Tikhonov regularization based on generalized Krylov subspace methods

Lothar Reichel [a], Fiorella Sgallari [b,1], Qiang Ye [c,2]

[a] *Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. E-mail:* `reichel@math.kent.edu`.

[b] *Department of Mathematics-CIRAM, University of Bologna, Via Saragozza 8, 40123 Bologna, Italy. E-mail:* `sgallari@dm.unibo.it`.

[c] *Department of Mathematics, University of Kentucky, Lexington, KY 40506, USA. E-mail:* `qye@ms.uky.edu`.

Abstract

We consider Tikhonov regularization of large linear discrete ill-posed problems with a regularization operator of general form and present an iterative scheme based on a generalized Krylov subspace method. This method simultaneously reduces both the matrix of the linear discrete ill-posed problem and the regularization operator. The reduced problem so obtained may be solved, e.g., with the aid of the singular value decomposition. Also, Tikhonov regularization with several regularization operators is discussed.

Key words: ill-posed problem, regularization operator, Tikhonov regularization, multiparameter regularization

1 Introduction

This paper considers the solution of linear systems of equations

$$A\boldsymbol{x} = \boldsymbol{b}, \qquad A \in \mathbb{R}^{n\times n}, \qquad \boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n, \tag{1}$$

with a large matrix of ill-determined rank. The singular values of A "cluster" at the origin. It follows that A is severely ill-conditioned and may be singular.

[1] Work partially supported by PRIN, grant 20083KLJEZ.
[2] Supported in part by NSF under grant DMS-0915062.

Linear systems of equations with a matrix of ill-determined rank frequently are referred to as linear discrete ill-posed problems. They arise, 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 of linear discrete ill-posed problems that arise in applications in science and engineering represents the available data and typically is contaminated by a measurement error $\boldsymbol{e} \in \mathbb{R}^n$, which we will refer to as "noise." Let $\hat{\boldsymbol{b}}$ denote the (unknown) noise-free vector associated with $\boldsymbol{b}$, i.e.,

$$\boldsymbol{b} = \hat{\boldsymbol{b}} + \boldsymbol{e}, \tag{2}$$

and assume that the unavailable noise-free system,

$$A\boldsymbol{x} = \hat{\boldsymbol{b}}, \tag{3}$$

is consistent. Let $\hat{\boldsymbol{x}}$ denote the solution of (3) of minimal Euclidean norm. We would like to determine an approximation of $\hat{\boldsymbol{x}}$ by computing a suitable approximate solution of (1).

Straightforward solution of (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 replaces (1) by a related problem, that is less sensitive to the error $\boldsymbol{e}$, and solves the latter. This replacement commonly is referred to as regularization. The possibly most popular regularization method, known as Tikhonov regularization, replaces (1) by the penalized least-squares problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \{\|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \mu\|L\boldsymbol{x}\|^2\}, \tag{4}$$

where the matrix $L \in \mathbb{R}^{p\times n}$ is a regularization operator, with $p \leq n$, and the scalar $\mu > 0$ is a regularization parameter. Throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm or the associated induced matrix norm. When L is the identity matrix, the Tikhonov minimization problem (4) is said to be in standard form. We assume that L is chosen so that

$$\mathcal{N}(A) \cap \mathcal{N}(L) = \{\boldsymbol{0}\},$$

where $\mathcal{N}(M)$ denotes the null space of the matrix M. Let M^* denote the adjoint of the matrix M. The normal equations associated with (4) are given by

$$(A^*A + \mu L^*L)\boldsymbol{x} = A^*\boldsymbol{b}, \tag{5}$$

which shows that (4) has the unique solution

$$\boldsymbol{x}_\mu = (A^*A + \mu L^*L)^{-1}A^*\boldsymbol{b} \tag{6}$$

for any $\mu > 0$. 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. [10], Groetsch [11], and Hansen [13] for discussions on Tikhonov regularization.

We would like to determine a suitable value of the regularization parameter $\mu > 0$ and an approximation of the associated solution $\boldsymbol{x}_\mu$ of (4). When the matrices A and L are of small to moderate size, this can be achieved with the aid of the Generalized Singular Value Decomposition (GSVD) of the matrix pair $\{A, L\}$; see [13]. In this paper, we are concerned with the situation when the matrices A and L are too large to apply the GSVD.

The Tikhonov minimization problem (4) can be transformed to standard form by the substitution $\boldsymbol{y} = L\boldsymbol{x}$. The matrix A in (4) then is replaced by $AL_A^\dagger$, where

$$L_A^\dagger = \left(I - (A(I - L^\dagger L))^\dagger A\right) L^\dagger \in \mathbb{R}^{n\times p}$$

is the A-weighted generalized inverse of L. Here $L^\dagger$ denotes the Moore-Penrose pseudoinverse of L, and I is the identity matrix; see e.g., Eldén [9] or Hansen [13, Section 2.3] for details. When A stems from the discretization of an ill-posed problem in one space-dimension, say, a Fredholm integral equation of the first kind on an interval, it is often possible to choose a regularization operator L with a special structure that allows efficient evaluation of matrix-vector products with $L_A^\dagger$ and $AL_A^\dagger$. In this situation it can be attractive to transform the minimization problem (4) to standard form, because this transformation simplifies subsequent computations, such as the determination of the regularization parameter; see, e.g., [4,5,21]. Examples of regularization operators L that allow efficient evaluation of matrix-vector products with $L_A^\dagger$ and $AL_A^\dagger$ include banded matrices, circulant matrices, orthogonal projections, and sparse nonsingular matrices that allow fast LU-factorization and fast forward- and back-substitution; see [6,9,13,18,21] for illustrations.

The present paper addresses the situation when L does not have a structure that allows efficient evaluation of matrix-vector products with $L_A^\dagger$ and $AL_A^\dagger$. This situation may arise, for instance, when A stems from the discretization of a Fredholm integral equation of the first kind in two or more space-dimensions. We will explore the application of a generalized Arnoldi process for the simultaneous orthogonal reduction of matrix pairs $\{A, L\}$ to pairs of smaller matrices presented in [17] for the solution of quadratic eigenvalue problems. The simultaneous orthogonal reduction of A and L to smaller matrices allows the

Tikhonov minimization problem (4) to be replaced by an analogous problem of smaller size, which can be solved conveniently by application of the GSVD or by some other methods. The matrices A and L do not have to be formed; only matrix-vector products with these matrices are needed. In particular, L is not required to possess any particular structure, except for being square. The latter requirement can be satisfied by zero-padding if necessary; see Section 2. The adjoints of A and L are not used. This can be advantageous in applications, where matrix-vector products with A are evaluated efficiently without forming the entries of the matrix. This situation arises, for instance, when matrix-vector products are computed by the fast multipole method or when (4) stems from the linearization of a nonlinear problem and A represents the Jacobian; see, e.g., [7] for a discussion on the evaluation of matrix-vector products with Jacobian matrices.

There are few efficient methods available for the solution of large-scale Tikhonov minimization problems (4) with a general regularization operator. While the computation of the solution (6) by the conjugate gradient or preconditioned conjugate gradient methods applied to the normal equations (5) is fairly straightforward, this approach often is computationally demanding. The reason for this is that, in general, a suitable value of the regularization parameter μ is not known a priori. Therefore, several systems (5) with different values of μ have to be solved. Recently, Kilmer et al. [15] described an inner-outer iterative method for the solution of (4) with a general matrix L. However, this method is not always cheap, since a fairly large number of inner iterations may be necessary. Moreover, matrix-vector product evaluations with the adjoints of A and L are required. We therefore investigate an alternative approach.

This paper is organized as follows. Section 2 reviews the generalized Arnoldi process described in [17] and extends it to matrix m-tuplets. Applications to the Tikhonov minimization problem (4) are discussed in Section 3. In some applications the null space of L is important. However, the reduction of L computed by the method of Section 3 generally is of full rank. This issue is addressed in Section 4. Computed examples can be found in Section 5, and Section 6 contains a few concluding remarks.

2 Generalized Arnoldi processes

The beginning of this section reviews the generalized Arnoldi process for matrix pairs described in [17]. It will be applied to the solution of (4). At the end of the section, we describe a generalization to matrix m-tuplets. It can be applied to multiparameter Tikhonov regularization problems.

We first discuss the generalized Arnoldi process for two square matrices A and

L. This process is based on the observation that for an arbitrary unit vector $\boldsymbol{q}_1 \in \mathbb{R}^n$, generically, there is an orthogonal matrix $Q = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n] \in \mathbb{R}^{n\times n}$, such that

$$Q^* A Q = H_A \equiv [h_{A;i,j}]_{i,j=1}^{n}, \qquad Q^* L Q = H_L \equiv [h_{L;i,j}]_{i,j=1}^{n} \tag{7}$$

with

$$h_{A;i,j} = 0 \text{ for } i \geq 2j+1, \qquad h_{L;i,j} = 0 \text{ for } i \geq 2j+2. \tag{8}$$

Thus, H_A and H_L are generalized Hessenberg matrices. The relations (7) and (8) yield the following algorithm; see [17] for details.

Algorithm 1 (Generalized Arnoldi process for matrix pairs $\{A, L\}$)

1. **Given** $\boldsymbol{q}_1$ **with** $\|\boldsymbol{q}_1\| = 1$**;**
2. $N := 1$;
3. **for** $j = 1, 2, \ldots, k$ **do**
4.   **if** $j > N$ **then exit;**
5. $\hat{\boldsymbol{q}} := A\boldsymbol{q}_j$;
6.   **for** $i = 1, 2, \ldots N$ **do**
7. $h_{A;i,j} := \boldsymbol{q}_i^* \hat{\boldsymbol{q}}$; $\hat{\boldsymbol{q}} := \hat{\boldsymbol{q}} - \boldsymbol{q}_i h_{A;i,j}$;
8.   **end for**
9. $h_{A;N+1,j} := \|\hat{\boldsymbol{q}}\|$;
10.   **if** $h_{A;N+1,j} > 0$ **then**
11. $N := N + 1$; $\boldsymbol{q}_N := \hat{\boldsymbol{q}}/h_{A;N,j}$;
12.   **end if**
13. $\hat{\boldsymbol{q}} := L\boldsymbol{q}_j$;
14.   **for** $i = 1, 2, \ldots N$ **do**
15. $h_{L;i,j} := \boldsymbol{q}_i^* \hat{\boldsymbol{q}}$; $\hat{\boldsymbol{q}} := \hat{\boldsymbol{q}} - \boldsymbol{q}_i h_{L;i,j}$;
16.   **end for**
17. $h_{L;N+1,j} := \|\hat{\boldsymbol{q}}\|$;
18.   **If** $h_{L;N+1,j} > 0$ **then**
19. $N := N + 1$; $\boldsymbol{q}_N := \hat{\boldsymbol{q}}/h_{A;N,j}$;
20.   **end if**
21. **end for**

The scalar N in the algorithm tracks the number of vectors $\boldsymbol{q}_i$ generated so far during the computations. Let α_k and β_k denote the values of N at the end of Lines 12 and 20, respectively, when $j = k$. It can be seen that $k \leq \alpha_k \leq \beta_k \leq \alpha_k + 1$. Thus, the first β_k columns of the the matrix Q and the leading $\alpha_k \times k$ and $\beta_k \times k$ principal submatrices of the matrices H_A and H_L, respectively, are generated. Generically, Algorithm 1 yields upon completion of k iterations the generalized Arnoldi decompositions

$$
\begin{aligned}
AQ_{(:,1:k)} &= Q_{(:,1:\alpha_k)} H_{A\,(1:\alpha_k,1:k)}, && (9)\\
LQ_{(:,1:k)} &= Q_{(:,1:\beta_k)} H_{L\,(1:\beta_k,1:k)}; && (10)
\end{aligned}
$$

see [17]. Here and throughout, we shall use MATLAB-like notation $X_{(i:j,k:\ell)}$ to denote the submatrix of X, consisting of the intersections of rows i to j and columns k to ℓ, and when $i:j$ is replaced by :, it means all rows, similarly for columns. We note that while $H_{A\,(1:\alpha_k,1:k)}$ and $H_{L\,(1:\beta_k,1:k)}$ are lower banded matrices, their lower bandwidths grow linearly with k.

The computations are terminated at Line 4 in the (rare) event that the recursions for Algorithm 1 break down. In this case, we obtain first N columns of Q with $\alpha_N = \beta_N = N$. This is a benign breakdown as we obtain the decompositions

$$
\begin{aligned}
AQ_{(:,1:N)} &= Q_{(:,1:N)} H_{A\,(1:N,1:N)},\\
LQ_{(:,1:N)} &= Q_{(:,1:N)} H_{L\,(1:N,1:N)}
\end{aligned}
$$

with the column space of $Q_{(:,1:N)}$ being an invariant subspace of both A and L.

An examination of Algorithm 1 shows that $\mathrm{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_\ell\}$ is generated from the sequence of vectors obtained by multiplying $\boldsymbol{q}_1$ by A and L in a periodic fashion,

$$
\begin{array}{ll}
\text{Group 0:} & \boldsymbol{q}_1,\\
\text{Group 1:} & A\boldsymbol{q}_1, L\boldsymbol{q}_1,\\
\text{Group 2:} & A^2\boldsymbol{q}_1, LA\boldsymbol{q}_1, AL\boldsymbol{q}_1, L^2\boldsymbol{q}_1,\\
\text{Group 3:} & A^3\boldsymbol{q}_1, LA^2\boldsymbol{q}_1, ALA\boldsymbol{q}_1, L^2A\boldsymbol{q}_1, A^2L\boldsymbol{q}_1, LAL\boldsymbol{q}_1, AL^2\boldsymbol{q}_1, L^3\boldsymbol{q}_1,\\
\vdots & \vdots
\end{array}
\qquad (11)
$$

ordered from top to bottom and from left to right. Let $\mathrm{g}\mathcal{K}_\ell(\{A, L\}, \boldsymbol{q}_1)$ denote the subspace spanned by the first ℓ vectors in the sequence (11). We refer to $\mathrm{g}\mathcal{K}_\ell(\{A, L\}, \boldsymbol{q}_1)$ as a generalized Krylov subspace. The following result, shown in [17], describes properties of the vectors $\boldsymbol{q}_i$ determined by Algorithm 1.

Theorem 2.1 *(1) If* $\dim \mathrm{g}\mathcal{K}_\ell(\{A, L\}, \boldsymbol{q}_1) = N$*, then*

$$
\mathrm{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_N\} = \mathrm{g}\mathcal{K}_\ell(\{A, L\}, \boldsymbol{q}_1).
$$

(2) If the j-loop of Algorithm 1 *runs to completion, then*

$$
\mathrm{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_N\} = \mathrm{g}\mathcal{K}_{2k+1}(\{A, L\}, \boldsymbol{q}_1).
$$

(3) If Algorithm 1 *is exited at* Line 4*, then for* $\ell \geq 2j-1$*,*

$$\mathrm{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_N\} = \mathrm{g}\mathcal{K}_\ell(\{A, L\}, \boldsymbol{q}_1).$$

If both the matrices A and L are symmetric, then the recursions of Algorithm 1 simplify and define the generalized Lanczos process; see [17] for details. The generalized Arnoldi process requires both matrices A and L to be square. However, many regularization operators L in the literature live in $\mathbb{R}^{p\times n}$ with $p < n$. They can be made square by appending $n-p$ zero rows. This approach to obtain a square regularization operator is discussed in [21]. The zero rows, of course, do not have to be stored.

The generalized Arnoldi process for matrix pairs described above can be extended to m-tuplets of $n \times n$ matrices $\{A_1, A_2, \ldots, A_m\}$. Recall that the basis for Algorithm 1 are the expressions (7) and (8). The following theorem presents the corresponding expressions for matrix m-tuplets. An algorithm, analogous to Algorithm 1, based on this theorem readily can be derived.

Theorem 2.2 *Let* $A_\ell \in \mathbb{R}^{n\times n}$*,* $1 \leq \ell \leq m$*. Given* $\boldsymbol{q}_1 \in \mathbb{R}^n$ *of unit length, there is an orthogonal matrix* $Q \in \mathbb{R}^{n\times n}$ *with* $Q\boldsymbol{e}_1 = \boldsymbol{q}_1$*, such that*

$$Q^* A_\ell Q = H_\ell \equiv [h_{\ell;i,j}] \quad \text{for } 1 \leq \ell \leq m, \tag{12}$$

with

$$h_{\ell;i,j} = 0 \quad \text{for } i \geq mj + m - \ell - 1,$$

where $\boldsymbol{e}_1 = [1, 0, \ldots, 0]^* \in \mathbb{R}^n$ *denotes the first axis vector. Thus, the matrices* H_ℓ *are generalized Hessenberg matrices.*

Theorem 2.2 and the corresponding reduction algorithm are of interest in the context of multiparameter Tikhonov regularization of the form

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \sum_{j=1}^{\ell-1} \mu_j \|L_j \boldsymbol{x}\|^2 \right\}, \tag{13}$$

where the matrices $L_j \in \mathbb{R}^{n\times n}$ are regularization operators and the scalars $\mu_j \geq 0$ regularization parameters. We will comment on this application in the following section. An analog of Theorem 2.2 holds for complex matrices; however, this generalization is not used in the present paper.

3 Tikhonov regularization based on the generalized Arnoldi process

We discuss the application of the decompositions (9) and (10) to the solution of large-scale Tikhonov minimization problems (4). Multiparameter Tikhonov regularization is considered at the end of the section.

Let the initial vector for Algorithm 1 be

$$\boldsymbol{q}_1 = \frac{A\boldsymbol{b}}{\|A\boldsymbol{b}\|}. \tag{14}$$

Then the algorithm determines the generalized Arnoldi decompositions (9) and (10). The indices α_k and β_k in these decompositions are the values of N at the end of Lines 12 and 20, respectively, at the last iteration. The initial vector (14) yields more accurate approximations of the desired solutions for many linear discrete ill-posed problems than the initial vector

$$\boldsymbol{q}_1 = \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|}. \tag{15}$$

A reason for this is that in our applications, the matrix A is a low-pass filter. Therefore, much high-frequency noise is damped in (14), but not in (15). A numerical illustration of the performance of Algorithm 1 for the initial vectors (14) and (15) is provided in Section 5.

When $L = I$, the method of the present paper with initial vector (14) simplifies to the Range Restricted Arnoldi-Tikhonov regularization method described in [16]. Computed examples in [16] illustrate the benefit of the initial vector (14) over (15) in this simpler setting.

Assume that the generalized Arnoldi decompositions (9) and (10) are available with $\boldsymbol{q}_1$ defined by (14). Let $\boldsymbol{x} \in \mathrm{g}\mathcal{K}_k(\{A, L\}, \boldsymbol{q}_1)$ and let $\boldsymbol{u} \in \mathbb{R}^k$ satisfy $\boldsymbol{x} = Q_{(:,1:k)}\boldsymbol{u}$. Then

$$\begin{aligned}
\|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \mu\|L\boldsymbol{x}\|^2 &= \|AQ_{(:,1:k)}\boldsymbol{u} - \boldsymbol{b}\|^2 + \mu\|LQ_{(:,1:k)}\boldsymbol{u}\|^2 \\
&= \|Q_{(:,1:\alpha_k)}H_{A\,(1:\alpha_k,1:k)}\boldsymbol{u} - Q_{(:,1:\alpha_k)}Q^*_{(:,1:\alpha_k)}\boldsymbol{b}\|^2 \\
&\quad + \|(I - Q_{(:,1:\alpha_k)}Q^*_{(:,1:\alpha_k)})\boldsymbol{b}\|^2 \\
&\quad + \mu\|Q_{(:,1:\beta_k)}H_{L\,(1:\beta_k,1:k)}\boldsymbol{u}\|^2 \\
&= \|H_{A\,(1:\alpha_k,1:k)}\boldsymbol{u} - Q^*_{(:,1:\alpha_k)}\boldsymbol{b}\|^2 \\
&\quad + \|(I - Q_{(:,1:\alpha_k)}Q^*_{(:,1:\alpha_k)})\boldsymbol{b}\|^2 \\
&\quad + \mu\|H_{L\,(1:\beta_k,1:k)}\boldsymbol{u}\|^2.
\end{aligned} \tag{16}$$

Thus, solving the minimization problem (4) over the generalized Krylov subspace $\mathrm{g}\mathcal{K}_k(\{A, L\}, \boldsymbol{q}_1)$ simplifies to the solution of the projected, generally fairly small, problem

$$\min_{\boldsymbol{u}\in\mathbb{R}^k}\{\|H_{A\,(1:\alpha_k,1:k)}\boldsymbol{u} - Q^*_{(:,1:\alpha_k)}\boldsymbol{b}\|^2 + \mu\|H_{L\,(1:\beta_k,1:k)}\boldsymbol{u}\|^2\}. \tag{17}$$

It may be solved, e.g., by first computing the GSVD of the matrix pair $\{H_{A\,(1:\alpha_k,1:k)}, H_{L\,(1:\beta_k,1:k)}\}$. However, typically, the matrix $H_{L\,(1:\beta_k,1:k)}$ is of full rank and not ill-conditioned. The solution therefore can be computed in a simpler fashion. Note that $\beta_k \geq k$ and introduce the QR factorization

$$H_{L\,(1:\beta_k,1:k)} = \widetilde{Q}_L\widetilde{R}_L, \tag{18}$$

where $\widetilde{Q}_L \in \mathbb{R}^{\beta_k\times k}$ has orthonormal columns and $\widetilde{R}_L \in \mathbb{R}^{k\times k}$ is upper triangular and nonsingular. Substituting (18) and $\boldsymbol{w} = \widetilde{R}_L\boldsymbol{u}$ into (17) yields a Tikhonov minimization problem in standard form,

$$\min_{\boldsymbol{w}\in\mathbb{R}^k}\{\|H_{A\,(1:\alpha_k,1:k)}\widetilde{R}_L^{-1}\boldsymbol{w} - Q^*_{(:,1:\alpha_k)}\boldsymbol{b}\|^2 + \mu\|\boldsymbol{w}\|^2\}. \tag{19}$$

Since the matrix $H_{A\,(1:\alpha_k,1:k)}\widetilde{R}_L^{-1}$ is not large, we may use its singular value decomposition to solve (19).

Various approaches can be used to determine a suitable value of the regularization parameter for this problem, including the L-curve, Generalized Cross Validation, and the discrepancy principle. The latter is applied below. We also may set $\mu = 0$ and truncate the SVD of $H_{A\,(1:\alpha_k,1:k)}\widetilde{R}_L^{-1}$.

Let $\boldsymbol{w}_\mu^{(k)}$ denote the solution of (19) for a fixed $\mu > 0$. Then the associated approximate solution of (4) is given by

$$\boldsymbol{x}_\mu^{(k)} = Q_{(:,1:k)}\widetilde{R}_L^{-1}\boldsymbol{w}_\mu^{(k)}. \tag{20}$$

The value of k is assumed to be small, but large enough, so that the approximate solutions $\boldsymbol{x}_\mu^{(k)}$ for μ-values of interest provide meaningful approximations of the corresponding solutions $\boldsymbol{x}_\mu$ of (4). The residual error associated with $\boldsymbol{x}_\mu^{(k)}$,

$$\boldsymbol{r}_\mu^{(k)} = \boldsymbol{b} - A\boldsymbol{x}_\mu^{(k)}, \tag{21}$$

commonly is referred to as the discrepancy.

Assume that a bound for the error $\boldsymbol{e}$ in $\boldsymbol{b}$,

$$\|\boldsymbol{e}\| \leq \varepsilon,$$

is known. We will describe a projected discrepancy principle for the determination of a suitable value of μ using this bound. Introduce the matrix

$$\breve{H} = H_{A\,(1:\alpha_k,1:k)}\widetilde{R}_L^{-1} \tag{22}$$

and let $P_{\mathcal{R}(\breve{H})}$ denote the orthogonal projector onto $\mathcal{R}(\breve{H})$, where $\mathcal{R}(\breve{H})$ denotes the range of $\breve{H}$. Then, in view of (20) and (21), we obtain, analogously to (16), that

$$\begin{aligned}
\|\boldsymbol{r}_\mu^{(k)}\|^2 &= \|\boldsymbol{b} - A\boldsymbol{x}_\mu^{(k)}\|^2 \\
&= \|\breve{H}\boldsymbol{w}_\mu^{(k)} - Q^*_{(:,1:\alpha_k)}\boldsymbol{b}\|^2 + \|(I - Q_{(:,1:\alpha_k)}Q^*_{(:,1:\alpha_k)})\boldsymbol{b}\|^2 \\
&= \|\breve{H}\boldsymbol{w}_\mu^{(k)} - P_{\mathcal{R}(\breve{H})}Q^*_{(:,1:\alpha_k)}\boldsymbol{b}\|^2 + \|P^\perp_{\mathcal{R}(\breve{H})}Q^*_{(:,1:\alpha_k)}\boldsymbol{b}\|^2 \\
&\quad + \|(I - Q_{(:,1:\alpha_k)}Q^*_{(:,1:\alpha_k)})\boldsymbol{b}\|^2.
\end{aligned}$$

We determine μ, so that

$$\|\breve{H}\boldsymbol{w}_\mu^{(k)} - P_{\mathcal{R}(\breve{H})}Q^*_{(:,1:\alpha_k)}\boldsymbol{b}\| = \eta\varepsilon, \tag{23}$$

where $\eta > 1$ is a user-specified constant independent of ε and $\boldsymbol{w}_\mu^{(k)}$ solves (19). We refer to this approach to determine μ as the projected discrepancy principle. Let μ_k denote the solution of (23). Clearly, $\|\boldsymbol{r}_{\mu_k}^{(k)}\| \geq \eta\varepsilon$. The computations may be carried out conveniently with the aid of the SVD of $\breve{H}$. We note that the error in $P_{\mathcal{R}(\breve{H})}Q^*_{(:,1:\alpha_k)}\boldsymbol{b}$ generally is of norm smaller than ε. Therefore, it may be possible to reduce the right-hand side in (23) before determining μ. However, lacking information about the error in $P_{\mathcal{R}(\breve{H})}Q^*_{(:,1:\alpha_k)}\boldsymbol{b}$, we determine μ as the solution of (23) in the numerical examples of Section 5.

The following theorem discusses properties of the left-hand side of (23) as a function of $\nu = 1/\mu$.

Theorem 3.1 *Assume that the generalized Arnoldi process Algorithm 1 with $\boldsymbol{q}_1$ defined by (14) completes k iterations. Let $\breve{H}$ be defined by (22), and introduce $\breve{\boldsymbol{b}} = P_{\mathcal{R}(\breve{H})}Q^*_{(:,1:\alpha_k)}\boldsymbol{b}$. Suppose that $\breve{H}^*\breve{\boldsymbol{b}} \neq \boldsymbol{0}$. Then the function*

$$\nu \to \phi^{(k)}(\nu) = \breve{\boldsymbol{b}}^*(\nu\breve{H}\breve{H}^* + I)^{-2}\breve{\boldsymbol{b}}, \qquad \nu \geq 0,$$

is strictly decreasing, convex, and

$$\phi^{(k)}(0) = \|\breve{\boldsymbol{b}}\|^2, \qquad \lim_{\nu\to\infty} \phi^{(k)}(\nu) = 0. \tag{24}$$

Moreover,

$$\phi^{(k)}(1/\mu) = \|\breve{H}\boldsymbol{w}_\mu^{(k)} - P_{\mathcal{R}(\breve{H})} Q^*_{(:,1:\alpha_k)}\boldsymbol{b}\|^2. \tag{25}$$

Proof. Related results are shown in [16, Theorem 2.1]. Substituting (20) into (21) and using the identity

$$I - \breve{H}(\breve{H}^*\breve{H} + \mu I)^{-1}\breve{H}^* = (\mu^{-1}\breve{H}\breve{H}^* + I)^{-1}$$

yields (25). The convexity and strict decrease of $\phi^{(k)}(\nu)$ are consequences of $\breve{H}^*\breve{\boldsymbol{b}} \neq \mathbf{0}$. □

As we can expect that $\eta\varepsilon << \|\boldsymbol{b}\|$, then for sufficiently large k, the inequality

$$\phi^{(k)}(0) > \eta^2\varepsilon^2 \tag{26}$$

holds. Indeed, for all reasonable problems, the inequality (26) holds for a modest value of k. In view of Theorem 3.1, the equation

$$\phi^{(k)}(\nu) = \eta^2\varepsilon^2 \tag{27}$$

has a unique solution $\hat{\nu}$. We use the regularization parameter $\hat{\mu} = 1/\hat{\nu}$ in (19) and compute the approximate solution $\boldsymbol{x}_{\hat{\mu}}^{(k)}$ of (4) using (20).

Many zero-finders can be used to solve (27). The monotonic decrease and convexity of $\phi^{(k)}(\nu)$ secure that Newton's method with an initial iterate $\nu_0 < \breve{\nu}$, such as $\nu_0 = 0$, converges monotonically and quadratically; see [20] for a discussion on zero-finders.

We turn to multiparameter Tikhonov regularization of the form (13). A nice discussion for small to moderately sized problems is provided by Brezinski et al. [3]. An extension of Algorithm 1 based on Theorem 2.2 can be used to determine the following decompositions, which are analogous to (9) and (10),

$$\begin{aligned} AQ_{(:,1:k)} &= Q_{(:,1:\alpha_k)} H_{A\,(1:\alpha_k,1:k)}, \\ L_j Q_{(:,1:k)} &= Q_{(:,1:\beta_k^{(j)})} H_{L_j\,(1:\beta_k^{(j)},1:k)}, \qquad j = 1, 2, \ldots, \ell - 1. \end{aligned}$$

Letting $\boldsymbol{x} = Q_{(:,1:k)}\boldsymbol{u}$, we proceed similarly as in the derivation of (17) and obtain that the minimization problem (13) over $\mathcal{R}(Q_{(:,1:k)})$ is equivalent to the reduced problem

$$\min_{\boldsymbol{u}\in\mathbb{R}^k} \left\{ \|H_{A\,(1:\alpha_k,1:k)}\boldsymbol{u} - Q^*_{(:,1:\alpha_k)}\boldsymbol{b}\|^2 + \sum_{j=1}^{\ell-1} \mu_j \|H_{L_j\,(1:\beta_k^{(j)},1:k)}\boldsymbol{u}\|^2 \right\},$$

which can be solved, e.g., as described in [3].

4 Selective regularization by augmentation

For some linear discrete ill-posed problems (1) the null space of the regularization operator L is important, because the solution component in $\mathcal{N}(L)$ is determined independently of $\mu > 0$ and therefore is not damped. It is often desirable that important features of the solution $\hat{\boldsymbol{x}}$ can be represented by vectors in $\mathcal{N}(L)$.

Example 4.1. Assume that $\hat{\boldsymbol{x}}$ can be approximated well by a uniform discretization of a linearly increasing real-valued function on a real interval. Then it is generally beneficial to choose a regularization operator L with

$$\mathcal{N}(L) = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n \end{bmatrix}.$$

A computed illustration can be found in [6, Example 6]. Many other examples are reported in [21]; see also Section 5. □

A drawback of the solution method described in Section 3 is that the regularization operator $H_{L\,(1:\beta_k,1:k)}$ of the reduced Tikhonov minimization problem (17) generally is of full rank, even though the regularization operator L in (4) might not be. Therefore, the solution component in $\mathcal{N}(L)$ is not determined independently of $\mu > 0$ when solving the reduced problem (17). This section describes how $\mathcal{N}(L)$ can be incorporated into the solution process by using the decomposition method described in [2].

Let the matrix $U \in \mathbb{R}^{n\times p}$, $p \ll n$, have orthonormal columns, which span $\mathcal{N}(L)$ or, more generally, let the columns represent components of the desired solution $\hat{\boldsymbol{x}}$ that should not be damped during the solution process. Compute the QR factorization

$$AU = WR,$$

where $W \in \mathbb{R}^{n\times p}$ has orthonormal columns and $R \in \mathbb{R}^{p\times p}$ is upper triangular. We assume that U is chosen so that R is of full rank and not ill-conditioned. Introduce the orthogonal projectors

$$P_U = UU^*, \qquad P_U^\perp = I - UU^*, \qquad P_W = WW^*, \qquad P_W^\perp = I - WW^*,$$

and partition the vector $\boldsymbol{x} = P_U\boldsymbol{x} + P_U^\perp\boldsymbol{x}$ and the linear system of equations (1) according to

$$P_W A P_U \boldsymbol{x} + P_W A P_U^\perp \boldsymbol{x} = P_W \boldsymbol{b}, \tag{28}$$
$$P_W^\perp A P_U^\perp \boldsymbol{x} = P_W^\perp \boldsymbol{b}, \tag{29}$$

where we have used the fact that $P_W^\perp A P_U = O$.

The system (28)-(29) is solved by back substitution. We first determine an approximate solution of (29) by the method described in Section 3. Since $P_W^\perp A P_U^\perp = P_W^\perp A$, we may ignore the projector $P_U^\perp$ in the left-hand side of (29). Moreover, we use the regularization operator $LP_U^\perp$ when computing an approximate solution of (29) by Tikhonov regularization. Using the operator $LP_U^\perp$, instead of L, yields a Tikhonov minimization problem, which ignores the solution component in $\mathcal{R}(U)$. Thus, we solve the minimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \{\|P_W^\perp A\boldsymbol{x} - P_W^\perp \boldsymbol{b}\|^2 + \mu\|LP_U^\perp \boldsymbol{x}\|^2\}. \tag{30}$$

Denote the computed approximate solution of (30) by $\boldsymbol{x}'$. Substituting $\boldsymbol{x}'$ into (28) yields the equation

$$P_W A P_U \boldsymbol{x} = P_W \boldsymbol{b} - P_W A \boldsymbol{x}',$$

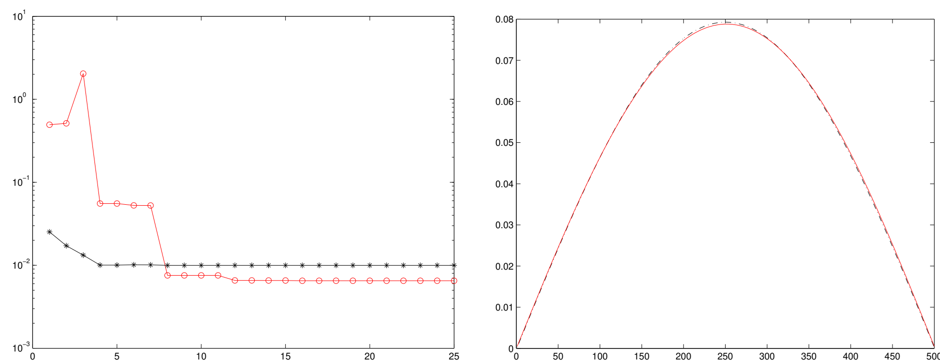
whose solution, $\boldsymbol{x}'' = P_U\boldsymbol{x}$, can be computed by a direct method, because $P_W A P_U$ can be represented by a small $p \times p$ matrix. Finally, the desired approximate solution of (1) is given by $\boldsymbol{x}'' + \boldsymbol{x}'$. Implementation details and some properties of this decomposition scheme are discussed in [2,21].

5 Numerical examples

We illustrate the performance of the solution method discussed with a few computed examples. The "noise-vector" $\boldsymbol{e}$ has in all examples normally distributed pseudorandom entries with mean zero, and is normalized to correspond to a chosen noise-level

$$\delta = \frac{\|\boldsymbol{e}\|}{\|\hat{\boldsymbol{b}}\|}.$$

Here $\hat{\boldsymbol{b}}$ denotes the noise-free right-hand side vector in (3). In all examples, the regularization parameter is determined by the discrepancy principle with the constant η in (23) set to 1.1. The first two examples illustrate the performance of our method when applied to the solution of two linear discrete ill-posed problems from Regularization Tools [14]. The last two examples show applications to the restoration of images that have been contaminated by blur and noise. The latter examples use a regularization operator, whose A-weighted pseudoinverse is unattractive for computation. The initial vector for the generalized Arnoldi process is given by (14), unless explicitly stated otherwise. All computations are carried out in MATLAB with about 16 significant decimal digits.

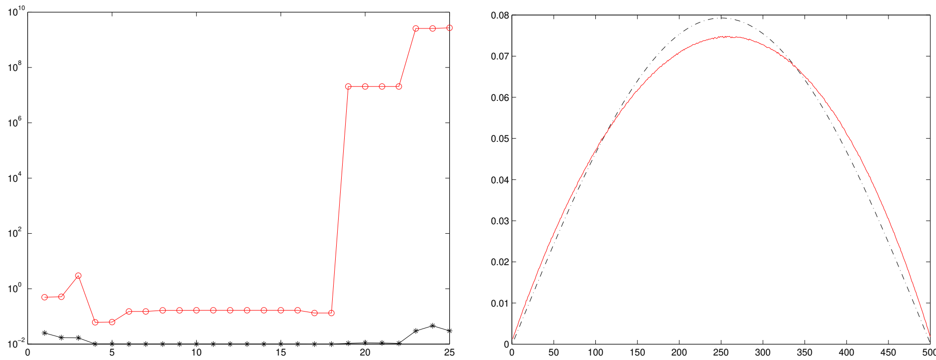


(a) (b)

Fig. 1. Example 5.1: Application of the tridiagonal regularization operator (32). (a) Norm of the relative error in the computed approximate solutions $\boldsymbol{x}_{\mu_k}^{(k)}$ marked by "o" (in red) and norm of the associated relative residual error $\boldsymbol{r}_{\mu_k}^{(k)}$ marked by "$*$" as functions of k. (b) The computed approximate solution $\boldsymbol{x}_{\mu_{15}}^{(15)}$ (red continuous graph) and the desired solution $\hat{\boldsymbol{x}}$ (black dash-dotted graph).

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

$$\int_0^{\pi/2} \kappa(\sigma,\tau)x(\sigma)d\sigma = b(\tau), \qquad 0 \le \tau \le \pi, \tag{31}$$



(a) (b)

Fig. 2. Example 5.1: Application of the regularization operator $L = I$. (a) Norm of the relative error in the computed approximate solutions $\boldsymbol{x}_{\mu_k}^{(k)}$ marked by "o" (in red) and norm of the associated relative residual error $\boldsymbol{r}_{\mu_k}^{(k)}$ marked by "$*$" as functions of k. (b) The computed approximate solution $\boldsymbol{x}_{\mu_4}^{(4)}$ (red continuous graph) and the desired solution $\hat{\boldsymbol{x}}$ (black dash-dotted graph).

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 code baart from [14] to discretize (31) by a Galerkin method with 500 orthonormal box functions as test and trial functions. The code produces the 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 noise-level $\delta = 1\cdot 10^{-2}$. The right-hand side $\boldsymbol{b}$ in the system (1) is obtained from (2). This example computes approximations of $\hat{\boldsymbol{x}}$ by the approach of Section 3 without augmenting the generalized Krylov subspaces generated by auxiliary vectors.

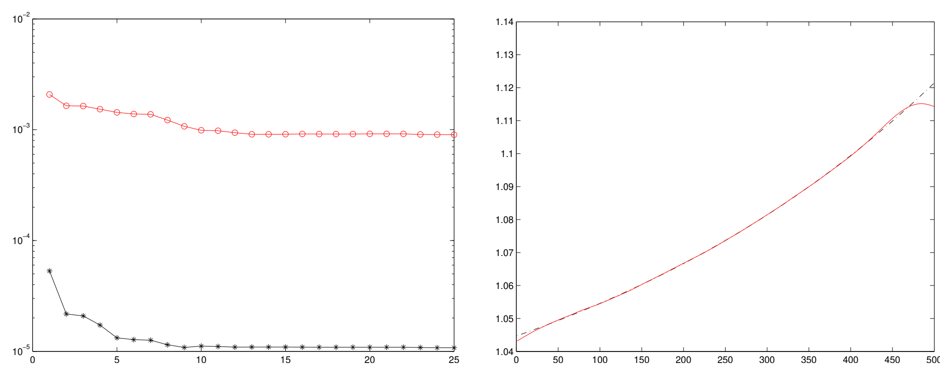
We first use the symmetric tridiagonal regularization operator

$$L = \begin{bmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & & & -1 \\ 0 & & & -1 & 2 \end{bmatrix} \in \mathbb{R}^{n\times n}. \tag{32}$$

Figure 1(a) displays the relative error $\|\boldsymbol{x}_{\mu_k}^{(k)} - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\|$ in the computed approximate solution $\boldsymbol{x}_{\mu_k}^{(k)}$ for $1 \le k \le 25$. These errors are marked by "o" (in red) and connected by line segments. The μ_k solve (27). Figure 1(a) also shows the associated relative residual errors $\|\boldsymbol{r}_{\mu_k}^{(k)}\|/\|\hat{\boldsymbol{b}}\|$, which are marked by "$*$" (in black) and connected by line segments. The figure shows the error in $\boldsymbol{x}_{\mu_k}^{(k)}$ to be essentially independent of the number of steps, k, carried out with the generalized Arnoldi process for $k \ge 8$. Figure 1(b) depicts the computed approximate solution $\boldsymbol{x}_{\mu_{15}}^{(15)}$ (red continuous graph) and the solution $\hat{\boldsymbol{x}}$ of the error-free linear system (3) (black dash-dotted graph). The former has relative error $6.5 \cdot 10^{-3}$.

Figure 2 is analogous to Figure 1 and displays results obtained when the regularization operator $L = I$ is used. This choice of L allows the application of the standard Arnoldi process with initial vector (14). Comparing Figures 1 and 2 shows that the tridiagonal regularization operator (32) yields approximate solutions of significantly higher quality. For instance, Figure 2(a) shows $\boldsymbol{x}_{\mu_4}^{(4)}$ to approximate $\hat{\boldsymbol{x}}$ the best; the relative error is $6.2 \cdot 10^{-2}$. Figure 2(b) displays $\boldsymbol{x}_{\mu_4}^{(4)}$ and $\hat{\boldsymbol{x}}$.

Finally, we consider the use of the initial vector (15) for the generalized Arnoldi process. Application of this initial vector and the regularization operator (32) yields the approximate solution $\boldsymbol{x}_{\mu_{15}}^{(15)}$ with relative error $7.0 \cdot 10^{-3}$. This error is larger than the error of the approximate solution displayed by Figure 1(b), which is obtained with the initial vector (14). Generally, generalized Krylov subspaces with initial vector (15) yield approximate solutions of worse quality than generalized Krylov subspaces with initial vector (14). $\square$

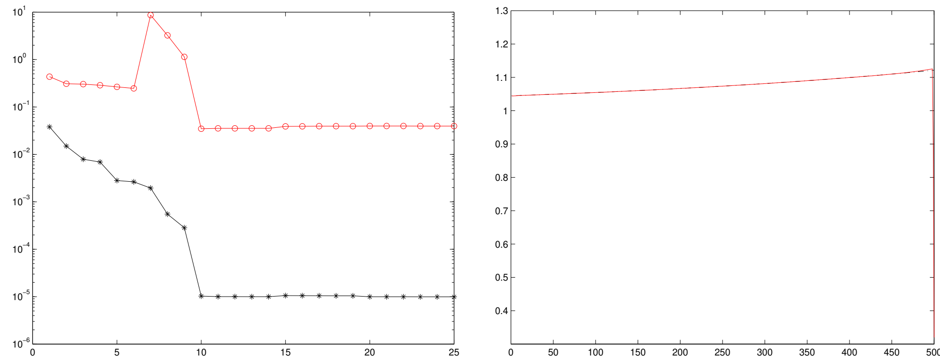


(a) (b)

Fig. 3. Example 5.2: Augmentation of the generalized Krylov subspaces generated by the vectors (34) and application of the regularization operator (35). (a) Norm of the relative error in the computed approximate solutions $\boldsymbol{x}_{\mu_k}^{(k)}$ marked by "o" (in red) and norm of the associated relative residual error $\boldsymbol{r}_{\mu_k}^{(k)}$ marked by "$*$" as functions of k. (b) The computed approximate solution $\boldsymbol{x}_{\mu_{25}}^{(25)}$ (red continuous graph) and the desired solution $\hat{\boldsymbol{x}}$ (black dash-dotted graph).

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

$$\int_0^1 k(s,t)x(t)dt = \exp(s) + (1-e)s - 1, \qquad 0 \leq s \leq 1, \tag{33}$$



(a) (b)

Fig. 4. Example 5.2: Application of the regularization operator (35) without augmentation. (a) Norm of the relative error in the computed approximate solutions $\boldsymbol{x}_{\mu_k}^{(k)}$ marked by "o" (in red) and norm of the associated relative residual error $\boldsymbol{r}_{\mu_k}^{(k)}$ marked by "$*$" as functions of k. (b) The computed approximate solution $\boldsymbol{x}_{\mu_{10}}^{(10)}$ (red continuous graph) and the desired solution $\hat{\boldsymbol{x}}$ (black dash-dotted graph).

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 [14]. This yields the matrix $A \in \mathbb{R}^{500\times 500}$. The code deriv2 also produces a scaled discrete approximation $\hat{\boldsymbol{x}} \in \mathbb{R}^{500}$ of the solution $x(t) = \exp(t)$ of (33). The error vector $\boldsymbol{e} \in \mathbb{R}^{500}$ has noise-level $1 \cdot 10^{-5}$. This example illustrates the possible benefit of augmentation of the generalized Krylov subspaces. We augment by the span of the vectors

$$\boldsymbol{w}_1 = [1, 1, \ldots, 1]^* \in \mathbb{R}^n, \qquad \boldsymbol{w}_2 = [1, 2, \ldots, n]^* \in \mathbb{R}^n. \tag{34}$$

This augmentation is meaningful, because the desired solution is the discretization of a smooth monotonically increasing function. We use the bidiagonal regularization operator

$$L = \begin{bmatrix} 1 & -1 & & & 0 \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & & -1 \\ 0 & & & & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}, \tag{35}$$

Figure 3 is analogous to the figures for Example 5.1. In particular, Figure 3(b) displays the computed approximate solution $\boldsymbol{x}_{\mu_{25}}^{(25)}$ with relative error $9.0 \cdot 10^{-4}$. Results without augmentation for the same regularization operator are shown in Figure 4. The most accurate approximate computed solution, $\boldsymbol{x}_{\mu_{10}}^{(10)}$, is shown in Figure 4(b).

This example shows that augmentation may increase the quality of the computed approximate solution. The regularization operators (32) and $L = I$ give worse approximate solutions for the present example than (35). We therefore do not display these errors. Finally, we note that for larger noise-levels than $1 \cdot 10^{-5}$, such as $\delta = 1 \cdot 10^{-4}$, the right-hand side $P_W^{\perp}\boldsymbol{b}$ of (29) is small enough to satisfy the discrepancy principle and no iterations have to be carried out; we then accept the solution zero of this system. This illustrates that augmentation may reduce the computational effort significantly. Since the purpose of the present example is to illustrate the performance of the generalized Arnoldi process, we choose the noise-level sufficiently small to secure that the system (29) has to be explicitly solved. $\square$

The following examples illustrate applications of our method to the restoration of two-dimensional gray-scale images that have been contaminated by blur and noise. The gray-scale images are represented by an array of $m \times m$ pixels, with each pixel stored in 8 bits. This allows pixel values in the interval $[0, 255]$. The pixels are ordered column-wise and stored in a vector of dimension $n = m^2$. Let $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ represent a blur- and noise-free image. We generate an associated blurred and noise-free image $\hat{\boldsymbol{b}}$ by multiplying $\hat{\boldsymbol{x}}$ by a block Toeplitz matrix $A \in \mathbb{R}^{n \times n}$ with Toeplitz blocks. The matrix A represents a Gaussian blurring operator and is generated with the MATLAB function `blur.m` from [14]. This function has two parameters `band` and `sigma`. The former specifies the half-bandwidth of the Toeplitz blocks and the latter the variance of the Gaussian point spread function. The larger `sigma` is, the more blurring. Enlarging `band` increases the storage requirement, the arithmetic work required for the evaluation of matrix-vector products with A, and to some extent the blurring. The blur- and noise-contaminated image $\boldsymbol{b} \in \mathbb{R}^n$ is obtained by adding a “noise-vector” $\boldsymbol{e} \in \mathbb{R}^n$ to $\hat{\boldsymbol{b}}$; cf. (2). We assume the blurring operator 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, which accurately approximates the blur- and noise-free image $\hat{\boldsymbol{x}}$.

A quantitative comparison of the restored images determined with different regularization operators L is provided by the Peak Signal-to-Noise Ratio (PSNR),

$$\mathrm{PSNR}(\boldsymbol{x}_{\mu_k}^{(k)}, \hat{\boldsymbol{x}}) = 20 \log_{10} \frac{255}{\|\boldsymbol{x}_{\mu_k}^{(k)} - \hat{\boldsymbol{x}}\|} \, \mathrm{dB},$$

(a) (b)

Fig. 5. Blur- and noise-free images for Examples 5.3 and 5.4.

where $\boldsymbol{x}_{\mu_k}^{(k)}$ is the computed restoration and the numerator, 255, is the largest pixel-value allowed. A high PSNR-value indicates that the restoration $\boldsymbol{x}_{\mu_k}^{(k)}$ is accurate. However, PSNR-values do not always agree with visual judgment. We therefore also display the restored images.

Example 5.3. We consider the restoration of a contaminated version of the blur- and noise-free 512×512-pixel image shown in Figure 5 (left). The performance of three regularization operators is compared: the identity, the Laplacian

$$\mathcal{L}(\boldsymbol{x}) = \Delta \boldsymbol{x}, \tag{36}$$

and the Perona-Malik operator

$$\mathcal{L}(\boldsymbol{x}) = \operatorname{div}(g(|\nabla \boldsymbol{x}|^2)\nabla \boldsymbol{x}), \tag{37}$$

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

$$g(s) = 1/(1 + s/\rho),$$

with $\rho > 0$ a small positive constant; see [19] for discussions on this choice of diffusivity.

We discretize (37) by finite differences. This gives a matrix $L(\boldsymbol{x})$ 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, denoted

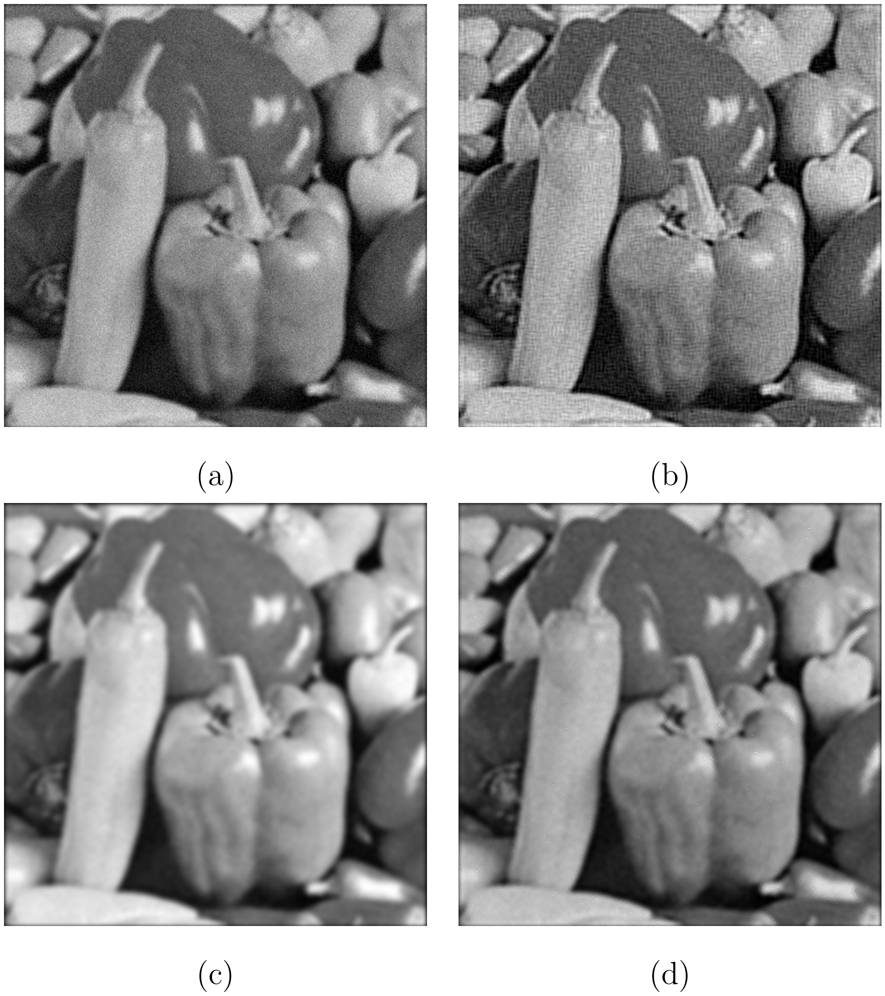


Fig. 6. Example 5.3: (a) Blurred and noisy image, images restored with (b) $L = I$, (c) $L = \Delta$, and (d) L the discrete Perona-Malik operator.

by (N, S, E, W). The row of $L(\boldsymbol{x})$ associated with pixel (i, j) is generically of the form

$$\{l_{i,j,S}, 0, \ldots, 0, l_{i,j,E}, -(l_{i,j,S} + l_{i,j,E} + l_{i,j,W} + l_{i,j,N}), l_{i,j,W}, 0, \ldots, 0, l_{i,j,N}\},$$

with elements

$$l_{i,j,E} = \frac{g_{i,j} + g_{i+1,j}}{2}, \qquad l_{i,j,W} = \frac{g_{i,j} + g_{i-1,j}}{2},$$

where $g_{i,j}$ represents the value of the discretization of the diffusivity $g(|\nabla \boldsymbol{x}|^2)$ at pixel (i, j). Partial derivatives are approximated by central finite differences, giving

$$g_{i,j} = g\left(\left(\frac{x_{i+1,j} - x_{i-1,j}}{2}\right)^2 + \left(\frac{x_{i,j+1} - x_{i,j-1}}{2}\right)^2\right),$$

where $x_{i,j}$ denotes the value of $\boldsymbol{x}$ at pixel (i, j). Expressions for $l_{i,j,S}$ and $l_{i,j,N}$ can be derived similarly; see [22] for details. Alternative discretizations are discussed in [12].

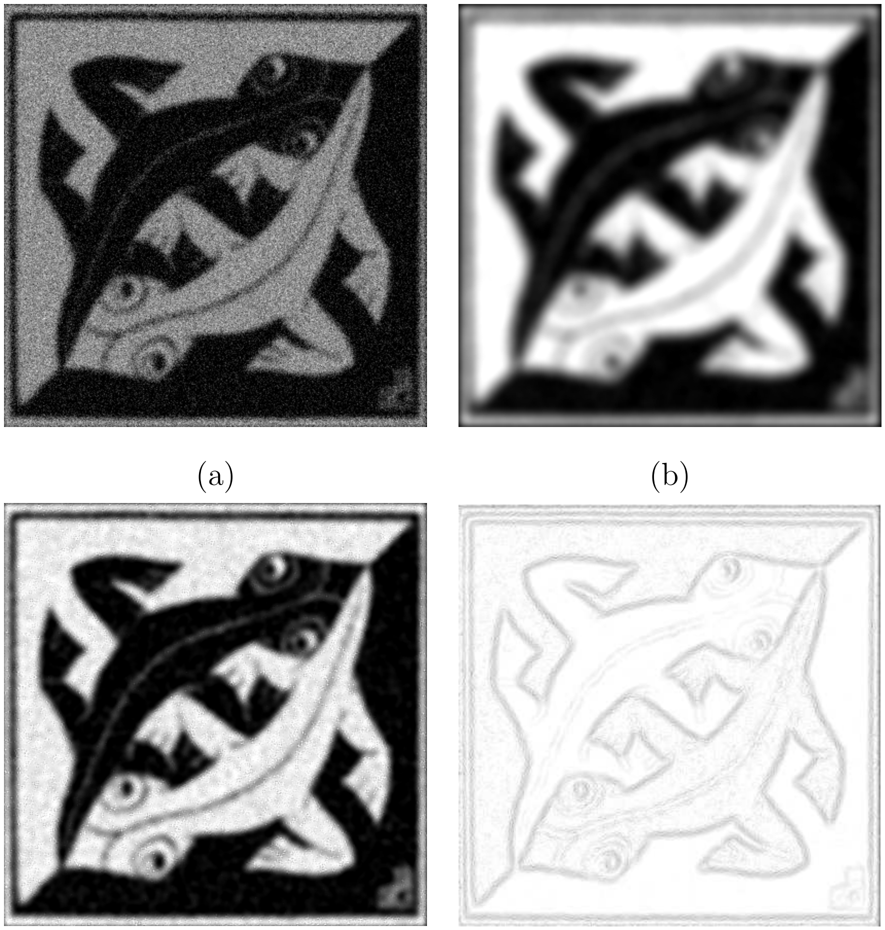
The choice of regularization operator L in (4) is important for the success of the restoration. A linearization of the Perona-Malik operator generally provides restorations of higher quality than the Laplacian (36), which typically yields over-smoothed restorations. We linearize the Perona-Malik operator by using an approximation of the desired image. Specifically, we carry out two iterations with the linearized Perona-Malik operator defined by the available blurred and noisy image to obtain an approximation, $\tilde{\boldsymbol{x}}$, of the desired image. A new linearized Perona-Malik operator then is defined by $L = L(\tilde{\boldsymbol{x}})$ and the generalized Arnoldi process is restarted with this regularization operator to improve $\tilde{\boldsymbol{x}}$. We have found this two-phase approach to determine a suitable regularization operator and an approximation of the noise- and blur-free image $\hat{\boldsymbol{x}}$ to perform well. In particular, it was not beneficial to carry out more than two initial iterations to determine the regularization operator $L = L(\tilde{\boldsymbol{x}})$ used in all subsequent iterations. Moreover, it was not worthwhile to update the regularization operator more than once.

Note that it is much easier and faster to evaluate matrix-vector products with L than to solve linear systems of equations with the matrix L. The use of the generalized Arnoldi method allows us to avoid the latter.

We generate a blurred and noisy image in the manner described above with blur parameters $\texttt{band} = 7$, $\texttt{sigma} = 5$ and noise-level $\delta = 1 \cdot 10^{-1}$. This image is displayed in Figure 6(a). The most accurate restorations for the regularization operators $L = I$ and $L = \Delta$ are obtained with $k = 7$ steps of the generalized Arnoldi process with the regularization parameter determined by (27). These restorations have PSNR-values 25.35 and 24.06, respectively. They are displayed in Figures 6(b) and (c). The best restoration with the Perona-Malik regularization operator is obtained with $k = 6$ steps of the generalized Arnoldi process. It has PSNR-value 25.53 and is shown in Figure 6(d). Visual inspection of the restored images shows the edge preserving Perona-Malik regularization operator to yield the most accurate restoration. $\Box$

Both restorations in the previous example show boundary effects. These often can be reduced by imposing appropriate boundary conditions; see, e.g., Donatelli et al. [8]. It is straightforward to implement suitable boundary conditions in our iterative scheme.

Example 5.4. We illustrate the performance of the regularization operators (36) and (37) when applied to the restoration of a blurred and noisy version of the 412×412-pixel image, with many thin details and very well defined edges, shown in Figure 6(b). Contamination by blur, determined by the parameters



(a) (b) (c) (d)

Fig. 7. Example 5.4: (a) Blurred and noisy image, (b) images restored with $L = \Delta$, (c) L the discrete Perona-Malik operator, (d) and edge map for the latter restoration.

`band` $= 9$ and `sigma` $= 3$ of the function `blur.m`, and by 30% additive Gaussian noise yields the image in Figure 7(a).

We compare the Laplacian regularization operator (36) with a linearized Perona-Malik regularization operator. The latter is defined similarly as in Example 5.3. For both regularization operators, the best restorations are obtained with $k = 6$ steps of the generalized Arnoldi process. The PSNR-values for the restored images determined with the Laplacian and linearized Perona-Malik operators are 11.81 and 15.24, respectively. The restorations are displayed in Figures 7(b) and (c). The over-smoothing produced by the Laplacian regularization operator is clearly visible. Figure 7(d) shows an edge map for the image restored with the Perona-Malik regularization operator. The edge map is seen to be accurate despite the severe noise and blur in the image that we restore. The edge map is determined with the edge detector of `gimp`, a public domain software tool for image processing. □

6 Conclusion

We have presented a new approach to Tikhonov regularization with general square regularization operators based on the generalized Arnoldi process. The use of the latter allows us to avoid the solution of linear systems of equations, which have the regularization operator as matrix; it suffices to evaluate matrix-vector products with the regularization operator. Numerical examples illustrate the good performance of the proposed approach.

Acknowledgment. We would like to thank Giuseppe Rodriguez for insightful comments and a referee for remarks.

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] 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.

[4] D. Calvetti, P. C. Hansen, and L. Reichel, *L-curve curvature bounds via Lanczos bidiagonalization*, Electron. Trans. Numer. Anal., 14 (2002), pp. 20–35.

[5] D. Calvetti and L. Reichel, *Tikhonov regularization of large linear problems*, BIT, 43 (2003), pp. 263–283.

[6] D. Calvetti, L. Reichel, and A. Shuibi, *Invertible smoothing preconditioners for linear discrete ill-posed problems*, Appl. Numer. Math., 54 (2005), pp. 135–149.

[7] 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.

[8] M. Donatelli, C. Estatico, A. Martinelli, and S. Serra-Capizzano, *Improved image deblurring with anti-reflective boundary conditions and re-blurring*, Inverse Problems, 22 (2006), pp. 2035–2053.

[9] L. Eldén, *A weighted pseudoinverse, generalized singular values, and constrained least squares problems*, BIT, 22 (1982), pp. 487–501.

[10] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.

[11] C. W. Groetsch, *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman, Boston, 1984.

[12] 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.

[13] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.

[14] P. C. Hansen, *Regularization tools version 4.0 for Matlab 7.3*, Numer. Algorithms, 46 (2007), pp. 189–194.

[15] 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.

[16] B. Lewis and L. Reichel, *Arnoldi-Tikhonov regularization methods*, J. Comput. Appl. Math., 226 (2009), pp. 92–102.

[17] 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.

[18] S. Morigi, L. Reichel, and F. Sgallari, *Orthogonal projection regularization operators*, Numer. Algorithms, 44 (2007), pp. 99–114.

[19] 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.

[20] L. Reichel and A. Shyshkov, *A new zero-finder for Tikhonov regularization*, BIT, 48 (2008), pp. 627–643.

[21] L. Reichel and Q. Ye, *Simple square smoothing regularization operators*, Electron. Trans. Numer. Anal., 33 (2009), pp. 63–83.

[22] 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.