.

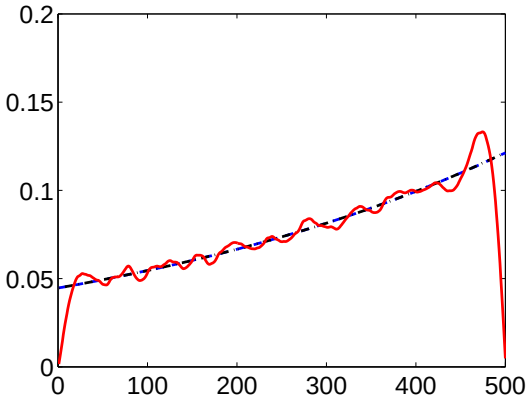


Fig. 5.2 Example 5.2. Computed approximate solutions for the test problem deriv2 with error level 10^{-3}. The red solid curve shows the approximate solution $x_{11,\mu_{11}}$ determined by FATR without augmentation and $\rho = \infty$, the blue dashed curve displays $x_{19,\mu_{19}}$ computed by FATR without augmentation and $\rho = 1/5$, and the black dash-dotted curve shows the $\hat{\boldsymbol{x}}$.

It is well known that augmentation of Krylov or generalized Krylov subspaces by suitable vectors can enhance the quality of the computed approximate solutions; see, e.g., [2,29,33] for illustrations. For the present example, it is beneficial to include an orthonormal basis for the span of the vectors

$$\boldsymbol{n}_1 = [1, 1, \ldots, 1]^T, \qquad \boldsymbol{n}_2 = [1, 2, \ldots, n]^T \tag{5.3}$$

in the solution subspace. However, for many linear discrete ill-posed problems it is not clear which vectors should be included in the solution subspace to improve the quality of the computed solution. It is therefore desirable that solution methods perform well also in this situation. The present example compares the performance of FATR with and without the inclusion of the span of the vectors (5.3) in the solution subspace. It is easy to adjust Algorithm 2.1 to include the span of the vectors (5.3) in the solution subspace. We first determine an orthonormal basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ of span$\{\boldsymbol{n}_1, \boldsymbol{n}_2\}$, and then skip the last assignment of line 2, as well as lines 17-26 for $j = 1, 2$.

Figure 5.2 shows the approximate solution $x_{11,\mu_{11}}$ (red solid curve) computed by FATR with $\rho = \infty$ and without augmentation (i.e., without inclusion of the vectors (5.3) in the solution subspace), the approximate solution $x_{19,\mu_{19}}$ (blue dashed curve) computed by FATR with $\rho = 1/5$ and without augmentation, and the desired solution $\hat{\boldsymbol{x}}$ (black dash-dotted curve). The relative errors and $\|B\boldsymbol{x}_{\ell,\mu_\ell}\|$ are reported in Table 5.2. There we also show results obtained by FATR and GATR when the solution subspace is augmented by an orthonormal basis for span$\{\boldsymbol{n}_1, \boldsymbol{n}_2\}$. The table shows that FATR for $\rho = 1/5$ and without augmentation yields a more accurate approximation of $\hat{\boldsymbol{x}}$ than the augmented FATR method with $\rho = \infty$. This is important, because it can be difficult to choose suitable vectors for augmentation. FATR with $\rho = 1$ and $\rho = 1/5$ also

is seen to yield better approximations of $\hat{\boldsymbol{x}}$ than GSVD and the GATR method with solution subspace augmented by the span of the vectors (5.3). □

5.2 Applications to image restoration

The following three examples illustrate the performance of the FATR method when applied to the restoration of images.

Example 5.3. We consider the restoration of a blurred gray-scale image contaminated by noise. The gray-scale image is represented by an array of 412×412 pixels, with each pixel stored as an 8-bit unsigned integer with a value in the interval $[0, 255]$. The pixels are ordered row-wise and stored in a vector of dimension $n = 412^2$. Let $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ represent the original uncontaminated image shown in Figure 5.3(a). This image has fine details and well-defined edges. We generate an associated blurred and noise-free image $\hat{\boldsymbol{b}}$ by multiplying $\hat{\boldsymbol{x}}$ by a Gaussian blurring matrix $A \in \mathbb{R}^{n\times n}$. This matrix is generated with the function blur from [18] with the parameters band $= 9$ and sigma $= 3$. The blur- and noise-contaminated image $\boldsymbol{b} \in \mathbb{R}^n$ is obtained by adding a vector $\boldsymbol{e} \in \mathbb{R}^n$, that represents 30% additive Gaussian noise, to $\hat{\boldsymbol{b}}$; cf. (1.2). This yields the blur- and noise-contaminated image displayed in Figure 5.3(b).

We assume the blurring matrix A, the contaminated image $\boldsymbol{b} \in \mathbb{R}^n$, and the noise level δ to be available, and we would like to determine a restoration that accurately approximates the blur- and noise-free image $\hat{\boldsymbol{x}}$, which is assumed not to be known. The peak signal-to-noise ratio (PSNR) is a commonly used measure of the quality of a restored image $\boldsymbol{x}_{\ell,\mu_\ell}$. It is defined as

$$\mathrm{PSNR}(\boldsymbol{x}_{\ell,\mu_\ell}, \hat{\boldsymbol{x}}) = 20\log_{10}\left(\frac{255}{\|\boldsymbol{x}_{\ell,\mu_\ell} - \hat{\boldsymbol{x}}\|}\right).$$

A larger PSNR-value indicates that the restoration $\boldsymbol{x}_{\ell,\mu_\ell}$ is of higher quality, however, in some cases this might not agree with visual judgment. We therefore also display the restored images.

We use two regularization matrices: the standard 5-point discretization of the Laplace operator,

$$B(x) = \Delta x, \tag{5.4}$$

using a five-point stencil, and a discretization of the Perona–Malik operator

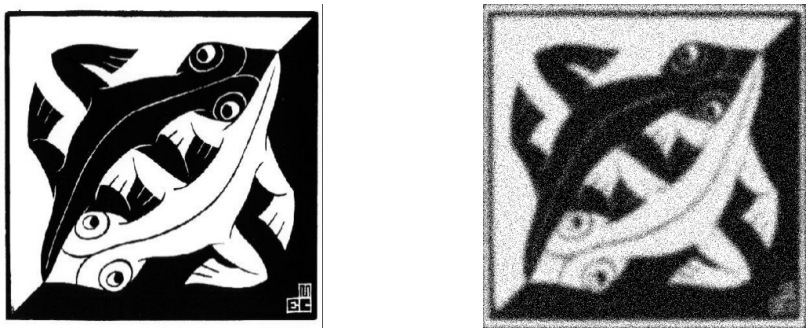
$$B(x) = \mathrm{div}(g(|\nabla x|^2)\,\nabla x), \tag{5.5}$$

where ∇x denotes the gradient of x considered as a real-valued function defined in $\mathbb{R}^2$. The diffusivity is given by

$$g(s) = \frac{\rho}{\rho + s}$$

with $\rho > 0$ a small positive constant; see [31] for discussions on this choice of diffusivity. We use the value $\rho = 10^{-5}$. The operator (5.5) is discretized by finite differences. This gives a matrix B with, generically, five nonvanishing

entries in each row. The entries in the row associated with pixel (i, j) are determined by the values of the image $\boldsymbol{x}$ at pixel (i, j) and at the four adjacent pixels in the horizontal and vertical directions; see [33] for details. Further discussions on this and alternative discretizations can be found in [16,36].



(a) (b)

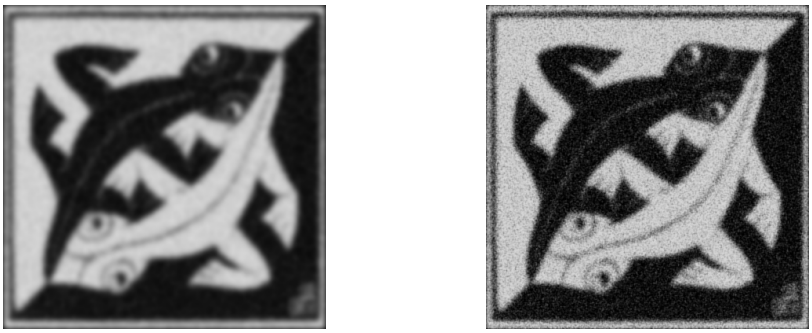
Fig. 5.3 Example 5.3. (a) Original image, (b) blur- and noise-contaminated image.

ρ	ℓ	PSNR	Figure
∞	2	12.11	Figure 5.4 (a)
1	2	12.47	Figure 5.4 (b)
1/2	3	12.55	Figure 5.4 (c)

Table 5.3 Example 5.3. PSNR-values of restorations determined by FATR with the discretized Laplacian as regularization matrix.

Figure 5.4 shows the restored images determined by FATR with the regularization matrix B chosen to be the standard five-point discretization of the Laplace operator. The PSNR-values for the restorations are listed in Table 5.3. When $\rho = \infty$, the best restoration is obtained after $\ell = 2$ steps; the restored image is shown in Figure 5.4(a), which is the most blurry restoration of the three images of Figure 5.4. It has PSNR=12.11. Letting $\rho = 1$, we again obtain the best restoration after only $\ell = 2$ steps. This restoration is displayed in Figure 5.4(b); its PSNR-value is 12.47. This image has less blurred edges than the image of Figure 5.4(a). Decreasing ρ to 1/2, we obtain, after three steps, the restoration with the largest PSNR-value: 12.55. It is depicted in Figure 5.4(c).

Figure 5.5(a) shows the restored image $\boldsymbol{x}_{4,\mu_4}$ determined by the FATR method with the discretized Perona–Malik (PM) regularization operator (5.5)



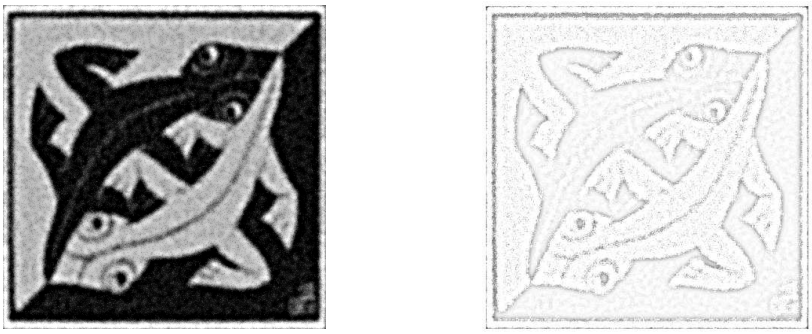
(a) (b)

(c)

Fig. 5.4 Example 5.3. Images restored by the FATR method with the discretized Laplacian as regularization matrix. (a) Restored image $\boldsymbol{x}_{2,\mu_2}$ obtained with $\rho = \infty$, (b) restored image $\boldsymbol{x}_{2,\mu_2}$ determined with $\rho = 1$, and (c) restored image $\boldsymbol{x}_{3,\mu_3}$ obtained with $\rho = 1/2$.

as regularization matrix, and $\rho = \infty$. The PM operator is linearized by first carrying out two steps of the range restricted GMRES method [30]. This yields a better approximation $\boldsymbol{x}_2$ of $\hat{\boldsymbol{x}}$ than $\boldsymbol{b}$. We use $\boldsymbol{x}_2$ to determine the discretized PM operator, which then is kept fixed during subsequent iterations. The PSNR-value for $\boldsymbol{x}_{4,\mu_4}$ is 13.53. This value is larger than the PSNR-values for the restorations computed with the discretized Laplace operator. The restored image is seen to have sharp edges. Figure 5.5(b) shows the edge map for the restoration of Figure 5.5(a). It is determined with the edge detector of gimp, a public domain software tool for image processing. The edge map is seen to be accurate despite the severe noise and blur in the image of Figure 5.3(b) that we restored.

Thus, FATR with $\rho = 1/2$ gives the best restoration when the regularization matrix is a discrete Laplacian, while $\rho = \infty$ yields the best restoration



(a) (b)

Fig. 5.5 Example 5.3. (a) Restored image $\boldsymbol{x}_{4,\mu_4}$ by FATR with PM operator (5.5) with $\rho = \infty$, (b) edges map for the restoration in (a).

when the regularization matrix is a discretization of the PM operator. This illustrates the benefit of a parameter ρ that determines the construction of the generalized Krylov subspace used. □

Example 5.4. This example discusses the solution of a large-scale problem that can be simplified by using its Kronecker product structure. The blurring matrix A in Example 5.3 is the Kronecker product of a matrix $M \in \mathbb{R}^{412\times 412}$ with itself. Let $L \in \mathbb{R}^{412\times 412}$ be the symmetric tridiagonal matrix that represents the second difference operator; thus a typical row of L has the nontrivial entries $\{-1, 2, -1\}$.

Consider the Tikhonov minimization problem

$$\min_{X\in\mathbb{R}^{412\times 412}} \{\|MXM - G\|_F^2 + \mu\|LXL\|_F^2\}, \tag{5.6}$$

where $G \in \mathbb{R}^{412\times 412}$ represents the available blur- and noise-contaminated image shown in Figure 5.3(b). Thus, the image represented by G is contaminated by 30% Gaussian noise and by the same blur as the contaminated image of Example 5.3. The norm $\|\cdot\|_F$ in (5.6) is the Frobenius norm.

Problem (5.6) can be expressed as

$$\min_{\boldsymbol{x}\in\mathbb{R}^{412^2}} \Big\{\|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \mu\|B\boldsymbol{x}\|^2\Big\}, \tag{5.7}$$

where $\boldsymbol{b}$ and $\boldsymbol{x}$ are vector representations of the matrices G and X, respectively, and $B = L \otimes L$. The Kronecker product structure of the matrices A and B makes it possible to solve the large problem (5.7) with the aid of the GSVD. Specifically, one can simplify the problem by computing the GSVD of the matrix pair $\{M, L\}$. Details of this approach are described by Bouhamidi and Jbilou [3]. Our interest in this problem stems from that it allows us to compare the solutions determined by GSVD and the FATR method.

Fig. 5.6 Example 5.4. Image $\boldsymbol{x}_{5,\mu_5}$ restored by the FATR method with a Kronecker product regularization matrix and $\rho = 1/3$.

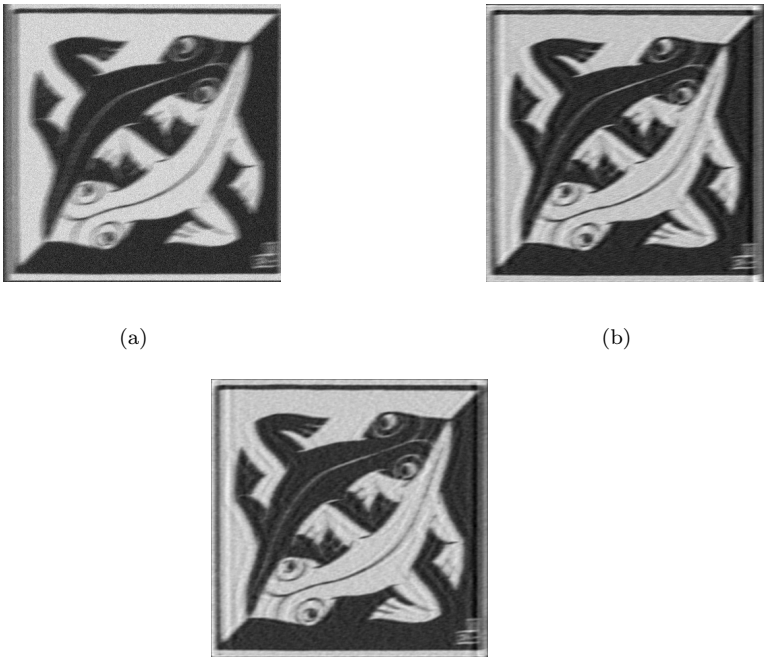
Method	ρ	ℓ	PSNR	Figure
FATR	1	5	16.90	
FATR	1	28	17.05	
FATR	1/3	5	16.91	Figure 5.6
FATR	1/3	50	16.99	
GSVD			16.88	

Table 5.4 Example 5.4. PSNR-values of restorations determined by FATR with the the Kronecker product of second order finite difference operators as regularization operator.

Table 5.4 shows FATR to yield restorations of about the same quality after 5 steps for both $\rho = 1$ and $\rho = 1/3$. The PSNR-value for the restorations increases with the number of steps. All restorations achieved by FATR have a larger PSNR-value than the GSVD restoration. Thus, also for this large-scale example, FATR is competitive with GSVD with respect to the quality of the restoration. Moreover, matrix-vector products with A and B can be evaluated efficiently by expoliting the Kronecker product structure. We remark that the regularization matrix B of the present example yields more accurate restorations than the regularization matrices of Example 5.3. This depends on the image to be restored and does not hold for all image restoration problems. □

Example 5.5. We consider the restoration of an image that has been contaminated by linear motion blur and noise. The point spread function (PSF) for motion blur is represented by a line segment of length r pixels in the direction of the motion. The angle θ (in degrees) specifies the direction; it is measured counter-clockwise from the positive horizontal axis. The PSF takes on the value r^{-1} on this segment and vanishes elsewhere. The PSF defines the blurring matrix A. The larger r, the more ill-conditioned is A, and the more difficult is the restoration task. In this example, we let $r = 15$ and $\theta = 10$, and use the same test image as in Example 5.4. Thus, the original blur- and noise-free image, shown in Figure 5.3(a), is represented by 412×412 pixels.

This image is contaminated by motion blur as well as by 10% Gaussian noise. Figure 5.7(a) displays the contaminated image, which is represented by the vector $\boldsymbol{b} \in \mathbb{R}^{412^2}$. We let the regularization matrix B be the standard 5-point discretization of the Laplace operator (5.4).



(a) (b)

(c)

Fig. 5.7 Example 5.5. (a) Image contaminated by motion blur and 10% Gaussian noise. (b) Restored image $\boldsymbol{x}_{21,\mu_{21}}$. (c) Restored image $\boldsymbol{x}_{100,\mu_{100}}$.

When applying Algorithm 2.1 to this restoration problem, the algorithm failed to determine a restoration that satisfies the discrepancy principle for several choices of the parameter ρ. A reason for this is that the solution subspaces generated are not suitable for this problem, because each matrix-vector product evaluation with A applies more motion blur in the direction θ. A remedy is to also evaluate matrix-vector products with the matrix A^* when defining the solution subspace. This matrix represents motion blur in the direction $-\theta$. This modification of Algorithm 2.1 easily can be implemented. Thus, after each inclusion of a vector $\boldsymbol{u}_{N(\boldsymbol{u})+1}$ in the solution subspace in line 18 of the algorithm, we also include the direction $A^*\boldsymbol{u}_{N(\boldsymbol{u})+1}$. Letting $\rho = 1$, the discrepancy principle can be satisfied after 21 steps of the algorithm. The solution subspace then is of dimension 21. Figure 5.7(b) displays the restored

image obtained. It has PSNR 15.75. The PSNR value of the computed restorations increases when carrying out more steps. For instance, after 100 steps, we obtain a restoration with PSNR 17.21. It is shown in Figure 5.7(c).

This example illustrates the ease of adjusting Algorithm 2.1 so that a suitable solution subspace is determined. Though, the ability to make a successful adjustment typically requires some insight into properties of the problem to be solved and the solution subspace generated. We remark that B^* can be introduced in Algorithm 2.1 in a similar way as A^*. This allows the solution of problems with a rectangular regularization matrix B. An illustration can be found in [37]. □

6 Conclusion

A new reduction method for large-scale Tikhonov regularization problems (1.4) is presented. It only requires the evaluation of matrix-vector products with the square matrices A and B in (1.4). Therefore, these matrices do not have to have particular structure. Numerical examples illustrate that the reduction method can determine approximate solutions of higher quality than Tikhonov regularization based on determining the GSVD of the matrix pair $\{A, B\}$ for problems that are small enough to allow the computation of the GSVD.

Acknowledgments. We would like to thank the referees for comments that lead to improvements of the presentation. This research is supported in part by NSF grant DMS-1115385.

References

1. M. L. Baart, *The use of auto-correlation for pseudo-rank determination in noisy ill-conditioned least-squares problems*, IMA J. Numer. Anal., 2 (1982), pp. 241–247.
2. J. Baglama and L. Reichel, *Decomposition methods for large linear discrete ill-posed problems*, J. Comput. Appl. Math., 198 (2007), pp. 332–342.
3. A. Bouhamidi and K. Jbilou, *A Sylvester–Tikhonov regularization method for image restoration*, J. Comput. Appl. Math., 206 (2007), pp. 86–98.
4. C. Brezinski, M. Redivo–Zaglia, G. Rodriguez, and S. Seatzu, *Multi-parameter regularization techniques for ill-conditioned linear systems*, Numer. Math., 94 (2003), pp. 203–224.
5. D. Calvetti and L. Reichel, *Tikhonov regularization of large linear problems*, BIT, 43 (2003), pp. 263–283.
6. T. F. Chan and K. R. Jackson, *Nonlinearly preconditioned Krylov subspace methods for discrete Newton algorithms*, SIAM J. Sci. Statist. Comput., 5 (1984), pp. 533–542.
7. J. W. Daniel, W. B. Gragg, L. Kaufman, and G. W. Stewart, *Reorthogonalization and stable algorithms for updating the Gram–Schmidt QR factorization*, Math. Comp., 30 (1976), pp. 772–795.
8. M. Donatelli, A. Neuman, and L. Reichel, *Square regularization matrices for large linear discrete ill-posed problems*, Numer. Linear Algebra Appl., 19 (2012), pp. 896–913.
9. M. Donatelli and L. Reichel, *Square smoothing regularization matrices with accurate boundary conditions*, J. Comput. Appl. Math., 272 (2014), pp. 334–349.
10. L. Dykes and L. Reichel, *On the reduction of Tikhonov minimization problems and the construction of regularization matrices*, Numer. Algorithms, 60 (2012), pp. 683–696.

11. L. Eldén, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT, 17 (1977), pp. 134–145.
12. L. Eldén, *A weighted pseudoinverse, generalized singular values, and constrained least squares problems*, BIT, 22 (1982), pp. 487–501.
13. H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
14. S. Gazzola, *Regularization techniques based on Krylov subspace methods for ill-posed linear systems*, Ph.D. thesis, Department of Mathematics, University of Padova, Padova, Italy, 2013.
15. S. Gazzola and P. Novati, *Muli-parameter Arnoldi–Tikhonov methods*, Electron. Trans. Numer. Anal., 40 (2013), pp. 452–475.
16. A. Handlovicová, K. Mikula, and F. Sgallari, *Semi-implicit complementary volume scheme for solving level set like equations in image processing and curve evolution*, Numer. Math., 93 (2003), pp. 675–695.
17. P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.
18. P. C. Hansen, *Regularization tools version 4.0 for Matlab 7.3*, Numer. Algorithms, 46 (2007), pp. 189–194.
19. T. A. Hearn and L. Reichel, *Application of denoising methods to regularization of ill-posed problems*, Numer. Algorithms, 66 (2014), pp. 761–777.
20. I. Hnětynková, M. Plešinger, and Z. Strakoš, *The regularizing effect of the Golub–Kahan iterative bidiagonalization and revealing the noise level in the data*, BIT Numer. Math., 49 (2009), pp. 669–696.
21. M. E. Hochstenbach and L. Reichel, *An iterative method for Tikhonov regularization with a general linear regularization operator*, J. Integral Equations Appl., 22 (2010), pp. 463–480.
22. M. E. Hochstenbach, L. Reichel, and X. Yu, *A Golub–Kahan-type reduction method for matrix pairs*, submitted for publication.
23. G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins University Press, Baltimore, 2013.
24. M. E. Kilmer, P. C. Hansen, and M. I. Español, *A projection-based approach to general-form Tikhonov regularization*, SIAM J. Sci. Comput., 29 (2007), pp. 315–330.
25. S. Kindermann, *Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems*, Electron. Trans. Numer. Anal., 38 (2011), pp. 233–257.
26. R.-C. Li and Q. Ye, *A Krylov subspace method for quadratic matrix polynomials with application to constrained least squares problems*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 405–428.
27. S. Lu and S. V. Pereverzyev, *Multi-parameter regularization and its numerical realization*, Numer. Math., 118 (2011), pp. 1–31.
28. S. Morigi, L. Reichel, and F. Sgallari, *Orthogonal projection regularization operators*, Numer. Algorithms, 44 (2007), pp. 99–114.
29. K. Morikuni, L. Reichel, and K. Hayami, *FGMRES for linear discrete ill-posed problems*, Appl. Numer. Math., 75 (2014), pp. 175–187.
30. A. Neuman, L. Reichel, and H. Sadok, *Implementations of range restricted iterative methods for linear discrete ill-posed problems*, Linear Algebra Appl., 436 (2012), pp. 3974–3990.
31. P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Analysis and Machine Intelligence, 12 (1990), pp. 629–639.
32. L. Reichel and G. Rodriguez, *Old and new parameter choice rules for discrete ill-posed problems*, Numer. Algorithms, 63 (2013), pp. 65–87.
33. L. Reichel, F. Sgallari, and Q. Ye, *Tikhonov regularization based on generalized Krylov subspace methods*, Appl. Numer. Math., 62 (2012), pp. 1215–1228.
34. L. Reichel and Q. Ye, *Simple square smoothing regularization operators*, Electron. Trans. Numer. Anal., 33 (2009), pp. 63–83.
35. Y. Saad, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Comput., 14 (1993), pp. 461–469.
36. J. Weickert, B. M. H. Romeny, and M. A. Viergever, *Efficient and reliable schemes for nonlinear diffusion filtering*, IEEE Trans. Image Processing, 7 (1998), pp. 398–410.

37. X. Yu, *Generalized Krylov subspace methods with applications*, Ph.D. thesis, Department of Mathematics, Kent State University, May 2014.
38. H. Zha, *Computing the generalized singular values/vectors of large sparse or structured matrix pairs*, Numer. Math., 72 (1996), pp. 391–417.