NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 0000; **00**:1–18
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nla

Square regularization matrices for large linear discrete ill-posed problems

Marco Donatelli[1], Arthur Neuman[2], Lothar Reichel[2*]

[1] *Dipartimento di Scienza e Alta Tecnologia, Università dell'Insubria, Via Valleggio 11, 22100 Como, Italy.*
[2] *Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA.*

SUMMARY

Large linear discrete ill-posed problems with contaminated data are often solved with the aid of Tikhonov regularization. Commonly used regularization matrices are finite difference approximations of a suitable derivative and are rectangular. This paper discusses the design of square regularization matrices that can be used in iterative methods based on the Arnoldi process for large-scale Tikhonov regularization problems. Copyright © 0000 John Wiley & Sons, Ltd.

Received ...

KEY WORDS: Tikhonov regularization; regularization matrix; Arnoldi process

1. INTRODUCTION

This paper is concerned with the computation of approximate solutions of large linear systems of equations

$$Ax = b \tag{1}$$

with a matrix $A \in \mathbb{R}^{n\times n}$ whose singular values "cluster" at the origin. In particular, A is severely ill-conditioned and may be singular. Linear systems of equations (1) with a matrix of this kind are commonly referred to as linear discrete ill-posed problems. They arise from the discretization of linear ill-posed problems, but also may originate in discrete form, for instance, in image restoration problems. The right-hand side $b \in \mathbb{R}^n$ represents available data and is assumed to be contaminated by an error $e \in \mathbb{R}^n$ caused by measurement inaccuracies, i.e.,

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

where $\hat{b}$ denotes the unavailable error-free vector associated with b. We will sometimes refer to the error e as "noise." The unknown error-free system

$$Ax = \hat{b} \tag{3}$$

is assumed to be consistent; we denote its solution of minimal Euclidean norm by $\hat{x}$. Our task is to determine an approximation of $\hat{x}$ by computing an approximate solution of (1). Due to the severe

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

Contract/grant sponsor: MIUR (for MD) and NSF (for LR); contract/grant number: 20083KLJEZ and DMS 111538.

ill-conditioning of A and the presence of the error $\boldsymbol{e}$ in $\boldsymbol{b}$, straightforward solution of (1) typically does not furnish a meaningful approximation of $\hat{\boldsymbol{x}}$.

The Tikhonov regularization method is one of the most popular approaches to determine an approximation of $\hat{\boldsymbol{x}}$. This method replaces the linear system of equations (1) by a penalized least-squares problem of the form

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

Here and throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm or the associated induced matrix norm. The matrix $L\in\mathbb{R}^{k\times n}$, $1\le k\le n$, is referred to as the regularization matrix and the scalar $\mu>0$ as the regularization parameter; see, e.g., Engl et al. [1] or Hansen [2] for discussions on Tikhonov regularization. We assume that

$$\mathcal{N}(A)\cap\mathcal{N}(L)=\{\boldsymbol{0}\}, \tag{5}$$

where $\mathcal{N}(M)$ denotes the null space of the matrix M. Then (4) has the unique solution

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

for any $\mu>0$, where the superscript ^{T} denotes transposition.

We are interested in the construction of square regularization matrices $L\in\mathbb{R}^{n\times n}$ that are convenient to use in iterative methods based on the Arnoldi process for the solution of large-scale minimization problems (4). The regularization matrices described also can be applied when $A\in\mathbb{R}^{m\times n}$ with $m\ne n$. When $m>n$, equation (1) should be considered a least-squares problem.

The value of the regularization parameter μ determines how sensitive $\boldsymbol{x}_\mu$ is to the error $\boldsymbol{e}$ in $\boldsymbol{b}$ and how close $\boldsymbol{x}_\mu$ is to $\hat{\boldsymbol{x}}$. A suitable value of μ generally is not known a priori, but has to be computed during the solution process. We assume a bound

$$\|\boldsymbol{e}\|\le\varepsilon$$

to be available. Then $\mu=\mu(\varepsilon)$ can be calculated with the help of the discrepancy principle, i.e., we determine $\mu(\varepsilon)>0$ so that

$$\|A\boldsymbol{x}_{\mu(\varepsilon)}-\boldsymbol{b}\|=\eta\varepsilon, \tag{7}$$

where $\eta>1$ is a user-supplied constant that is independent of ε. One can show that

$$\lim_{\varepsilon\searrow 0}\boldsymbol{x}_{\mu(\varepsilon)}=\hat{\boldsymbol{x}};$$

see [1] for a proof in a Hilbert space setting. Thus, our computational task is to determine both a value $\mu=\mu(\varepsilon)$, such that $\boldsymbol{x}_{\mu(\varepsilon)}$ given by (6) satisfies (7), and an approximation of the vector $\boldsymbol{x}_{\mu(\varepsilon)}$.

Common choices of regularization matrices are the identity matrix and scaled rectangular finite difference matrices, such as

$$L=\frac{1}{2}\begin{bmatrix} 1 & -1 & & & & 0 \\ & 1 & -1 & & & \\ & & 1 & -1 & & \\ & & & \ddots & \ddots & \\ 0 & & & & 1 & -1 \end{bmatrix}\in\mathbb{R}^{(n-1)\times n} \tag{8}$$

and

$$L=\frac{1}{4}\begin{bmatrix} -1 & 2 & -1 & & & 0 \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ 0 & & & -1 & 2 & -1 \end{bmatrix}\in\mathbb{R}^{(n-2)\times n}. \tag{9}$$

The regularization matrices (8) and (9) are associated with finite difference approximations of the first and second derivatives in one space-dimension, respectively. Many illustrations of iterative methods applied to Tikhonov regularization with $L = I$ can be found in, e.g., [3, 4, 5, 6]. A suitable choice of regularization matrix may enhance the quality of the computed approximation of $\hat{\boldsymbol{x}}$ significantly; see [7, 8, 9, 10, 11] for examples and discussions, as well as Section 4.

The null space of regularization matrices is important, because vectors in the null space are not damped in (4). The regularization matrices (8) and (9) have the null spaces

$$\mathcal{N}(L) = \operatorname{span}\{[1, 1, \ldots, 1]^T\}$$

and

$$\mathcal{N}(L) = \operatorname{span}\{[1, 1, \ldots, 1]^T, [1, 2, \ldots, n]^T\},$$

respectively. Generally, one seeks to use regularization matrices with a null space that avoids damping of known important features of the desired solution $\hat{\boldsymbol{x}}$.

Tikhonov regularization problems (4) with $L = I$ are said to be in *standard form*, while problems with $L \neq I$ are said to be in *general form*. Large-scale problems in standard form have the advantage over large-scale problems in general form that they can be reduced to smaller size by application of a few steps of a Krylov subspace method, such as the Arnoldi process or Lanczos bidiagonalization, independently of the value of μ. Tikhonov regularization problems in general form can be transformed into standard form. However, the regularization matrix L has to be of particular simple structure to make it feasible to evaluate the matrix-vector products required when applying Krylov subspace methods to the transformed problem.

It is the purpose of the present paper to discuss the construction of novel square regularization matrices L that allow fairly inexpensive computation of the matrix-vector products required for the approximate solution of the transformed Tikhonov minimization problem in standard form. Since both A and L are square, the matrix in the standard-form minimization problem will be square. This allows application of Arnoldi-type iterative methods, which may require fewer matrix-vector product evaluations than the commonly used Lanczos bidiagonalization method; see, e.g., [12] for recent illustrations. When A is rectangular, iterative methods based on Lanczos bidiagonalization can be applied together with the square regularization matrices of the present paper. Our work is based on results by Björck [13] and Eldén [14], who discuss iterative solution of large-scale Tikhonov minimization problem by methods based on Lanczos bidiagonalization and the transformation of these problems to standard form. More recently, Hansen and Jensen [9] described a novel approach to apply rectangular regularization matrices L with the Arnoldi process. Their approach is interesting, but does not allow evaluation of the norm of the residual error associated with each iterate without explicit computation of the residual error, which requires an extra matrix-vector product evaluation with the matrix A. Knowledge of the norm of the residual error is important when the regularization parameter μ is determined by the discrepancy principle or by the L-curve method. Other constructions of square regularization matrices are presented in [11].

We remark that small to medium-sized minimization problems (4) can be solved conveniently by first computing the generalized singular value decomposition (GSVD) of the matrix pair $\{A, L\}$. This paper is concerned with the development of solution methods for large-scale minimization problems, for which the computation of the GSVD of $\{A, L\}$ is too expensive to be attractive. We also note that certain large-scale problems benefit from the use of a regularization matrix that has a structure that does not allow efficient transformation to standard form. Special iterative methods have been developed for this situation; see, e.g., [15, 16, 17]. The present paper is interested in the case when the regularization matrix has an exploitable structure.

This paper is organized as follows. Section 2 reviews the approach of Eldén [14] to transform Tikhonov minimization problems in general form (4) to standard form. This transformation uses the A-weighted pseudoinverse, $L_A^\dagger$, associated with the matrix pair $\{A, L\}$. For general regularization matrices L, the A-weighted pseudoinverse $L_A^\dagger$ is unattractive to use in iterative methods. We describe new ways to construct square regularization matrices L with useful matrices $L_A^\dagger$. Section 3 discusses how large-scale minimization problems (4) in standard form can be reduced to small problems by a few steps of the range restricted Arnoldi-type process introduced in [18]. The implementation discussed in [18, 19] is superior to the one applied in [12]. Computed examples are presented in Section 4. In these examples both the number of steps of a Krylov subspace method and

DOI: 10.1002/nla

the value of the regularization parameter are determined with the aid of the discrepancy principle. We note, however, that the regularization matrices of the present paper also can be applied when the number of iterations and the value of the regularization parameter are determined by other techniques, such as by the L-curve or generalized cross validation; see, e.g., [2] for discussions on these techniques. Concluding remarks are found in Section 5.

2. THE CONSTRUCTION OF SQUARE REGULARIZATION MATRICES

We first summarize the method described by Eldén [14] for transforming the Tikhonov minimization problem in general form (4) to standard form. Consider the regularization matrix $L \in \mathbb{R}^{k\times n}$, with $1 \leq k \leq n$, and let $L^\dagger$ denote its Moore-Penrose pseudoinverse. Then the A-weighted pseudoinverse of L, introduced by Eldén [14], is given by

$$L_A^\dagger = \left(I - (A(I-L^\dagger L))^\dagger A\right) L^\dagger \in \mathbb{R}^{n\times k}. \tag{10}$$

Define the vectors

$$\begin{aligned} \bar{\boldsymbol{x}} &= L\boldsymbol{x}, \\ \boldsymbol{x}^{(0)} &= \left(A(I-L^\dagger L)\right)^\dagger \boldsymbol{b}, \\ \bar{\boldsymbol{b}} &= \boldsymbol{b} - A\boldsymbol{x}^{(0)}. \end{aligned}$$

Eldén [14] showed that the Tikhonov regularization problem (4) is equivalent to the minimization problem in standard form,

$$\min_{\bar{\boldsymbol{x}}\in\mathbb{R}^k} \{\|AL_A^\dagger \bar{\boldsymbol{x}} - \bar{\boldsymbol{b}}\|^2 + \mu\|\bar{\boldsymbol{x}}\|^2\}. \tag{11}$$

The solution $\boldsymbol{x}_\mu$ of (4), given by (6), can be recovered from the solution $\bar{\boldsymbol{x}}_\mu$ of (11) according to

$$\boldsymbol{x}_\mu = L_A^\dagger \bar{\boldsymbol{x}}_\mu + \boldsymbol{x}^{(0)}. \tag{12}$$

We note for future reference that

$$\|AL_A^\dagger \bar{\boldsymbol{x}}_\mu - \bar{\boldsymbol{b}}\| = \|A\boldsymbol{x}_\mu - \boldsymbol{b}\|. \tag{13}$$

In order to be able to solve (11) by an iterative method based on the Arnoldi process, the matrix $AL_A^\dagger$ has to be square. This requires $L_A^\dagger$, and therefore also $L^\dagger$, to be square. For this reason, we are interested in the construction square regularization matrices L.

Let the columns of $W \in \mathbb{R}^{n\times \ell}$ form an orthonormal basis for $\mathcal{N}(L)$ and introduce the orthogonal projector $P_{\mathcal{N}(L)}$ onto $\mathcal{N}(L)$. Then

$$I - L^\dagger L = P_{\mathcal{N}(L)} = WW^T.$$

The dimension ℓ of the null space of regularization matrices L of interest to us is fairly small, say, $\ell \leq 4$. Therefore, it is quite inexpensive to compute the QR factorization

$$AW = UR, \tag{14}$$

where $U \in \mathbb{R}^{n\times\ell}$ has orthonormal columns and $R \in \mathbb{R}^{\ell\times\ell}$ is upper triangular. It follows from (5) that R is nonsingular. Therefore the matrix $A(I-L^\dagger L) = URW^T$ has the Moore-Penrose pseudoinverse

$$(A(I-L^\dagger L))^\dagger = WR^{-1}U^T. \tag{15}$$

This representation can be established, e.g., by observing that the matrix in the right-hand side satisfies Penrose's four conditions for a pseudoinverse; see, e.g., [20, Theorem 1.2.11] for details.

DOI: 10.1002/nla

Substituting (15) into (10) yields

$$\begin{aligned} L_A^\dagger &= (I - WR^{-1}U^TA)L^\dagger, \\ AL_A^\dagger &= (I - UU^T)AL^\dagger. \end{aligned} \tag{16}$$

We conclude that matrix-vector products with the matrices $L_A^\dagger$ and $AL_A^\dagger$ can be evaluated quite inexpensively provided that

(i) $\mathcal{N}(L)$ is of small dimension ℓ and has an explicitly known basis,

(ii) the restriction of A to $\mathcal{N}(L)$ is not very ill-conditioned, and

(iii) matrix-vector products with the Moore-Penrose pseudoinverse $L^\dagger$ are inexpensive to compute.

The remainder of this section describes new ways to construct square regularization matrices with these properties. The null spaces chosen in the numerical examples of Section 4 are made up of discretizations of linear or slowly oscillating functions. The restrictions of the matrices A of interest to these null spaces are fairly well conditioned.

Let $\widetilde{L} \in \mathbb{R}^{n\times n}$ be a square nonsingular matrix that allows rapid solution of linear systems of equations with this matrix, e.g.,

$$\widetilde{L} = \frac{1}{2}\begin{bmatrix} 1 & -1 & & & & \mathbf{0} \\ & 1 & -1 & & & \\ & & 1 & -1 & & \\ & & & \ddots & \ddots & \\ & & & & 1 & -1 \\ \mathbf{0} & & & & & 1 \end{bmatrix} \in \mathbb{R}^{n\times n} \tag{17}$$

or

$$\widetilde{L} = \frac{1}{4}\begin{bmatrix} 2 & -1 & & & & \mathbf{0} \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ \mathbf{0} & & & & -1 & 2 \end{bmatrix} \in \mathbb{R}^{n\times n}. \tag{18}$$

These matrices are obtained by appending and prepending suitable rows to the rectangular regularization matrices (8) or (9). Linear systems of equations with the matrix (17) can be solved by back substitution in only $\mathcal{O}(n)$ arithmetic floating point operations. Also, linear systems of equations with the matrix (18) can be solved in only $\mathcal{O}(n)$ arithmetic floating point operations with the aid of Cholesky factorization.

However, the matrices (17) and (18) are not very useful as regularization matrices, because they typically can be applied successfully only if the desired solution $\hat{\boldsymbol{x}}$ has suitable boundary behavior; see [8] for illustrations. For instance, the regularization matrix (18) should be used only when the first and last components of the desired solution $\hat{\boldsymbol{x}}$ are close to zero. We now will describe modifications of the matrices (17) and (18) that give useful square regularization matrices.

Let $W \in \mathbb{R}^{n\times\ell}$ have orthonormal columns and introduce the square regularization matrix

$$L = \widetilde{L}(I - WW^T), \tag{19}$$

where $\widetilde{L}$ is defined by (17) or (18). Then $\mathcal{N}(L) = \mathcal{R}(W)$, where $\mathcal{R}(W)$ denotes the range of W. The regularization matrix (19) can be applied successfully to a much wider class of problems than the regularization matrices (17) or (18). The following theorem yields an expression for the pseudoinverse of regularization matrices of the form (19).

DOI: 10.1002/nla

Theorem 1
Let $\widetilde{L} \in \mathbb{R}^{n\times n}$ be nonsingular and let $W \in \mathbb{R}^{n\times \ell}$ have orthonormal columns. Then the matrix (19) defined by $\widetilde{L}$ and W has the Moore-Penrose pseudoinverse

$$L^{\dagger} = (I - WW^T)\widetilde{L}^{-1}(I - QQ^T), \tag{20}$$

where the columns of $Q \in \mathbb{R}^{n\times \ell}$ form an orthonormal basis for $\mathcal{N}((I - WW^T)\widetilde{L}^T)$. Moreover,

$$AL_A^{\dagger} = (I - UU^T)A\widetilde{L}^{-1}(I - QQ^T), \tag{21}$$

where the matrix $U \in \mathbb{R}^{n\times \ell}$ is given by (14).

Proof
Let $\boldsymbol{d} \in \mathbb{R}^n$ be arbitrary and consider the least-squares problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|L\boldsymbol{x} - \boldsymbol{d}\|.$$

The minimal-norm solution is given by $L^{\dagger}\boldsymbol{d}$. It also can be determined as the minimal-norm solution of the consistent linear system of equations

$$\widetilde{L}(I - WW^T)\boldsymbol{x} = P_{\mathcal{R}(\widetilde{L}(I-WW^T))}\boldsymbol{d}. \tag{22}$$

The orthogonal projector in the right-hand side can be conveniently represented by observing that $\mathcal{R}(\widetilde{L}(I - WW^T))$ is orthogonal to $\mathcal{N}((I - WW^T)\widetilde{L}^T)$. The latter space is of (low) dimension ℓ; the columns of the matrix $\widetilde{L}^{-T}W$ form a basis for this space. Let the matrix $Q \in \mathbb{R}^{n\times \ell}$ be obtained by orthogonalizing the columns of $\widetilde{L}^{-T}W$. Then

$$P_{\mathcal{R}(\widetilde{L}(I-WW^T))} = I - QQ^T.$$

The vector $\widetilde{L}^{-1}(I - QQ^T)\boldsymbol{d}$ is a solution of (22). The solution of minimal norm is orthogonal to $\mathcal{N}(L)$; it is given by $(I - WW^T)\widetilde{L}^{-1}(I - QQ^T)\boldsymbol{d}$. This shows (20).

The expression (21) follows by substituting (20) into (16) and the observation that $(I - UU^T)A(I - WW^T) = (I - UU^T)A$. □

A simple way to obtain a square regularization matrix from a rectangular matrix $L \in \mathbb{R}^{k\times n}$, $k < n$, such as (8) or (9), is to append or prepend $n - k$ rows of zeros. This approach has been used in [11]. The following corollary yields an expression for the pseudoinverse. A similar result is discussed by Björck [13, Section 6].

Corollary 1
Let $L \in \mathbb{R}^{k\times n}$ with $1 \le k < n$ be of full rank. Define the matrix $\widetilde{L} \in \mathbb{R}^{n\times n}$ by enlarging L by $n - k$ rows, so that $\widetilde{L}$ is nonsingular. We generally will append or prepend the new rows to L. Let the matrix $\widetilde{L}_0 \in \mathbb{R}^{n\times n}$ be obtained by setting the entries in the new rows of $\widetilde{L}$ to zero. Then

$$\widetilde{L}_0^{\dagger} = P_{\mathcal{N}(L)}^{\perp}\widetilde{L}^{-1}. \tag{23}$$

Proof
The result can be shown similarly as Theorem 1, with the additional observation that the projector $P_{\mathcal{R}(\widetilde{L}_0)}$ is not required. The latter follows from the fact that components of $P_{\mathcal{R}(\widetilde{L}_0)}\boldsymbol{v}$ in nonvanishing rows of $\widetilde{L}_0$ are the same as those of $\boldsymbol{v}$ for all $\boldsymbol{v} \in \mathbb{R}^n$. □

Example 2.1 Let the regularization matrix L be defined by (9) and let $\widetilde{L}$ be the square extension (18). We can rapidly evaluate expressions of the form $\widetilde{L}^{-1}\boldsymbol{x}$ for $\boldsymbol{x} \in \mathbb{R}^n$ by using the Cholesky factorization of $\widetilde{L}$. Matrix-vector products with the Moore-Penrose pseudoinverse (23) therefore are simple to compute when an orthonormal basis for $\mathcal{N}(L)$ is available. □

DOI: 10.1002/nla

Example 2.2 The regularization matrix (9) also can be extended to a square matrix by only appending rows. For instance, we may use the upper triangular matrix

$$\widetilde{L} = \frac{1}{4}\begin{bmatrix} -1 & 2 & -1 & & & & \mathbf{0} \\ & -1 & 2 & -1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \\ \mathbf{0} & & & & & -1 \end{bmatrix} \in \mathbb{R}^{n\times n}. \tag{24}$$

The expression $\widetilde{L}^{-1}\boldsymbol{x}$, $\boldsymbol{x} \in \mathbb{R}^n$, can be evaluated by back substitution. □

Let L and $\widetilde{L}$ be the matrices of Corollary 1. The condition number of the matrix L, denoted by $\kappa(L)$, is the quotient of the largest and smallest singular values of L. It follows, e.g. from [21, Cor. 8.6.3], that

$$\kappa(L) \le \kappa(\widetilde{L}).$$

The size of $\kappa(\widetilde{L})$ depends on how $\widetilde{L}$ is constructed from L.

Example 2.3 Let $L \in \mathbb{R}^{100\times 98}$ be defined by (9). Then $\kappa(L) = 1.8 \cdot 10^3$. The condition number of the symmetric extension (18) is $4.1 \cdot 10^3$, and the condition number of the upper triangular extension (24) is $1.2 \cdot 10^4$. □

The following theorem shows that rows can be appended or prepended to obtain a square matrix from a rectangular one so that the condition number does not increase.

Theorem 2
Let the conditions of Corollary 1 hold and let the orthonormal columns of the matrix $W \in \mathbb{R}^{n\times(n-k)}$ form a basis for $\mathcal{N}(L)$. Denote the largest and smallest singular values of L by σ_1 and σ_k, respectively, and let s be a constant such that $\sigma_k \le s \le \sigma_1$. Define the matrix $\widetilde{L}$ by appending or prepending the matrix sW^T to L. Then

$$\kappa(L) = \kappa(\widetilde{L}).$$

Proof
Introduce the singular value decomposition $L = \breve{U}\breve{\Sigma}\breve{V}^T$, where the matrices $\breve{U} \in \mathbb{R}^{k\times k}$ and $\breve{V} \in \mathbb{R}^{n\times n}$ are orthogonal, and $\breve{\Sigma} = [\,\Sigma\,|\,0\,] \in \mathbb{R}^{k\times n}$ with

$$\Sigma = \mathrm{diag}[\sigma_1, \sigma_2, \ldots, \sigma_k] \in \mathbb{R}^{k\times k}, \qquad \sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_k > 0.$$

Then $\breve{V} = [\,V\,|\,W\,]$, where $V \in \mathbb{R}^{n\times k}$, and $L = \breve{U}\Sigma V^T$. The singular value decomposition of $\widetilde{L}$ obtained by appending sW^T to L is given by

$$\widetilde{L} = \begin{bmatrix} L \\ sW^T \end{bmatrix} = \begin{bmatrix} \breve{U} & 0 \\ 0 & I_{n-k} \end{bmatrix}\begin{bmatrix} \Sigma & 0 \\ 0 & sI_{n-k} \end{bmatrix}\breve{V}^T$$

up to a permutation of the singular values. Similarly, the singular value decomposition of $\widetilde{L}$ obtained by prepending sW^T to L is given by

$$\widetilde{L} = \begin{bmatrix} sW^T \\ L \end{bmatrix} = \begin{bmatrix} I_{n-k} & 0 \\ 0 & \breve{U} \end{bmatrix}\begin{bmatrix} sI_{n-k} & 0 \\ 0 & \Sigma \end{bmatrix}\begin{bmatrix} W^T \\ V^T \end{bmatrix}$$

up to a permutation of the singular values. The bounds on s secure that $\kappa(\widetilde{L}) = \sigma_1/\sigma_k$. This completes the proof. □

Thus, appending or prepending the matrix W^T, suitably scaled, to L produces a square nonsingular matrix $\widetilde{L}$ with the same condition number as L. Linear systems of equations with the

DOI: 10.1002/nla

matrix $\widetilde{L}$ can be solved efficiently when a QR factorization of L^T can be determined inexpensively. This is the case in the important situation when L has small bandwidth, such as the when L is given by (8) or (9). Let $L^T = \hat{Q}\hat{R}$ be a QR factorization, where $\hat{Q} \in \mathbb{R}^{n\times k}$ has orthonormal columns and $\hat{R} \in \mathbb{R}^{k\times k}$ is upper triangular. Then

$$\widetilde{L}^T = \left[\begin{array}{cc} W & \hat{Q} \end{array}\right] \left[\begin{array}{cc} sI_{n-k} & 0 \\ 0 & \hat{R} \end{array}\right]. \tag{25}$$

Linear systems of equations with the matrix $\widetilde{L}$ can be solved rapidly by using the factorization (25). The solution of linear systems of equations with a matrix $\widetilde{L}$ obtained by appending sW^T to L can be solved similarly.

3. THE RANGE RESTRICTED ARNOLDI PROCESS

This section discusses the approximate solution of the Tikhonov minimization problem in standard form (11). We outline the application of the range restricted Arnoldi process described in [18, 19]. This method, when applied to the matrix

$$\bar{A} = A\widetilde{L}_A^{\dagger} \tag{26}$$

with initial vector $\bar{\boldsymbol{b}}$, yields after p steps the decomposition

$$\bar{A}V_p = W_{p+2}H_{p+2,p}, \tag{27}$$

where the columns of $W_{p+2} \in \mathbb{R}^{n\times(p+2)}$ form an orthonormal basis for the Krylov subspace

$$\mathcal{K}_{p+2}(\bar{A}, \bar{\boldsymbol{b}}) = \mathrm{span}\{\bar{\boldsymbol{b}}, \bar{A}\bar{\boldsymbol{b}}, \ldots, \bar{A}^{p+1}\bar{\boldsymbol{b}}\}$$

and the columns of $V_p \in \mathbb{R}^{n\times p}$ form an orthonormal basis for the Krylov subspace $\mathcal{K}_p(\bar{A}, \bar{A}\bar{\boldsymbol{b}})$. Moreover, $W_{p+2}\boldsymbol{e}_1 = \bar{\boldsymbol{b}}/\|\bar{\boldsymbol{b}}\|$, where $\boldsymbol{e}_1 = [1, 0, \ldots, 0]^T$ denotes the first axis vector. The matrix $H_{p+2,p} \in \mathbb{R}^{(p+2)\times p}$ vanishes below the sub-subdiagonal. The dominating computational work for determining the decomposition (27) for large matrices $\bar{A}$ is the evaluation of $p+1$ matrix-vector products with $\bar{A}$. Note that the matrix (26) is not explicitly formed; only matrix-vector products with this matrix are evaluated.

Substituting $\bar{\boldsymbol{x}} = V_p\bar{\boldsymbol{y}}$, with V_p defined by (27), into (11) yields the reduced Tikhonov regularization problem

$$\min_{\bar{\boldsymbol{y}}\in\mathbb{R}^p} \left\| \left[\begin{array}{c} H_{p+2,p} \\ \mu^{1/2}I \end{array}\right] \bar{\boldsymbol{y}} - \boldsymbol{e}_1\|\bar{\boldsymbol{b}}\| \right\|.$$

Its solution, $\bar{\boldsymbol{y}}_{\mu,p}$, can be computed by QR factorization of the matrix. Then $\bar{\boldsymbol{x}}_{\mu,p} = V_p\bar{\boldsymbol{y}}_{\mu,p}$ satisfies

$$\bar{\boldsymbol{x}}_{\mu,p} = \operatorname*{argmin}_{\bar{\boldsymbol{x}}\in\mathcal{K}_p(\bar{A},\bar{A}\bar{\boldsymbol{b}})} \{\|\bar{A}\bar{\boldsymbol{x}} - \bar{\boldsymbol{b}}\|^2 + \mu\|\bar{\boldsymbol{x}}\|^2\}, \tag{28}$$

and the corresponding approximate solution of (4) is given by

$$\boldsymbol{x}_{\mu,p} = \widetilde{L}_A^{\dagger}\bar{\boldsymbol{x}}_{\mu,p} + \boldsymbol{x}^{(0)}; \tag{29}$$

cf. (12).

Proposition 1

The function

$$\phi_p(\mu) = \|\bar{\boldsymbol{b}}\|^2\boldsymbol{e}_1^T(\mu^{-1}H_{p+2,p}H_{p+2,p}^T + I)^{-2}\boldsymbol{e}_1 \tag{30}$$

is strictly increasing and satisfies

$$\phi_p(\mu) = \|A\boldsymbol{x}_{\mu,p} - \boldsymbol{b}\|^2. \tag{31}$$

DOI: 10.1002/nla

Moreover, the equation

$$\phi_p(\mu) = \tau \tag{32}$$

has a unique solution $0 < \mu < \infty$ for any τ with

$$\|P_{\mathcal{N}(H^T_{p+2,p})}\boldsymbol{e}_1\|^2\|\bar{\boldsymbol{b}}\|^2 < \tau < \|\bar{\boldsymbol{b}}\|^2, \tag{33}$$

where $P_{\mathcal{N}(H^T_{p+2,p})}$ denotes the orthogonal projector onto the null space of $H^T_{p+2,p}$.

Proof
The representation (30) shows that ϕ_p is increasing as well as the bounds (33). By [5, Section 2], we have

$$\phi_p(\mu) = \|A\widetilde{L}_A^\dagger\bar{\boldsymbol{x}}_{\mu,p} - \bar{\boldsymbol{b}}\|^2,$$

and (31) now is a consequence of (13). □

Corollary 2

$$\|P_{\mathcal{N}(H^T_{p+3,p+1})}\boldsymbol{e}_1\| \leq \|P_{\mathcal{N}(H^T_{p+2,p})}\boldsymbol{e}_1\|. \tag{34}$$

Proof
It follows from (30) and (33) that

$$\lim_{\mu\searrow 0}\phi_p(\mu) = \|P_{\mathcal{N}(H^T_{p+2,p})}\boldsymbol{e}_1\|^2\|\bar{\boldsymbol{b}}\|^2.$$

Moreover, by (28) and (31), we have

$$\phi_p(0) = \|\bar{A}\bar{\boldsymbol{x}}_{0,p} - \bar{\boldsymbol{b}}\|^2 = \min_{\bar{\boldsymbol{x}}\in\mathcal{K}_p(\bar{A},\bar{A}\bar{\boldsymbol{b}})}\|\bar{A}\bar{\boldsymbol{x}} - \bar{\boldsymbol{b}}\|^2.$$

Now $\mathcal{K}_p(\bar{A},\bar{A}\bar{\boldsymbol{b}}) \subset \mathcal{K}_{p+1}(\bar{A},\bar{A}\bar{\boldsymbol{b}})$ yields $\phi_{p+1}(0) \leq \phi_p(0)$. This shows (34). □

We have to choose p large enough in order for the computed approximate solution (29) of (4) to be able to satisfy the discrepancy principle

$$\|A\boldsymbol{x}_{\mu,p} - \boldsymbol{b}\| = \eta\varepsilon; \tag{35}$$

cf. (7). Typically, a fairly small value of p suffices; see Section 4. We let $p_{\min}$ denote the smallest integer p such that

$$\lim_{\mu\searrow 0}\phi_p(\mu) < \eta^2\varepsilon^2.$$

It follows from (30) that the function $\mu \to \phi_p(1/\mu)$ is decreasing and convex for $0 < \mu < \infty$. It therefore may be beneficial to solve

$$\phi_p(1/\mu) = \eta^2\varepsilon^2 \tag{36}$$

instead of (32) with $\tau = \eta^2\varepsilon^2$. However, for large-scale problems the computational effort required to determine a value of μ so that (35) holds is negligible compared with the work needed for evaluating matrix-vector products. We therefore will not dwell on the computation of μ further.

We remark that large-scale Tikhonov regularization problems in standard form (11) also can be reduced to smaller problems by other Krylov subspace methods, such as by Lanczos bidiagonalization. A comparison of reductions by the range restricted Arnoldi process (though based on a different implementation than the one outlined above), by Lanczos bidiagonalization, and by the standard Arnoldi process is reported in [12]. Computed examples in [12] show reduction by the range restricted Arnoldi process to require fewer matrix-vector product evaluations than reduction by Lanczos bidiagonalization, and to give higher accuracy than reduction by the standard Arnoldi process.

DOI: 10.1002/nla

4. COMPUTED EXAMPLES

We illustrate the performance of the regularization matrices discussed with some numerical 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

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

Here $\hat{\boldsymbol{b}}$ is the noise-free right-hand side vector in (3). We let $\eta = 1.01$ in (7) in all examples. The computations are carried out in MATLAB with about 16 significant decimal digits.

We reduce the Tikhonov minimization problem to standard form (11) with the range restricted Arnoldi process described in Section 3 and report results for $p = p_{\min}$ and $p = p_{\min} + 1$ steps, where $p_{\min}$ is the smallest number of steps, such that the discrepancy principle (35) can be satisfied. The regularization parameter μ is computed by applying Newton's method to the solution of equation (36). Details about the implementation of Newton's method with a function similar to $\phi_p(1/\mu)$ can be found in [12]; see also [17] for a discussion on zero-finders.

This section compares the performance of several square regularization matrices. Let L_1 and L_2 be the finite difference matrices (8) and (9), respectively. More generally, $L_q \in \mathbb{R}^{(n-q)\times n}$ denotes the finite difference matrix associated with a finite difference approximation with equidistant nodes of the qth derivative in one space-dimension. Then

$$\mathcal{N}(L_q) = \text{span}\{\boldsymbol{n}_0, \boldsymbol{n}_1, \ldots, \boldsymbol{n}_{q-1}\}, \quad \boldsymbol{n}_i = [1^i, 2^i, \ldots, n^i]^T.$$

As described in Section 2, the matrix L_q can be extended to a square matrix $\widetilde{L}_q$ by appending or prepending q suitable rows. We will append or prepend rows to L_q so that $\widetilde{L}_q$ is Toeplitz. The matrices $\widetilde{L}_q$ so obtained are invertible. In particular, $\widetilde{L}_1$ and $\widetilde{L}_2$ are given by (17) and (18), respectively. The matrices $\widetilde{L}_{q,0}$ are obtained by appending q zero rows to L_q; they are of the type considered in Corollary 1. We denote the matrix (19) by $L(\widetilde{L}, W)$, and the matrix $\widetilde{L}$ in Theorem 2 by

$$\widetilde{L}_j^{(\text{post})} = \begin{bmatrix} L_j \\ sW^T \end{bmatrix} \quad \text{or} \quad \widetilde{L}_j^{(\text{pre})} = \begin{bmatrix} sW^T \\ L_j \end{bmatrix},$$

where $\mathcal{R}(W) = \mathcal{N}(L_j)$ and $s \approx \sigma_{n/2}(L_j)$.

Our comparison also includes the square regularization matrix $(\hat{C}_2^\dagger D_\delta^{-1})^\dagger \in \mathbb{R}^{n\times n}$ introduced in [11]. Here $\hat{C}_2^\dagger$ is the pseudoinverse of the circulant matrix associated with the finite difference approximation of the second derivative, with the three smallest eigenvalues set to zero. Further,

$$D_\delta = \text{diag}[\delta, 1, 1, \ldots, 1, \delta] \in \mathbb{R}^{n\times n}, \quad \delta = 1 \cdot 10^{-8}.$$

The regularization matrix $(\hat{C}_2^\dagger D_\delta^{-1})^\dagger$ is designed not to damp slow oscillations and linear growth in the computed approximate solution very much; see [11] for details.

Example 4.1 We consider the Fredholm integral equation of the first kind

$$\int_0^1 k(s,t)x(t)dt = \exp(s) + (1-\mathrm{e})s - 1, \quad 0 \le s \le 1, \tag{37}$$

where

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

This equation is discussed, e.g., by Delves and Mohamed [22]. We discretize the integral equation by a Galerkin method with orthonormal box functions as test and trial functions using the MATLAB program deriv2 from Regularization Tools [23]. The program yields a symmetric indefinite matrix

DOI: 10.1002/nla

Table I. Example 4.1: Errors in approximate solutions computed by Tikhonov regularization based on the range restricted Arnoldi process for the problem deriv2 with exponential solution, $\nu = 1 \cdot 10^{-3}$, $W = \mathrm{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3]$, and $p = p_{\min}$.

Regularization matrix	p	$\lVert \boldsymbol{x}_{\mu,p} - \hat{\boldsymbol{x}} \rVert$	$\lVert \boldsymbol{x}_{\mu,p+1} - \hat{\boldsymbol{x}} \rVert$
I	8	$1.8683 \cdot 10^{-1}$	$1.9080 \cdot 10^{-1}$
$\widetilde{L}_{2,0}$	1	$3.4009 \cdot 10^{-3}$	$3.3780 \cdot 10^{-3}$
$\widetilde{L}_{3,0}$	1	$3.1915 \cdot 10^{-3}$	$2.6732 \cdot 10^{-3}$
$(\hat{C}_2^{\dagger} D_{\delta}^{-1})^{\dagger}$	2	$3.1354 \cdot 10^{-3}$	$2.9775 \cdot 10^{-3}$
$L(I, W)$	1	$2.4255 \cdot 10^{-3}$	$7.8445 \cdot 10^{-3}$
$L(\widetilde{L}_2, W)$	1	$2.2524 \cdot 10^{-3}$	$2.6032 \cdot 10^{-3}$
$L(\widetilde{L}_3, W)$	1	$7.1758 \cdot 10^{-4}$	$2.5515 \cdot 10^{-3}$
$\widetilde{L}_2^{(\mathrm{post})}$	1	$3.4009 \cdot 10^{-3}$	$3.3780 \cdot 10^{-3}$
$\widetilde{L}_3^{(\mathrm{post})}$	1	$3.1915 \cdot 10^{-3}$	$2.6732 \cdot 10^{-3}$

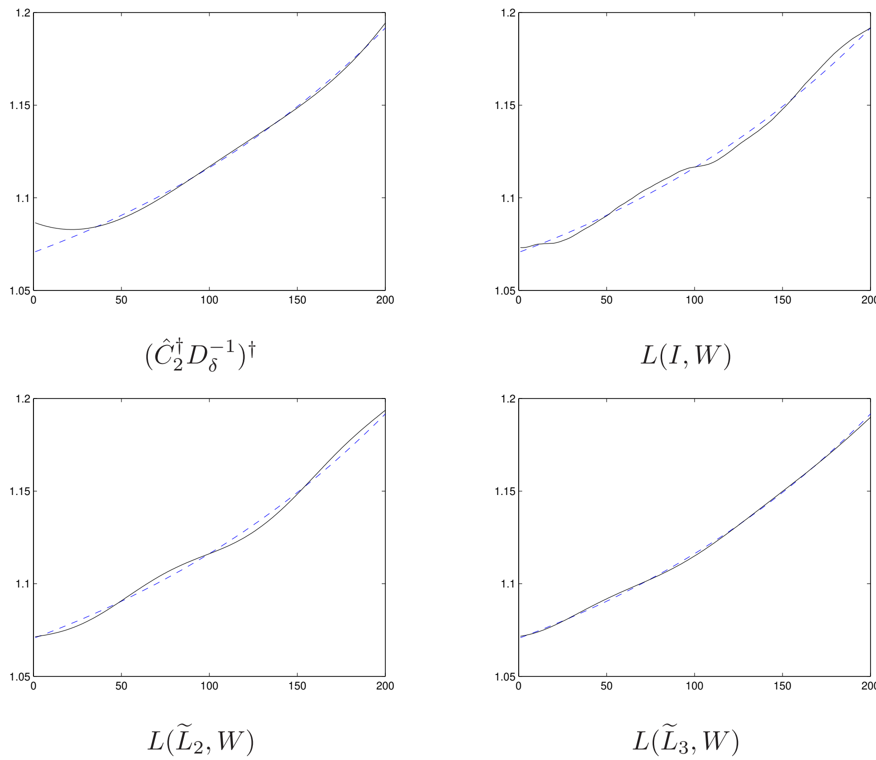


Figure 1. Example 4.1: Approximate solutions $\boldsymbol{x}_{\mu,p_{\min}}$ (continuous graphs) computed by Tikhonov regularization based on the range restricted Arnoldi process for the problem deriv2 with exponential solution, $\nu = 1 \cdot 10^{-3}$, and $W = \mathrm{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3]$. The dashed graphs depict the desired solution $\hat{\boldsymbol{x}}$.

$A \in \mathbb{R}^{200 \times 200}$ and a scaled discrete approximation $\hat{\boldsymbol{x}} \in \mathbb{R}^{200}$ of the solution $x(t) = \exp(t)$ of (37). The error-free right-hand side vector is given by $\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}$, and the right-hand side vector $\boldsymbol{b}$ in (1) is obtained by (2). The noise-level is $\nu = 1 \cdot 10^{-3}$.

Table I reports results determined by Tikhonov regularization based on the range restricted Arnoldi process for several square regularization matrices. The number of Arnoldi steps, $p = p_{\min}$, is the smallest number that allows (36) to be satisfied. The norm of the error in the computed Tikhonov solution, $\boldsymbol{x}_{\mu,p}$, is shown in the penultimate column. The last column shows the error in $\boldsymbol{x}_{\mu,p+1}$ obtained by carrying out one more step of the range restricted Arnoldi process and then determining a value of μ that satisfies (36) with p replaced by $p + 1$. We observe that the quality of

DOI: 10.1002/nla

Table II. Example 4.2: Errors in approximate solutions computed by Tikhonov regularization based on the range restricted Arnoldi process for the problem deriv2 with exponential solution, $\nu = 1 \cdot 10^{-5}$, $W = \mathrm{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3]$, and $p = p_{\min}$.

Regularization matrix	p	$\lVert \boldsymbol{x}_{\mu,p} - \hat{\boldsymbol{x}} \rVert$	$\lVert \boldsymbol{x}_{\mu,p+1} - \hat{\boldsymbol{x}} \rVert$
I	22	$9.7162 \cdot 10^{-2}$	$9.7744 \cdot 10^{-2}$
$\widetilde{L}_{2,0}$	10	$1.8805 \cdot 10^{-3}$	$1.9045 \cdot 10^{-3}$
$\widetilde{L}_{3,0}$	4	$3.0066 \cdot 10^{-4}$	$3.0063 \cdot 10^{-4}$
$(\hat{C}_2^{\dagger} D_{\delta}^{-1})^{\dagger}$	2	$2.2457 \cdot 10^{-4}$	$2.2341 \cdot 10^{-4}$
$L(I, W)$	2	$2.9875 \cdot 10^{-4}$	$3.0685 \cdot 10^{-4}$
$L(\widetilde{L}_2, W)$	3	$2.9072 \cdot 10^{-4}$	$2.9077 \cdot 10^{-4}$
$L(\widetilde{L}_3, W)$	6	$2.7453 \cdot 10^{-4}$	$2.6909 \cdot 10^{-4}$
$\widetilde{L}_2^{(\mathrm{pre})}$	10	$1.8071 \cdot 10^{-3}$	$1.8305 \cdot 10^{-3}$
$\widetilde{L}_3^{(\mathrm{post})}$	4	$3.0066 \cdot 10^{-4}$	$3.0063 \cdot 10^{-4}$

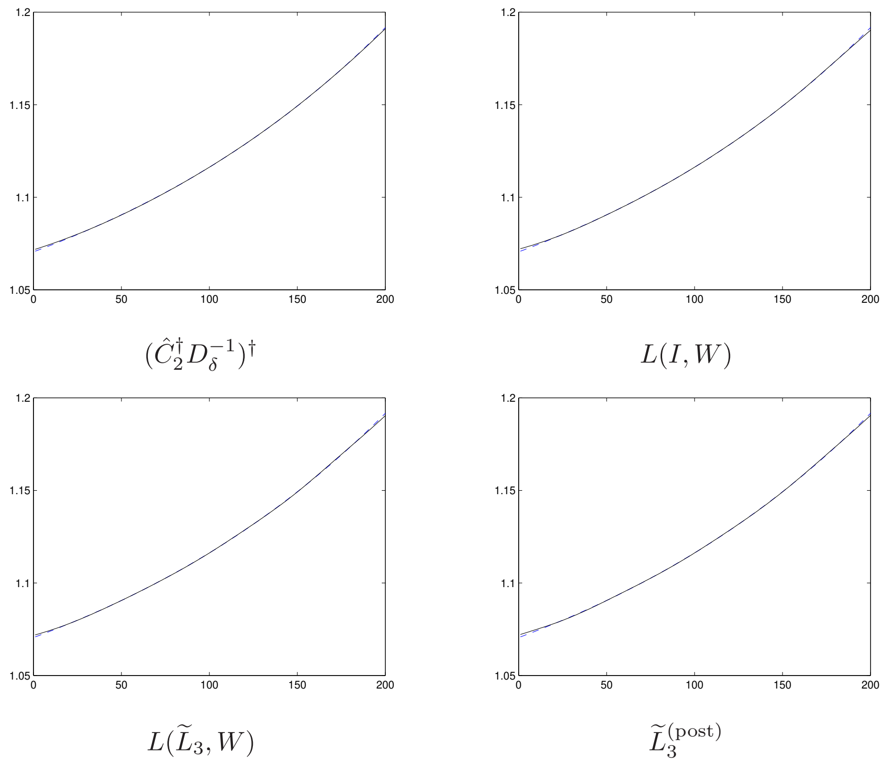


Figure 2. Example 4.2: Approximate solutions $\boldsymbol{x}_{\mu,p_{\min}}$ (continuous graphs) computed by Tikhonov regularization based on the range restricted Arnoldi process for the problem deriv2 with exponential solution, $\nu = 1 \cdot 10^{-5}$, and $W = \mathrm{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3]$. The dashed graphs depict the desired solution $\hat{\boldsymbol{x}}$.

the computed approximation of the desired solution $\hat{\boldsymbol{x}}$ generally does not improve significantly by taking more than the minimal number of steps, $p = p_{\min}$, required to satisfy (36).

Figure 1 displays $\hat{\boldsymbol{x}}$ and the most accurate computed approximate solutions. The best approximation of $\hat{\boldsymbol{x}}$ is obtained with $L(\widetilde{L}_3, W)$, where $W = \mathrm{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3]$ denotes a matrix with orthonormal columns that span $\mathcal{R}([\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3])$. $\Box$

Example 4.2 We consider the test problem of Example 4.1 with less noise in $\boldsymbol{b}$; the noise-level is $\nu = 1 \cdot 10^{-5}$. The reduced error has the effect that more steps, $p = p_{\min}$, of the range restricted Arnoldi process have to be carried out in order to be able to satisfy (36) than in Example 4.1.

DOI: 10.1002/nla

Table III. Example 4.3: Errors in approximate solutions computed by Tikhonov regularization based on the range restricted Arnoldi process for the problem phillips with linear and exponential components added to the solution, $\nu = 1 \cdot 10^{-2}$, $W = \text{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2]$, and $p = p_{\min}$.

Regularization matrix	p	$\lVert \boldsymbol{x}_{\mu,p} - \hat{\boldsymbol{x}} \rVert$	$\lVert \boldsymbol{x}_{\mu,p+1} - \hat{\boldsymbol{x}} \rVert$
I	7	$4.1097 \cdot 10^{-2}$	$3.0391 \cdot 10^{-2}$
$\widetilde{L}_{2,0}$	3	$6.5757 \cdot 10^{-3}$	$6.6729 \cdot 10^{-3}$
$(\hat{C}_2^{\dagger} D_{\delta}^{-1})^{\dagger}$	3	$1.4137 \cdot 10^{-2}$	$1.5416 \cdot 10^{-2}$
$L(I, W)$	2	$5.6219 \cdot 10^{-3}$	$5.7454 \cdot 10^{-3}$
$L(\widetilde{L}_1, W)$	5	$5.2157 \cdot 10^{-3}$	$5.7061 \cdot 10^{-3}$
$L(\widetilde{L}_2, W)$	3	$6.3756 \cdot 10^{-3}$	$6.4003 \cdot 10^{-3}$
$\widetilde{L}_2^{(\text{pre})}$	3	$6.4576 \cdot 10^{-3}$	$6.6729 \cdot 10^{-3}$

Table II shows the regularization matrix $(\hat{C}_2^{\dagger} D_{\delta}^{-1})^{\dagger}$ to give the best approximation of $\hat{\boldsymbol{x}}$, though many of the regularization matrices yield about the same accuracy. Figure 2 displays approximate solutions computed with several regularization matrices (continuous graphs). These graphs are hardly distinguishable from those for $\hat{\boldsymbol{x}}$ (dashed graphs).

We remark that efficient evaluation of a matrix-vector product with the regularization matrix $(\hat{C}_2^{\dagger} D_{\delta}^{-1})^{\dagger}$ requires the use of the fast Fourier transform and, therefore, $\mathcal{O}(n \log(n))$ arithmetic floating point operations, while the evaluation of matrix-vector products with the other regularization matrices of Tables I and II requires only $\mathcal{O}(n)$ arithmetic floating point operations. Among the latter regularization matrices, $L(\widetilde{L}_3, W)$ with $W = \text{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3]$ yields the best approximations of $\hat{\boldsymbol{x}}$ in Table II. This regularization matrix performs well for many noise-levels ν. $\Box$

Example 4.3 Consider the Fredholm integral equation of the first kind

$$\int_{-6}^{6} \kappa(\tau, \sigma) x(\sigma) d\sigma = g(\tau), \qquad -6 \leq \tau \leq 6, \tag{38}$$

with kernel and solution given by

$$\begin{aligned} \kappa(\tau, \sigma) &= x(\tau - \sigma), \\ x(\sigma) &= \begin{cases} 1 + \cos(\frac{\pi}{3}\sigma), & \text{if } |\sigma| < 3, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

The right-hand side $g(\tau)$ is defined by (38). This integral equation is discussed by Phillips [24]. The MATLAB code phillips in [23] determines a discretization by a Galerkin method with orthonormal box functions. A discretization of a scaled solution, $\boldsymbol{x}_0$, also is provided. The matrix $A \in \mathbb{R}^{500 \times 500}$ determined by phillips represents the discretized integral operator. We add a discretization of the function $1 + \exp(\frac{1}{12}(t + 6))$, $t \in [-6, 6]$, to the vector $\boldsymbol{x}_0$ to obtain a slowly oscillatory and increasing solution $\hat{\boldsymbol{x}}$. The noise-free right-hand side is given by $\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}$ and the noise level is $\nu = 1 \cdot 10^{-2}$.

Table III shows $L(\widetilde{L}_1, W)$ with $W = \text{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2]$ to give the smallest approximation error. However, also other square regularization matrices, such as $L_2^{(\text{pre})}$, yield quite accurate approximations of $\hat{\boldsymbol{x}}$; see Figure III. $\Box$

Example 4.4 The Fredholm integral equation of the first kind,

$$\int_{0}^{\pi/2} \kappa(\sigma, \tau) x(\sigma) d\sigma = g(\tau), \qquad 0 \leq \tau \leq \pi,$$

with $\kappa(\sigma, \tau) = \exp(\sigma \cos(\tau))$, $g(\tau) = 2\sinh(\tau)/\tau$, and solution $x(\sigma) = \sin(\sigma)$ is discussed by Baart [25]. We use the MATLAB code baart in [23] to determines a discretization by a Galerkin method

DOI: 10.1002/nla

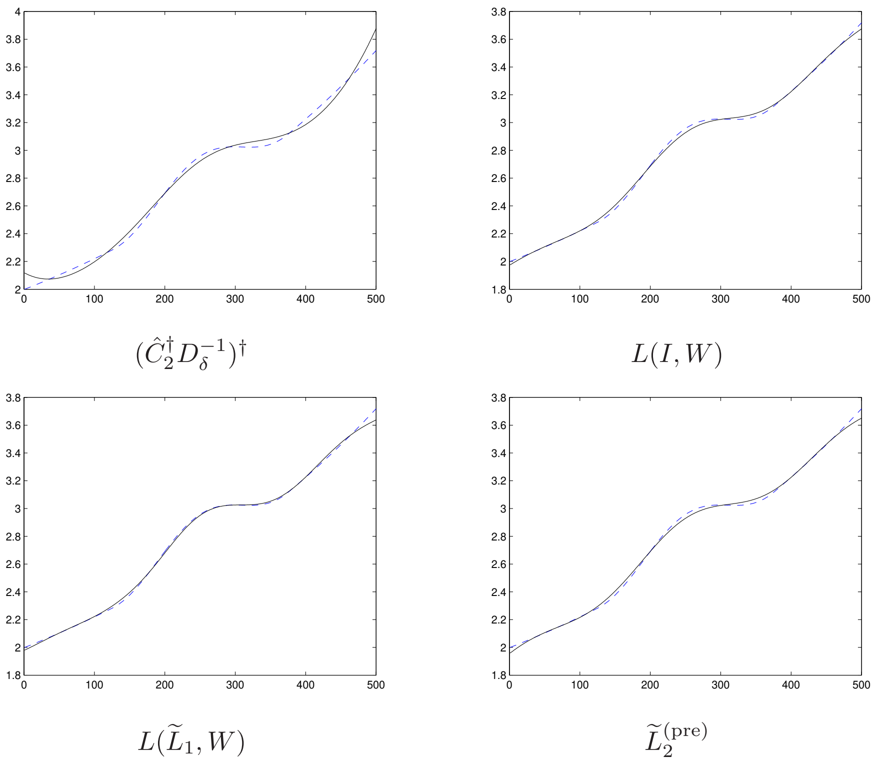


Figure 3. Example 4.3: Approximate solutions $\boldsymbol{x}_{\mu,p_{\min}}$ (continuous graphs) computed by Tikhonov regularization based on the range restricted Arnoldi process for the problem phillips with linear and exponential components added to the solution, $\nu = 1 \cdot 10^{-2}$, and $W = \mathrm{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2]$. The dashed graphs depict the desired solution $\hat{\boldsymbol{x}}$.

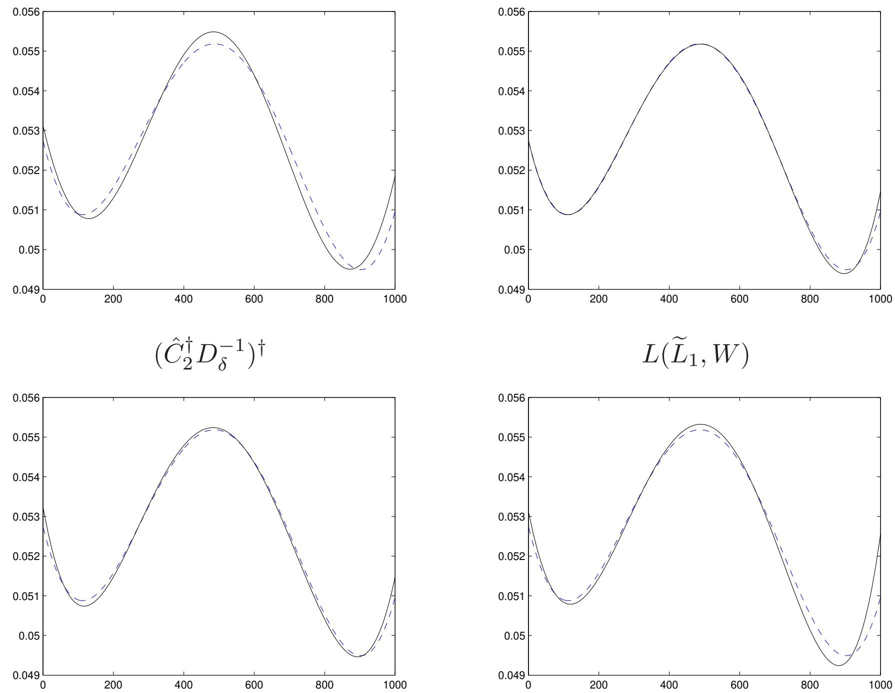
Table IV. Example 4.4: Errors in approximate solutions computed by Tikhonov regularization based on the range restricted Arnoldi process for the problem baart with a quadratic component added to the solution, $\nu = 3 \cdot 10^{-6}$, $W = \mathrm{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2]$, and $p = p_{\min}$.

Regularization matrix	p	$\Vert\boldsymbol{x}_{\mu,p} - \hat{\boldsymbol{x}}\Vert$	$\Vert\boldsymbol{x}_{\mu,p+1} - \hat{\boldsymbol{x}}\Vert$
I	5	$6.2541 \cdot 10^{-3}$	$6.2604 \cdot 10^{-3}$
$\widetilde{L}_{1,0}$	3	$5.5382 \cdot 10^{-3}$	$7.3538 \cdot 10^{-3}$
$\widetilde{L}_{2,0}$	3	$8.1223 \cdot 10^{-3}$	$8.2636 \cdot 10^{-3}$
$(\hat{C}_2^{\dagger} D_{\delta}^{-1})^{\dagger}$	2	$5.7755 \cdot 10^{-3}$	$5.8425 \cdot 10^{-3}$
$L(I, W)$	3	$6.5072 \cdot 10^{-3}$	$6.5873 \cdot 10^{-3}$
$L(\widetilde{L}_1, W)$	2	$1.4750 \cdot 10^{-3}$	$3.5012 \cdot 10^{-3}$
$L(\widetilde{L}_2, W)$	2	$2.3624 \cdot 10^{-3}$	$5.4850 \cdot 10^{-3}$
$\widetilde{L}_1^{(\mathrm{post})}$	3	$5.5382 \cdot 10^{-3}$	$7.3538 \cdot 10^{-3}$
$\widetilde{L}_2^{(\mathrm{post})}$	3	$8.1223 \cdot 10^{-3}$	$8.2636 \cdot 10^{-3}$

with 1000 orthonormal box functions. The code produces the matrix $A \in \mathbb{R}^{1000 \times 1000}$ and a scaled discrete approximation $\boldsymbol{x}_0$ of $x(\tau)$. We add a discretization of the function $\frac{1}{19}(\frac{16}{\pi^2}(t - \frac{\pi}{4})^2 - \frac{t}{15\pi}t)$, $t \in [0, \pi/2]$, to the vector $\boldsymbol{x}_0$ to obtain a slowly oscillatory solution $\hat{\boldsymbol{x}}$. The noise-free right-hand side is given by $\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}$ and the noise level is $\nu = 3 \cdot 10^{-6}$.

Table IV and Figure 4 show $L(\tilde{L}_1, W)$ with $W = \mathrm{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2]$ to yield the best approximation of $\hat{\boldsymbol{x}}$. We remark that this regularization matrix also performs well for larger noise levels. Fewer iterations are required when the right-hand side vector $\boldsymbol{b}$ is contaminated by more noise. $\square$

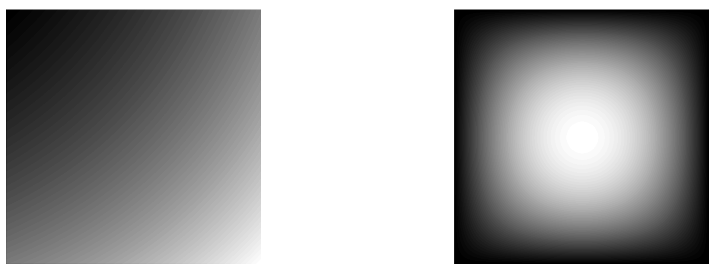
DOI: 10.1002/nla



$(\hat{C}_2^{\dagger} D_{\delta}^{-1})^{\dagger}$ $L(\widetilde{L}_1, W)$

$L(\widetilde{L}_2, W)$ $\widetilde{L}_1^{(\mathrm{post})}$

Figure 4. Example 4.4: Approximate solutions $\boldsymbol{x}_{\mu,p_{\min}}$ (continuous graphs) computed by Tikhonov regularization based on the range restricted Arnoldi process for the problem baart with a quadratic component added to the solution, $\nu = 3 \cdot 10^{-6}$, and $W = \mathrm{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2]$. The dashed graphs depict the desired solution $\hat{\boldsymbol{x}}$.



$\widehat{X}$ $\widehat{B}$

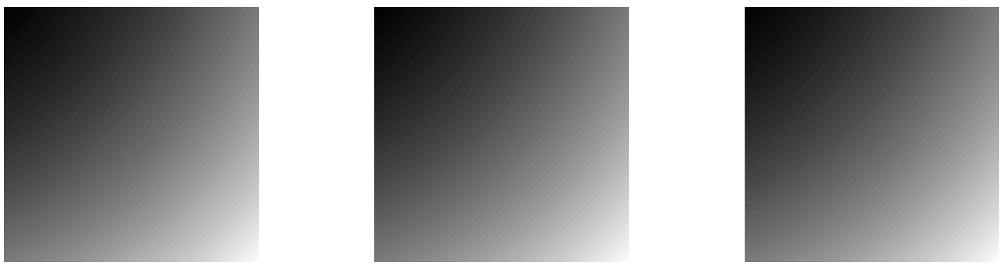
Figure 5. Example 4.5: The desired solution $\widehat{X}$ and the right-hand side of the matrix equation (39) obtained by from the MATLAB function from deriv2 with exponential solution; $\nu = 1 \cdot 10^{-4}$.

Example 4.5 We consider a version of the Fredholm integral equation of the first kind (37) in two space-dimensions. Let $A^{(1)} \in \mathbb{R}^{500\times 500}$ and $\hat{\boldsymbol{x}}^{(1)} \in \mathbb{R}^{500}$ be computed with the MATLAB program deriv2 from Regularization Tools [23] and define the matrix $A = A^{(1)} \otimes A^{(1)}$ and the error-free solution $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}^{(1)} \otimes \hat{\boldsymbol{x}}^{(1)}$, where $\otimes$ denotes tensor product. The error-free right-hand side vector is given by $\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}$ and the right-hand side vector $\boldsymbol{b}$ in (1) is obtained by (2). The noise-level is $\nu = 1 \cdot 10^{-4}$.

DOI: 10.1002/nla

Table V. Example 4.5: Errors $\|\boldsymbol{x}_{\mu,p} - \hat{\boldsymbol{x}}\|$ in approximate solutions $\boldsymbol{x}_{\mu,p}$ computed by Tikhonov regularization based on the range restricted Arnoldi process with $L = \left((L^{(1)})^\dagger \otimes I + I \otimes (L^{(1)})^\dagger\right)^\dagger$ for a problem in two space-dimensions obtained as tensor product of the problem deriv2 in one space-dimension with exponential solution, $\nu = 1 \cdot 10^{-4}$, $W^{(1)} = \mathrm{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3]$, and $p = p_{\min}$.

$L^{(1)}$	p	$\|\boldsymbol{x}_{\mu,p} - \hat{\boldsymbol{x}}\|$	$\|\boldsymbol{x}_{\mu,p+1} - \hat{\boldsymbol{x}}\|$
$\widetilde{L}_{2,0}$	10	$7.0405 \cdot 10^{-3}$	$7.0324 \cdot 10^{-3}$
$\widetilde{L}_{3,0}$	1	$4.5125 \cdot 10^{-4}$	$4.6314 \cdot 10^{-4}$
$(\hat{C}_2^\dagger D_\delta^{-1})^\dagger$	25	$1.5972 \cdot 10^{-2}$	$5.8027 \cdot 10^{-3}$
$L(I, W^{(1)})$	1	$3.8113 \cdot 10^{-4}$	$3.8112 \cdot 10^{-4}$
$L(\widetilde{L}_2, W^{(1)})$	1	$3.7949 \cdot 10^{-4}$	$3.7949 \cdot 10^{-4}$
$L(\widetilde{L}_3, W^{(1)})$	2	$3.5573 \cdot 10^{-4}$	$3.5980 \cdot 10^{-4}$
$\widetilde{L}_2^{(\mathrm{post})}$	10	$7.0405 \cdot 10^{-3}$	$7.0324 \cdot 10^{-3}$
$\widetilde{L}_3^{(\mathrm{post})}$	1	$4.5125 \cdot 10^{-4}$	$4.6314 \cdot 10^{-4}$



$L^{(1)} = \widetilde{L}_{3,0}$ $L^{(1)} = L(I, W^{(1)})$ $L^{(1)} = L(\widetilde{L}_3, W^{(1)})$

Figure 6. Example 4.5: Approximate solutions $X_{\mu,p_{\min}}$ computed by Tikhonov regularization based on the range restricted Arnoldi process with L defined by (40) for a problem in two space-dimensions obtained as a tensor product of the problem deriv2 in one space-dimension with exponential solution, $\nu = 1 \cdot 10^{-4}$, $W^{(1)} = \mathrm{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3]$.

$L^{(1)} = \widetilde{L}_{3,0}$ $L^{(1)} = L(I, W^{(1)})$ $^{(1)} = L(\widetilde{L}_3, W^{(1)})$

Figure 7. Example 4.5: Component-wise absolute error, $|\widehat{X} - X_{\mu,p_{\min}}|$, with $X_{\mu,p_{\min}}$ computed by Tikhonov regularization based on the range restricted Arnoldi process with L defined by (40) for a problem in two space-dimensions obtained as a tensor product of the problem deriv2 in one space-dimension with exponential solution, $\nu = 1 \cdot 10^{-4}$, $W^{(1)} = \mathrm{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3]$.

The vectors $\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}} \in \mathbb{R}^{500^2}$ can be stored in 500×500 matrices; in MATLAB notation

$$\widehat{X} = \mathsf{reshape}(\hat{\boldsymbol{x}}, [500,\ 500]), \qquad \widehat{B} = \mathsf{reshape}(\hat{\boldsymbol{b}}, [500,\ 500]).$$

Hence, the matrix-vector relation (3) is equivalent to the matrix-matrix relation

$$A^{(1)}\widehat{X}A^{(1)} = \widehat{B}.$$

Similarly, the system (1) can be written as

$$A^{(1)}XA^{(1)} = B. \tag{39}$$

The matrices $X, \widehat{X}, B, \widehat{B}$ represent discretizations of surfaces. Figure 5 shows the desired solution $\widehat{X}$ and the observed right-hand side B of (39). The graphs are scaled so that black is the minimum value and white the maximum value.

Consider the representation (1) of (39). We would like to choose a regularization matrix L in (4) such that matrix-vector products with $L^\dagger$ can be carried out efficiently. We therefore require $L^\dagger$ to have a tensor product structure similar to that of A. Then matrix-vector products with the matrix $L^\dagger$ can be evaluated similarly as in equation (39) for the matrix A. Note that it is not necessary that L have a tensor product structure.

Let $L^{(1)}$ be one of the regularization matrices considered in Example 4.1 for the analogous problem in one space-dimension. We remark that the "natural" regularization matrix $L = L^{(1)} \otimes I + I \otimes L^{(1)}$ is unattractive to use, because matrix-vector products with $L^\dagger$ are expensive to evaluate. Instead, we use the regularization matrix

$$L = \left((L^{(1)})^\dagger \otimes I + I \otimes (L^{(1)})^\dagger\right)^\dagger. \tag{40}$$

Then $L^\dagger \hat{\boldsymbol{x}}$ can be evaluated fairly efficiently as $(L^{(1)})^\dagger \widehat{X} + \widehat{X}(L^{(1)})^\dagger$.

The evaluation of matrix-vector products with the matrix (16) requires that an orthonormal basis for $\mathcal{N}(L)$ be explicitly known. Let the columns of $W^{(1)}$ form an orthonormal basis for $\mathcal{N}(L^{(1)})$. Then the columns of $W = W^{(1)} \otimes W^{(1)}$ form an orthonormal basis for $\mathcal{N}(L)$. Moreover, the QR factorization $A^{(1)}W^{(1)} = U^{(1)}R^{(1)}$, cf. (14), yields the QR factorization $AW = (U^{(1)} \otimes U^{(1)})(R^{(1)} \otimes R^{(1)})$. The other computations are similar to those of Example 4.1.

Table V shows $L^{(1)} = L(\widetilde{L}_3, W^{(1)})$ with $W^{(1)} = \mathrm{orth}[\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3]$ to give the smallest approximation error. The choice $L^{(1)} = (\hat{C}_2^\dagger D_\delta^{-1})^\dagger$ does not seem to be robust enough for this example. Figure 6 shows some of the best restorations and Figure 7 displays the absolute error of each solution component. The differences in the computed solutions is clearly visible. □

5. CONCLUSION

The numerical examples of the previous section show square regularization matrices constructed as described by Theorems 1 and 2, as well as by Corollary 1, to perform well. Their ease of application and the possibility of choosing the null space makes them attractive to use in a variety of situations. In particular, they also may be useful for Tikhonov regularization problems (4) with a rectangular matrix A.

REFERENCES

1. Engl HW, Hanke M, Neubauer A. *Regularization of Inverse Problems*. Kluwer: Dordrecht, 1996.
2. Hansen PC. *Rank-Deficient and Discrete Ill-Posed Problems*. SIAM: Philadelphia, 1998.
3. Calvetti D, Golub GH, Reichel L. Estimation of the L-curve via Lanczos bidiagonalization. *BIT* 1999; **39**:603–619.
4. Hanke M, Hansen PC. Regularization methods for large-scale problems. *Surveys on Mathematics for Industry* 1993; **3**:253–315.
5. Hochstenbach ME, McNinch N, Reichel L. Discrete ill-posed least-squares problems with a solution norm constraint. *Linear Algebra and Its Applications*, in press.
6. Reichel L, Sadok H, Shyshkov A. Greedy Tikhonov regularization for large linear ill-posed problems. *International Journal of Computer Mathematics* 2007; **84**:1151–1166.
7. Brezinski C, Redivo-Zaglia M, Rodriguez G, Seatzu S. Multi-parameter regularization techniques for ill-conditioned linear systems. *Numerische Mathematik* 2003; **94**:203–228.

DOI: 10.1002/nla

8. Calvetti D, Reichel L, Shuibi A. Invertible smoothing preconditioners for linear discrete ill-posed problems. *Applied Numerical Mathematics* 2005; **54**:135–149.
9. Hansen PC, Jensen TK. Smoothing norm preconditioning for regularizing minimum residual methods. *SIAM Journal on Matrix Analysis and Applcations* 2006; **29**:1–14.
10. Morigi S, Reichel L, Sgallari F. Orthogonal projection regularization operators. *Numerical Algorithms* 2007; **44**:99–114.
11. Reichel L, Ye Q. Simple square smoothing regularization operators. *Electronic Transactions on Numerical Analysis* 2009; **33**:63–83.
12. Lewis B, Reichel L. Arnoldi-Tikhonov regularization methods. *Journal of Computational and Applied Mathematics* 2009; **226**:92–102.
13. Björck Å. A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations. *BIT* 1988; **18**:659–670.
14. Eldén L. A weighted pseudoinverse, generalized singular values, and constrained least squares problems. *BIT* 1982; **22**:487–501.
15. Hochstenbach ME, Reichel L. An iterative method for Tikhonov regularization with a general linear regularization operator. *Journal of Integral Equations and Applications* 2010; **22**:463–480.
16. Kilmer ME, Hansen PC, Español MI. A projection-based approach to general-form Tikhonov regularization. *SIAM Journal on Scientific Computing* 2007; **29**:315–330.
17. Lampe J, Reichel L, Voss H. Large-scale Tikhonov regularization via reduction by orthogonal projection. *Linear Algebra and Its Applications*, in press.
18. Neuman A, Reichel L, Sadok H. Implementations of range restricted iterative methods for linear discrete ill-posed problems. *Linear Algebra and Its Applications*, in press.
19. Neuman A, Reichel L, Sadok H. Algorithms for range restricted iterative methods for linear discrete ill-posed problems. *Numerical Algorithms* 2012; **59**:325–331.
20. Björck Å. *Numerical Methods for Least Squares Problems*. SIAM: Philadelphia, 1996.
21. Golub GH, Van Loan CF. *Matrix Computations*. Johns Hopkins University Press: Baltimore, 1996.
22. Delves LM, Mohamed JL. *Computational Methods for Integral Equations*. Cambridge University Press: Cambridge, 1985.
23. Hansen PC. Regularization tools version 4.0 for MATLAB 7.3. *Numerical Algorithms* 2007; **46**:189–194.
24. Phillips DL. A technique for the numerical solution of certain integral equations of the first kind. *Journal of the ACM* 1962; **9**:84–97.
25. Baart ML. The use of auto-correlation for pseudo-rank determination in noisy ill-conditioned least-squares problems. *IMA Journal of Numerical Analysis* 1982; **2**:241–247.

DOI: 10.1002/nla