REGULARIZATION MATRICES DETERMINED BY MATRIX NEARNESS PROBLEMS

GUANGXIN HUANG*, SILVIA NOSCHESE†, AND LOTHAR REICHEL‡

Abstract. This paper is concerned with the solution of large-scale linear discrete ill-posed problems with error-contaminated data. Tikhonov regularization is a popular approach to determine meaningful approximate solutions of such problems. The choice of regularization matrix in Tikhonov regularization may significantly affect the quality of the computed approximate solution. This matrix should be chosen to promote the recovery of known important features of the desired solution, such as smoothness and monotonicity. We describe a novel approach to determine regularization matrices with desired properties by solving a matrix nearness problem. The constructed regularization matrix is the closest matrix in the Frobenius norm with a prescribed null space to a given matrix. Numerical examples illustrate the performance of the regularization matrices so obtained.

Key words. Tikhonov regularization, regularization matrix, matrix nearness problem.

AMS subject classifications. 65R30, 65F22, 65F10.

1. Introduction. We are concerned with the computation of an approximate solution of linear least-squares problems of the form

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|K\boldsymbol{x}-\boldsymbol{b}\|, \qquad K\in\mathbb{R}^{m\times n}, \quad \boldsymbol{b}\in\mathbb{R}^m, \tag{1.1}$$

with a large matrix K with many singular values of different orders of magnitude close to the origin. In particular, K is severely ill-conditioned and may be singular. Linear least-squares problems with a matrix of this kind often 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 vector $\boldsymbol{b}$ of linear discrete ill-posed problems that arise in applications typically represents measured data that is contaminated by an unknown error $\boldsymbol{e}\in\mathbb{R}^m$.

Let $\widehat{\boldsymbol{b}}\in\mathbb{R}^m$ denote the unknown error-free vector associated with $\boldsymbol{b}$, i.e.,

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

and let $\widehat{\boldsymbol{x}}$ be the solution of the unavailable linear system of equations

$$K\boldsymbol{x}=\widehat{\boldsymbol{b}}, \tag{1.3}$$

which we assume to be consistent. If K is singular, then $\widehat{\boldsymbol{x}}$ denotes the solution of minimal Euclidean norm.

Let $K^\dagger$ denote the Moore–Penrose pseudoinverse of K. The solution of minimal Euclidean norm of (1.1), given by

$$K^\dagger\boldsymbol{b}=K^\dagger\widehat{\boldsymbol{b}}+K^\dagger\boldsymbol{e}=\widehat{\boldsymbol{x}}+K^\dagger\boldsymbol{e},$$

*Geomathematics Key Laboratory of Sichuan, Chengdu University of Technology, Chengdu, 610059, P. R. China. Email: huangx@cdut.edu.cn. Research supported by a grant from CSC, Research Fund Project (NS2014PY08) of SUSE and the young scientific research backbone teachers of CDUT (KYGG201309).

†SAPIENZA Università di Roma, P.le A. Moro, 2, I-00185 Roma, Italy. E-mail: noschese@mat.uniroma1.it. Research supported by a grant from SAPIENZA Università di Roma.

‡Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. E-mail: reichel@math.kent.edu. Research supported in part by NSF grant DMS-1115385.

typically is not a useful approximation of $\widehat{\boldsymbol{x}}$ due to severe propagation of the error $\boldsymbol{e}$. This depends on the large norm of $K^\dagger$. Therefore, one generally replaces the least-squares problem (1.1) by a nearby problem, whose solution is less sensitive to the error $\boldsymbol{e}$. This replacement is known as regularization. One of the most popular regularization methods is due to Tikhonov. This method replaces (1.1) by a penalized least-squares problem of the form

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

where $L \in \mathbb{R}^{p\times n}$ is referred to as a regularization matrix and the scalar $\mu > 0$ as a regularization parameter; see, e.g., [1, 9, 11]. Throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm or the spectral matrix norm. We assume that the matrices K and L satisfy

$$\mathcal{N}(K) \cap \mathcal{N}(L) = \{\mathbf{0}\}, \tag{1.5}$$

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

$$\boldsymbol{x}_\mu = (K^TK + \mu L^TL)^{-1}K^T\boldsymbol{b}$$

for any $\mu > 0$. The superscript ^{T} denotes transposition. When L is the identity matrix, the Tikhonov minimization problem (1.4) is said to be in *standard form*, otherwise it is in *general form*. We are interested in minimization problems (1.4) in general form.

The value of $\mu > 0$ in (1.4) determines how sensitive $\boldsymbol{x}_\mu$ is to the error $\boldsymbol{e}$, how close $\boldsymbol{x}_\mu$ is to the desired solution $\widehat{\boldsymbol{x}}$, and how small the residual error $\boldsymbol{b} - K\boldsymbol{x}_\mu$ is. A suitable value of μ generally is not explicitly known and has to be determined during the solution process.

Minimization problems (1.4) in general form with matrices K and L of small to moderate size can be conveniently solved with the aid of the Generalized Singular Value Decomposition (GSVD) of the matrix pair $\{K, L\}$; see, e.g., [7, 11]. We are interested in developing solution methods for large-scale minimization problems (1.4). These problems have to be solved by an iterative method. However, the regularization matrices L derived also may be useful for problems of small size.

Common choices of regularization matrices L in (1.4) when the least-squares problem (1.1) is obtained by discretizing a Fredholm integral equation of the first kind in one space-dimension are the $n \times n$ identity matrix I_n, and scaled finite difference approximations of the first derivative operator,

$$L_1 = \frac{1}{2}\begin{bmatrix} 1 & -1 & & & & 0 \\ & 1 & -1 & & & \\ & & 1 & -1 & & \\ & & & \ddots & \ddots & \\ 0 & & & & 1 & -1 \end{bmatrix} \in \mathbb{R}^{(n-1)\times n}, \tag{1.6}$$

as well as of the second derivative operator,

$$L_2 = \frac{1}{4}\begin{bmatrix} -1 & 2 & -1 & & & 0 \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ 0 & & & -1 & 2 & -1 \end{bmatrix} \in \mathbb{R}^{(n-2)\times n}. \tag{1.7}$$

The null spaces of these matrices are

$$\mathcal{N}(L_1) = \text{span}\{[1, 1, \ldots, 1]^T\} \tag{1.8}$$

and

$$\mathcal{N}(L_2) = \text{span}\{[1, 1, \ldots, 1]^T, [1, 2, \ldots, n]^T\}. \tag{1.9}$$

The regularization matrices L_1 and L_2 damp fast oscillatory components of the solution $\boldsymbol{x}_\mu$ of (1.4) more than slowly oscillatory components. This can be seen by comparing Fourier coefficients of the vectors $\boldsymbol{x}$, $L_1\boldsymbol{x}$, and $L_2\boldsymbol{x}$; see, e.g., [21]. These matrices therefore are referred to as *smoothing regularization matrices*. Here we think of the vector $\boldsymbol{x}_\mu$ as a discretization of a continuous real-valued function. The use of a smoothing regularization matrix can be beneficial when the desired solution $\widehat{\boldsymbol{x}}$ is a discretization of a smooth function.

The regularization matrix L in (1.4) should be chosen so that known important features of the desired solution $\widehat{\boldsymbol{x}}$ of (1.3) can be represented by vectors in $\mathcal{N}(L)$, because these vectors are not damped by L. For instance, if the solution is known to be the discretization at equidistant points of a smooth monotonically increasing function, then it may be appropriate to use the regularization matrix (1.7), because its null space contains the discretization of linear functions. Several approaches to construct regularization matrices with desirable properties are described in the literature; see, e.g., [4, 5, 6, 13, 16, 18, 21]. Many of these approaches are designed to yield square modifications of the matrices (1.6) and (1.7) that can be applied in conjunction with iterative solution methods based on the Arnoldi process. We will discuss the Arnoldi process more below.

The present paper describes a new approach to the construction of square regularization matrices. It is based on determining the closest matrix with a prescribed null space to a given square nonsingular matrix. For instance, the given matrix may be defined by appending a suitable row to the finite difference matrix (1.6) to make the matrix nonsingular, and then prescribing a null space, say, (1.8) or (1.9). The distance between matrices is measured with the Frobenius norm,

$$\|A\|_F := \sqrt{\langle A, A\rangle}, \qquad A \in \mathbb{R}^{p\times n},$$

where the inner product between matrices is defined by

$$\langle A, B\rangle := \text{Trace}(B^T A), \qquad A, B \in \mathbb{R}^{p\times n}.$$

Our reason for using the Frobenius norm is that the solution of the matrix nearness problem considered in this paper can be determined with fairly little computations in this setting.

We remark that commonly used regularization matrices in the literature, such as (1.6) and (1.7), are rectangular. Our interest in square regularization matrices stems from the fact that they allow the solution of (1.4) by iterative methods that are based on the Arnoldi process. Application of the Arnoldi process to the solution of Tikhonov minimization problems (1.4) was first described in [2]; a recent survey is provided by Gazzola et al. [10]. We are interested in being able to use iterative solution methods that are based on the Arnoldi process because they only require the computation of matrix-vector products with the matrix A and, therefore, typically require fewer matrix-vector product evaluations than methods that demand the computation of

matrix-vector products with both A and A^T, such as methods based on Golub–Kahan bidiagonalization; see, e.g., [15] for an example.

This paper is organized as follows. Section 2 discusses matrix nearness problems of interest in the construction of the regularization matrices. The application of regularization matrices obtained by solving these nearness problems is discussed in Section 3. Krylov subspace methods for the computation of an approximate solution of (1.4), and therefore of (1.1), are reviewed in Section 4, and a few computed examples are presented in Section 5. Concluding remarks can be found in Section 6.

2. Matrix nearness problems. This section investigates the distance of a matrix to the closest matrix with a prescribed null space. For instance, we are interested in the distance of the invertible square bidiagonal matrix

$$L_{1,\delta} = \frac{1}{2}\begin{bmatrix} 1 & -1 & & & & 0 \\ & 1 & -1 & & & \\ & & 1 & -1 & & \\ & & & \ddots & \ddots & \\ & & & & 1 & -1 \\ 0 & & & & 0 & \delta \end{bmatrix} \in \mathbb{R}^{n\times n} \tag{2.1}$$

with $\delta > 0$ to the closest matrix with the same null space as the rectangular matrix (1.6). Regularization matrices of the form (2.1) with $\delta > 0$ small have been considered in [4]; see also [13] for a discussion.

Square regularization matrices have the advantage over rectangular ones that they can be used together with iterative methods based on the Arnoldi process for Tikhonov regularization [2, 10] as well as in GMRES-type methods [17]. These applications have spurred the development of a variety of square regularization matrices. For instance, it has been proposed in [21] that a zero row be appended to the matrix (1.6) to obtain the square regularization matrix

$$L_{1,0} = \frac{1}{2}\begin{bmatrix} 1 & -1 & & & & 0 \\ & 1 & -1 & & & \\ & & 1 & -1 & & \\ & & & \ddots & \ddots & \\ & & & & 1 & -1 \\ 0 & & & & 0 & 0 \end{bmatrix} \in \mathbb{R}^{n\times n}$$

with the same null space. Among the questions that we are interested in is whether there is a square regularization matrix that is closer to the matrix (2.1) than $L_{1,0}$ and has the same null space as the latter matrix. Throughout this paper $\mathcal{R}(A)$ denotes the range of the matrix A.

PROPOSITION 2.1. *Let the matrix $V \in \mathbb{R}^{n\times\ell}$ have $1 \le \ell < n$ orthonormal columns and define the subspace $\mathcal{V} := \mathcal{R}(V)$. Let $\mathcal{B}$ denote the subspace of matrices $B \in \mathbb{R}^{p\times n}$ whose null space contains $\mathcal{V}$. Then $BV = 0$ and the matrix*

$$\widehat{A} := A(I_n - VV^T) \tag{2.2}$$

satisfies the following properties:

1. *$\widehat{A} \in \mathcal{B}$;*
2. *if $A \in \mathcal{B}$, then $\widehat{A} \equiv A$;*

3. if $B \in \mathcal{B}$, then $\langle A - \widehat{A}, B\rangle = 0$.

Proof. We have $\widehat{A}V = 0$, which shows the first property. The second property implies that $AV = 0$, from which it follows that

$$\widehat{A} = A - AVV^T = A.$$

Finally, for any $B \in \mathcal{B}$, we have

$$\langle A - \widehat{A}, B\rangle = \mathrm{Trace}(B^T AVV^T) = 0,$$

where the last equality follows from the cyclic property of the trace. □

The following result is a consequence of Proposition 2.1.

COROLLARY 2.2. *The matrix (2.2) is the closest matrix to A in $\mathcal{B}$ in the Frobenius norm. The distance between the matrices A and (2.2) is $\|AVV^T\|_F$.*

The matrix closest to a given matrix with a prescribed null space also can be characterized in a different manner that does not require an orthonormal basis of the null space. It is sometimes convenient to use this characterization.

PROPOSITION 2.3. *Let $\mathcal{B}$ be the subspace of matrices $B \in \mathbb{R}^{p\times n}$ whose null space contains $\mathcal{R}(V)$, where $V \in \mathbb{R}^{n\times \ell}$ is a rank-ℓ matrix. Then the closest matrix to A in $\mathcal{B}$ in the Frobenius norm is AP, where*

$$P = I_n - V\Omega^{-1}V^T \tag{2.3}$$

with $\Omega = V^TV$.

Proof. Since the columns of V are linearly independent, the matrix Ω is positive definite and, hence, invertible. It follows that P is an orthogonal projector with null space $\mathcal{R}(V)$. The desired result now follows from Proposition 2.1. □

It follows from Proposition 2.1 and Corollary 2.2 with $V = n^{-1/2}[1, 1, \ldots, 1]^T$, or from Proposition 2.3, that the closest matrix to $L_{1,\delta}$ with null space $\mathcal{N}(L_1)$ is $L_{1,\delta}P$, where $P = [P_{h,k}] \in \mathbb{R}^{n\times n}$ is the orthogonal projector given by

$$P_{h,k} = \begin{cases} -\dfrac{1}{n}, & h \neq k, \\[2ex] \dfrac{n-1}{n}, & h = k. \end{cases}$$

Hence,

$$L_{1,\delta}P = \frac{1}{2}\begin{bmatrix} 1 & -1 & & & & 0 \\ & 1 & -1 & & & \\ & & 1 & -1 & & \\ & & & \ddots & \ddots & \\ & & & & 1 & -1 \\ -\frac{\delta}{n} & -\frac{\delta}{n} & \cdots & \cdots & -\frac{\delta}{n} & (1-\frac{1}{n})\delta \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

Thus, $\|L_{1,\delta} - L_{1,\delta}P\|_F = \frac{\delta}{2\sqrt{n}}$ is smaller than $\|L_{1,\delta} - L_{1,0}\|_F = \frac{\delta}{2}$.

We turn to square tridiagonal regularization matrices. The matrix

$$L_{2,0} = \frac{1}{4}\begin{bmatrix} 0 & 0 & 0 & & & 0 \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ 0 & & & 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{n\times n}$$

with the same null space as (1.7) is considered in [6, 21]. We can apply Propositions 2.1 or 2.3 to determine whether this matrix is the closest matrix to

$$\widetilde{L}_2 = \frac{1}{4}\begin{bmatrix} 2 & -1 & & & & 0 \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ 0 & & & & -1 & 2 \end{bmatrix} \in \mathbb{R}^{n\times n}$$

with the null space (1.9). We also may apply Proposition 2.4 below, which is analogous to Proposition 2.3 in that that no orthonormal basis for the null space is required. The result can be shown by direct computations.

PROPOSITION 2.4. *Given $A \in \mathbb{R}^{p\times n}$, the closest matrix to A in the Frobenius norm with a null space containing the linearly independent vectors $v^{(1)}, v^{(2)} \in \mathbb{R}^n$ is given by $A(I_n - C)$, where*

$$C_{i,j} = \frac{\|v^{(1)}\|^2 v_i^{(2)} v_j^{(2)} - [v_i^{(2)} v_j^{(1)} + v_i^{(1)} v_j^{(2)}](v^{(1)}, v^{(2)}) + \|v^{(2)}\|^2 v_i^{(1)} v_j^{(1)}}{\|v^{(1)}\|^2 \|v^{(2)}\|^2 - (v^{(1)}, v^{(2)})^2}. \tag{2.4}$$

It follows easily from Proposition 2.4 that the closest matrix to $\widetilde{L}_2$ with null space $\mathcal{N}(L_2)$ is $\widetilde{L}_2 P$, where $P = [P_{h,k}] \in \mathbb{R}^{n\times n}$ is an orthogonal projector defined by

$$P_{h,k} = \delta_{h,k} - \frac{2(n+1)(-3h+2n+1) + 6k(2h-n-1)}{n(n+1)(n-1)}, \quad h,k = 1,\ldots,n. \tag{2.5}$$

The regularization matrices constructed above are generally nonsymmetric. We are also interested in determining the distance between a given nonsingular symmetric matrix, such as $\widetilde{L}_2$, and the closest symmetric matrix with a prescribed null space, such as (1.9). The following results shed light on this.

PROPOSITION 2.5. *Let the matrix $A \in \mathbb{R}^{n\times n}$ be symmetric, let $V \in \mathbb{R}^{n\times \ell}$ have $1 \le \ell < n$ orthonormal columns and define the subspace $\mathcal{V} := \mathcal{R}(V)$. Let $\mathcal{B}_{sym}$ denote the subspace of symmetric matrices $B \in \mathbb{R}^{n\times n}$ whose null space contains $\mathcal{V}$. Then $BV = 0$ and the matrix*

$$\widehat{A} = (I_n - VV^T)A(I_n - VV^T) \tag{2.6}$$

satisfies the following properties:

1. *$\widehat{A} \in \mathcal{B}_{sym}$;*
2. *if $A \in \mathcal{B}_{sym}$, then $\widehat{A} \equiv A$;*
3. *if $B \in \mathcal{B}_{sym}$, then $\langle A - \widehat{A}, B\rangle = 0$.*

Proof. We have $\widehat{A} = \widehat{A}^T$ and $\widehat{A}V = 0$, which shows the first property. The second property implies that $AV = V^T A = 0$, from which it follows that

$$\widehat{A} = A - VV^T A - AVV^T + VV^T AVV^T = A.$$

Finally, for any $B \in \mathcal{B}_{sym}$, it follows from the cyclic property of the trace that

$$\langle A - \widehat{A}, B\rangle = \mathrm{Trace}(BVV^T A + BAVV^T - BVV^T AVV^T) = 0.$$

□

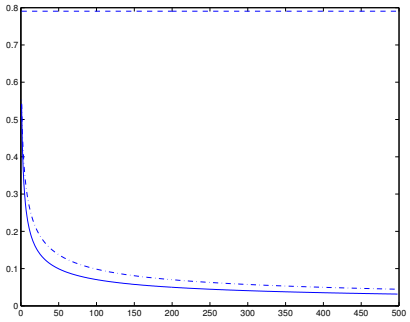

FIG. 2.1. *Distances* $\|\widetilde{L}_2 - L_{2,0}\|_F$ *(dashed curve),* $\|\widetilde{L}_2 - P\widetilde{L}_2P\|_F$ *(dash-dotted curve), and* $\|\widetilde{L}_2 - \widetilde{L}_2P\|_F$ *(solid curve) as a function of the matrix order* n*.*

COROLLARY 2.6. *The matrix (2.6) is the closest matrix to* A *in* $\mathcal{B}_{sym}$ *in the Frobenius norm. The distance between the matrices* A *and (2.6) is given by* $\|VV^TAVV^T - VV^TA - AVV^T\|_F$*.*

Proposition 2.5 characterizes the closest matrix in $\mathcal{B}_{sym}$ to a given symmetric matrix $A \in \mathbb{R}^{n\times n}$. The following proposition provides another characterization that does not explicitly use an orthonormal basis for the prescribed null space. The result follows from Proposition 2.5 and Corollary 2.6 in a straightforward manner.

PROPOSITION 2.7. *Let* $\mathcal{B}_{sym}$ *be the subspace of symmetric matrices* $B \in \mathbb{R}^{n\times n}$ *whose null space contains* $\mathcal{R}(V)$*, where* $V \in \mathbb{R}^{n\times\ell}$ *is a rank-*ℓ *matrix. Then the closest matrix to the symmetric matrix* A *in* $\mathcal{B}_{sym}$ *in the Frobenius norm is* PAP*, where* $P \in \mathbb{R}^{n\times n}$ *is defined by (2.3).*

We are interested in determining the closest symmetric matrix to $\widetilde{L}_2$ with null space in (1.9). It is given by $P\widetilde{L}_2P$, with P defined in (2.5). One has

$$\|\widetilde{L}_2 - \widetilde{L}_2P\|_F < \|\widetilde{L}_2 - P\widetilde{L}_2P\|_F < \|\widetilde{L}_2 - L_{2,0}\|_F = \frac{\sqrt{10}}{4}.$$

Figure 2.1 displays the three distances for increasing matrix dimensions.

3. Application of the regularization matrices. In this section we discuss the use of regularization matrices of the form $L = \tilde{L}P$ and $L = P\tilde{L}P$ in the Tikhonov minimization problem (1.4), where P is an orthogonal projector and $\tilde{L}$ is nonsingular. We solve the problem (1.4) by transforming it to standard form in two steps. First, we let $\boldsymbol{y} = P\boldsymbol{x}$ and then set $\boldsymbol{z} = \tilde{L}\boldsymbol{y}$. Following Eldén [8] or Morigi et al. [16], we express the Tikhonov minimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ \|K\boldsymbol{x} - \boldsymbol{b}\|^2 + \mu\|\tilde{L}P\boldsymbol{x}\|^2 \right\} \tag{3.1}$$

in the form

$$\min_{\boldsymbol{y}\in\mathbb{R}^n} \left\{ \|K_1\boldsymbol{y} - \boldsymbol{b}_1\|^2 + \mu\|\tilde{L}\boldsymbol{y}\|^2 \right\}, \tag{3.2}$$

where

$$K_1 = KP_K^\dagger, \qquad P_K^\dagger = (I_n - (K(I_n - P^\dagger P))^\dagger K)P$$

and

$$\boldsymbol{b}_1 = \boldsymbol{b} - K\boldsymbol{x}^{(0)}, \qquad \boldsymbol{x}^{(0)} = (K(I_n - P^\dagger P))^\dagger \boldsymbol{b}.$$

Let the columns of V_ℓ form an orthonormal basis for the desired null space of L. Then $P = I_n - V_\ell V_\ell^T$. Determine the QR factorization

$$KV_\ell = QR, \tag{3.3}$$

where $Q \in \mathbb{R}^{m\times\ell}$ has orthonormal columns and $R \in \mathbb{R}^{\ell\times\ell}$ is upper triangular. It follows from (1.5) that R is nonsingular, and we obtain

$$P_K^\dagger = I_n - V_\ell R^{-1} Q^T K, \qquad KP_K^\dagger = (I_m - QQ^T)K. \tag{3.4}$$

These formulas are convenient to use in iterative methods for the solution of (3.2); see [16] for details. Let $\boldsymbol{y}_\mu$ solve (3.2). Then the solution of (3.1) is given by $\boldsymbol{x}_\mu = P_K^\dagger \boldsymbol{y}_\mu + \boldsymbol{x}^{(0)}$.

We turn to the solution of (3.2). This minimization problem can be expressed in standard form

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} \left\{ \|K_2\boldsymbol{z} - \boldsymbol{b}_1\|^2 + \mu\|\boldsymbol{z}\|^2 \right\}, \tag{3.5}$$

where $K_2 = K_1\tilde{L}^{-1}$. Let $\boldsymbol{z}_\mu$ solve (3.5). Then the solution of (3.2) is given by $\boldsymbol{y}_\mu = \tilde{L}^{-1}\boldsymbol{z}_\mu$. In actual computations, we evaluate $\tilde{L}^{-1}\boldsymbol{z}$ by solving a linear system of equations with $\tilde{L}$. We can similarly solve the problem (1.4) with $L = P\tilde{L}P$ by transforming it to standard form in three steps, where the first two steps are the same as above and the last step is similar to the first step of the case with $L = \tilde{L}P$.

It is desirable that the matrix $\tilde{L}$ not be very ill-conditioned to avoid severe error propagation when solving linear systems of equations with this matrix. For instance, the condition number of the regularization matrix $L_{1,\delta}$, defined by (2.1), depends on the parameter $\delta > 0$. Clearly, the condition number of $L_{1,\delta}$, defined as the ratio of the largest and smallest singular value of the matrix, is large for $\delta > 0$ "tiny" and of moderate size for $\delta = 1$. In the computations reported in Section 5, we use the latter value.

4. Krylov subspace methods and the discrepancy principle. A variety of Krylov subspace iterative methods are available for the solution of the Tikhonov minimization problem (3.5); see, e.g., [2, 3, 10, 17] for discussions and references. The discrepancy principle is a popular approach to determining the regularization parameter μ when a bound ε for the norm of the error $\boldsymbol{e}$ in $\boldsymbol{b}$ is known, i.e., $\|\boldsymbol{e}\| \le \varepsilon$. It can be shown that the error in $\boldsymbol{b}_1$ satisfies the same bound. The discrepancy principle prescribes that $\mu > 0$ be chosen so that the solution $\boldsymbol{z}_\mu$ of (3.5) satisfies

$$\|K_2\boldsymbol{z}_\mu - \boldsymbol{b}_1\| = \eta\varepsilon, \tag{4.1}$$

where $\eta > 1$ is a constant independent of ε. This is a nonlinear equation of μ.

We can determine an approximation of $\boldsymbol{z}_\mu$ by applying an iterative method to the linear system of equations

$$K_2\boldsymbol{z} = \boldsymbol{b}_1 \tag{4.2}$$

and terminating the iterations sufficiently early. This is simpler than solving (3.5), because it circumvents the need to solve the nonlinear equation (4.1) for μ. We

therefore use this approach in the computed examples of Section 5. Specifically, we apply the Range Restricted GMRES (RRGMRES) iterative method described in [17]. At the kth step, this method computes an approximate solution $\boldsymbol{z}_k$ of (4.2) as the solution of the minimization problem

$$\min_{\boldsymbol{z}\in\mathcal{K}_k(K_2,K_2\boldsymbol{b}_1)} \|K_2\boldsymbol{z}-\boldsymbol{b}_1\|,$$

where $\mathcal{K}_k(K_2,K_2\boldsymbol{b}_1) := \mathrm{span}\{K_2\boldsymbol{b}_1, K_2^2\boldsymbol{b}_1,\ldots,K_2^k\boldsymbol{b}_1\}$ is a Krylov subspace. The discrepancy principle prescribes that the iterations with RRGMRES be terminated as soon as an iterate $\boldsymbol{z}_k$ that satisfies

$$\|K_2\boldsymbol{z}_k-\boldsymbol{b}_1\| \le \eta\varepsilon \tag{4.3}$$

has been computed. The number of iterations required to satisfy this stopping criterion generally increases as ε is decreased. Using the transformation from $\boldsymbol{z}_\mu$ to $\boldsymbol{x}_\mu$ described in Section 3, we transform $\boldsymbol{z}_k$ to an approximate solution $\boldsymbol{x}_k$ of (1.1). Further details can be found in [17]. Here we only note that $\|K_2\boldsymbol{z}_k-\boldsymbol{b}_1\|$ can be computed without explicitly evaluating the matrix-vector product $K_2\boldsymbol{z}_k$.

5. Numerical examples. We illustrate the performance of regularization matrices of the form $L=\tilde{L}P$ and $L=P\tilde{L}P$. The error vector $\boldsymbol{e}$ has in all examples normally distributed pseudorandom entries with mean zero, and is normalized to correspond to a chosen noise level

$$\nu := \frac{\|\boldsymbol{e}\|}{\|\widehat{\boldsymbol{b}}\|},$$

where $\widehat{\boldsymbol{b}}$ denotes the error-free right-hand side vector in (1.3). We let $\eta=1.01$ in (4.3) in all examples. Throughout this section P_1 and P_2 denote orthogonal projectors with null spaces (1.8) and (1.9), respectively. All computations are carried out on a computer with an Intel Core i5-3230M @ 2.60GHz processor and 8GB ram using MATLAB R2012a. The computations are done with about 15 significant decimal digits.

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

$$\int_{-6}^{6}\kappa(\tau,\sigma)x(\sigma)d\sigma = g(\tau), \qquad -6\le\tau\le 6, \tag{5.1}$$

with kernel and solution given by

$$\kappa(\tau,\sigma) := x(\tau-\sigma)$$

and

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

This equation is discussed by Phillips [19]. We use the MATLAB code phillips from [12] to discretize (5.1) by a Galerkin method with 200 orthonormal box functions as test and trial functions. The code produces the matrix $K\in\mathbb{R}^{200\times 200}$ and a scaled discrete approximation of $x(\sigma)$. Adding $\boldsymbol{n}_1=[1,1,\ldots,1]^T$ to the latter yields the vector $\widehat{\boldsymbol{x}}\in\mathbb{R}^{200}$ with which we compute the error free right-hand side $\widehat{\boldsymbol{b}} := K\widehat{\boldsymbol{x}}$.

reg. mat.	# iterations k	# mat.-vec. prod.	$\|\boldsymbol{x}_k - \widehat{\boldsymbol{x}}\|/\|\widehat{\boldsymbol{x}}\|$
		noise level $\nu = 1 \cdot 10^{-2}$	
I	4	5	$3.5 \cdot 10^{-2}$
$L_{1,0}$	3	4	$6.5 \cdot 10^{-3}$
$L_{1,\delta}P_1$	5	7	$5.1 \cdot 10^{-3}$
$L_{2,0}$	3	4	$6.6 \cdot 10^{-3}$
$\tilde{L}_2P_2$	4	7	$9.5 \cdot 10^{-3}$
$P_2\tilde{L}_2P_2$	1	7	$1.5 \cdot 10^{-2}$
		noise level $\nu = 1 \cdot 10^{-3}$	
I	9	10	$1.7 \cdot 10^{-2}$
$L_{1,0}$	3	4	$4.5 \cdot 10^{-3}$
$L_{1,\delta}P_1$	7	9	$1.2 \cdot 10^{-3}$
$\tilde{L}_{2,0}$	3	4	$4.5 \cdot 10^{-3}$
$\tilde{L}_2P_2$	5	8	$4.1 \cdot 10^{-3}$
$P_2\tilde{L}_2P_2$	5	11	$1.4 \cdot 10^{-2}$
		noise level $\nu = 1 \cdot 10^{-4}$	
I	10	11	$6.1 \cdot 10^{-3}$
$L_{1,0}$	6	7	$2.8 \cdot 10^{-3}$
$L_{1,\delta}P_1$	9	11	$2.0 \cdot 10^{-3}$
$\tilde{L}_{2,0}$	6	7	$2.8 \cdot 10^{-3}$
$\tilde{L}_2P_2$	7	10	$2.1 \cdot 10^{-3}$
$P_2\tilde{L}_2P_2$	6	12	$3.9 \cdot 10^{-3}$

TABLE 5.1

Example 5.1: Number of iterations, number of matrix-vector product evaluations with the matrix K, and relative error in approximate solutions $\boldsymbol{x}_k$ determined by truncated iteration with RRGMRES using the discrepancy principle and different regularization matrices for several noise levels.

This provides an example of a problem for which it is undesirable to damp the $\boldsymbol{n}_1$-component in the computed approximation of $\widehat{\boldsymbol{x}}$.

The error vector $\boldsymbol{e} \in \mathbb{R}^{200}$ is generated as described above and normalized to correspond to different noise levels $\nu \in \{1 \cdot 10^{-2}, 1 \cdot 10^{-3}, 1 \cdot 10^{-4}\}$. The data vector $\boldsymbol{b}$ in (1.1) is obtained from (1.2).

Table 5.1 displays results obtained with RRGMRES for several regularization matrices and different noise levels, and Figure 5.1 shows three computed approximate solutions obtained for the noise level $\nu = 1 \cdot 10^{-4}$. The iterations are terminated by the discrepancy principle (4.3). From Table 5.1 and Figure 5.1, we can see that the regularization matrix $L = L_{1,\delta}P_1$ yields the best approximation of $\widehat{\boldsymbol{x}}$. The worst approximation is obtained when no regularization matrix is used with RRGMRES. This situation is denoted by $L = I$. Table 5.1 shows both the number of iterations and the number of matrix-vector product evaluations with the matrix K. The fact that the latter number is larger depends on the ℓ matrix-vector product evaluations with K required to evaluate the left-hand side of (3.3) and the matrix-vector product with K needed for evaluating the product of $P_K^\dagger$ with a vector; cf. (3.4). □

Example 5.2. Regard the Fredholm integral equation of the first kind,

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

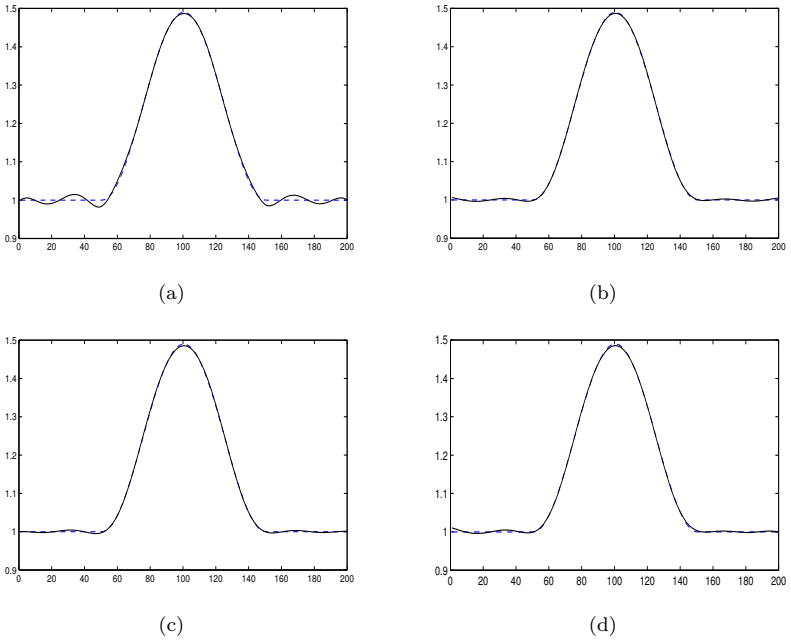

FIG. 5.1. *Example 5.1: Continuous curves: Computed approximate solutions $\boldsymbol{x}_k$ determined by truncated iteration with RRGMRES using the discrepancy principle. The noise level is $\nu = 1 \cdot 10^{-4}$. (a) Iterate $\boldsymbol{x}_{10}$ determined without regularization matrix ($L = I$), (b) iterate $\boldsymbol{x}_9$ determined with the regularization matrix $L = L_{1,\delta} P_1$, (c) iterate $\boldsymbol{x}_7$ determined with the regularization matrix $L = \tilde{L}_2 P_2$, and (d) iterate $\boldsymbol{x}_6$ determined with the regularization matrix $L = P_2 \tilde{L}_2 P_2$. The dashed curves show the desired solution $\widehat{\boldsymbol{x}}$.*

where

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

We discretize (5.2) by a Galerkin method with orthonormal box functions as test and trial functions using the MATLAB program deriv2 from [12]. This program yields a symmetric indefinite matrix $K \in \mathbb{R}^{200\times 200}$ and a scaled discrete approximation of the solution $x(t) = \exp(t)$ of (5.2). Adding $\boldsymbol{n}_1 = [1, 1, \ldots, 1]^T$ yields the vector $\widehat{\boldsymbol{x}} \in \mathbb{R}^{200}$ with which we compute the error-free right-hand side $\widehat{\boldsymbol{b}} := K\widehat{\boldsymbol{x}}$. Error vectors $\boldsymbol{e} \in \mathbb{R}^{200}$ are constructed similarly as in Example 5.1, and the data vector $\boldsymbol{b}$ in (1.1) is obtained from (1.2).

Table 5.2 shows results obtained with RRGMRES for different regularization matrices. The performance for three noise levels is displayed. The iterations are terminated with the aid of the discrepancy principle (4.3). When $L = \tilde{L}_{2,0}$, $L = \tilde{L}_2 P_2$ or $L = P_2 \tilde{L}_2 P_2$, and the noise level is $\nu = 1 \cdot 10^{-2}$, as well as when $L = P_2 \tilde{L}_2 P_2$, and the noise level is $\nu = 1 \cdot 10^{-3}$ or $\nu = 1 \cdot 10^{-4}$, the initial residual $\boldsymbol{r}_0 := \boldsymbol{b} - A\boldsymbol{x}^{(0)}$ satisfies the discrepancy principle and no iterations are carried out. Figure 5.2 shows computed approximate solutions obtained for the noise level $\nu = 1 \cdot 10^{-4}$ with the regularization matrix $L = \tilde{L}_2 P_2$ and without regularization matrix. Table 5.2 and Figure 5.2 show the regularization matrix $L = \tilde{L}_2 P_2$ to give the most accurate approximations of the

reg. mat.	# iterations k	# mat.-vec. prod.	$\|\boldsymbol{x}_k - \widehat{\boldsymbol{x}}\|/\|\widehat{\boldsymbol{x}}\|$
		noise level $v = 1 \cdot 10^{-2}$	
I	4	5	$2.4 \cdot 10^{-1}$
$L_{1,0}$	1	2	$1.1 \cdot 10^{-2}$
$L_{1,\delta}P_1$	1	3	$2.6 \cdot 10^{-2}$
$\tilde{L}_{2,0}$	0	1	$3.1 \cdot 10^{-3}$
$\tilde{L}_2P_2$	0	3	$3.1 \cdot 10^{-3}$
$P_2\tilde{L}_2P_2$	0	6	$1.1 \cdot 10^{-2}$
		noise level $v = 1 \cdot 10^{-3}$	
I	8	9	$1.5 \cdot 10^{-1}$
$L_{1,0}$	3	4	$7.3 \cdot 10^{-3}$
$L_{1,\delta}P_1$	7	9	$4.6 \cdot 10^{-2}$
$\tilde{L}_{2,0}$	1	2	$1.9 \cdot 10^{-3}$
$\tilde{L}_2P_2$	1	4	$1.7 \cdot 10^{-3}$
$P_2\tilde{L}_2P_2$	0	6	$1.1 \cdot 10^{-3}$
		noise level $v = 1 \cdot 10^{-4}$	
I	13	14	$1.0 \cdot 10^{-1}$
$L_{1,0}$	2	3	$5.6 \cdot 10^{-3}$
$L_{1,\delta}P_1$	26	28	$8.0 \cdot 10^{-2}$
$\tilde{L}_{2,0}$	2	3	$1.4 \cdot 10^{-3}$
$\tilde{L}_2P_2$	3	6	$1.2 \cdot 10^{-3}$
$P_2\tilde{L}_2P_2$	0	6	$9.5 \cdot 10^{-5}$

TABLE 5.2

Example 5.2: Number of iterations, number of matrix-vector product evaluations with the matrix K, and relative error in approximate solutions $\boldsymbol{x}_k$ determined by truncated iteration with RRGMRES using the discrepancy principle and different regularization matrices for several noise levels.

desired solution $\widehat{\boldsymbol{x}}$. We remark that addition of the vector $\boldsymbol{n}_1$ to to the solution vector determined by the program deriv2 enhances the benefit of using a regularization matrix different from the identity. The benefit would be even larger, if a larger multiple of the vector $\boldsymbol{n}_1$ were added to the solution. □

The above examples illustrate the performance of regularization matrices suggested by the theory developed in Section 2. Other combinations of nonsingular regularization matrices and orthogonal projectors also can be applied. For instance, the regularization matrix $L = \tilde{L}_2P_1$ performs as well as $L = \tilde{L}_2P_2$ when applied to the solution of the problem of Example 5.1.

6. Conclusion. This paper presents a novel method to determine regularization matrices via the solution of a matrix nearness problem. Numerical examples illustrate the effectiveness of the regularization matrices so obtained. While all examples used the discrepancy principle to determine a suitable regularized approximate solution of (1.1), other parameter choice rules also can be applied; see, e.g., [14, 20] for discussions and references.

Acknowledgment. SN is grateful to Paolo Buttà for valuable discussions and comments on part of the present work. The authors would like to thank a referee for comments.

REFERENCES

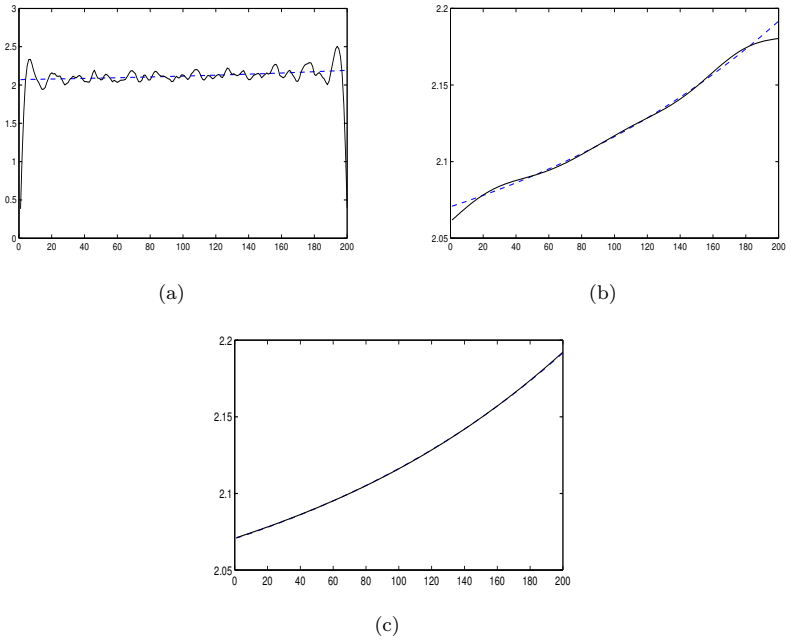

FIG. 5.2. *Example 5.2: Continuous curves: Computed approximate solutions $\boldsymbol{x}_k$ determined by truncated iteration with RRGMRES using the discrepancy principle. The noise level is $\nu = 1 \cdot 10^{-4}$. (a) Iterate $\boldsymbol{x}_{13}$ determined without regularization matrix ($L := I$), (b) iterate $\boldsymbol{x}_3$ determined with the regularization matrix $L = \tilde{L}_2 P_2$ and (c) iterate $\boldsymbol{x}_0$ determined with the regularization matrix $L = P_2 \tilde{L}_2 P_2$. The dashed curves show the desired solution $\widehat{\boldsymbol{x}}$.*

[1] C. Brezinski, M. Redivo–Zaglia, G. Rodriguez, and S. Seatzu, *Extrapolation techniques for ill-conditioned linear systems*, Numer. Math., 81 (1998), pp. 1–29.
[2] D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari, *Tikhonov regularization and the L-curve for large discrete ill-posed problems*, J. Comput. Appl. Math., 123 (2000), pp. 423–446.
[3] D. Calvetti and L. Reichel, *Tikhonov regularization of large linear problems*, BIT, 43 (2003), pp. 263–283.
[4] D. Calvetti, L. Reichel, and A. Shuibi, *Invertible smoothing preconditioners for linear discrete ill-posed problems*, Appl. Numer. Math., 54 (2005), pp. 135–149.
[5] M. Donatelli, A. Neuman, and L. Reichel, *Square regularization matrices for large linear discrete ill-posed problems*, Numer. Linear Algebra Appl., 19 (2012), pp. 896–913.
[6] M. Donatelli and L. Reichel, *Square smoothing regularization matrices with accurate boundary conditions*, J. Comput. Appl. Math., 272 (2014), pp. 334–349.
[7] L. Dykes, S. Noschese, L. Reichel, *Rescaling the GSVD with application to ill-posed problems*, Numer. Algorithms, 68 (2015), pp. 531–545.
[8] L. Eldén, *A weighted pseudoinverse, generalized singular values, and constrained least squares problems*, BIT, 22 (1982), pp. 487–501.
[9] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
[10] S. Gazzola, P. Novati, and M. R. Russo, *On Krylov projection methods and Tikhonov regularization*, Electron. Trans. Numer. Anal., 44 (2015), pp. 83–123.
[11] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.
[12] P. C. Hansen, *Regularization tools version 4.0 for MATLAB 7.3*, Numer. Algorithms, 46 (2007), pp. 189–194.
[13] P. C. Hansen and T. K. Jensen, *Smoothing norm preconditioning for regularizing minimum residual methods*, SIAM J. Matrix Anal. Appl., 29 (2006), pp. 1–14.
[14] S. Kindermann, *Discretization independent convergence rates for noise level-free parameter*

choice rules for the regularization of ill-conditioned problems, Electron. Trans. Numer. Anal., 40 (2013), pp. 58–81.

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

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

[17] A. Neuman, L. Reichel, and H. Sadok, *Implementations of range restricted iterative methods for linear discrete ill-posed problems*, Linear Algebra Appl., 436 (2012), pp. 3974–3990.

[18] S. Noschese and L. Reichel, *Inverse problems for regularization matrices*, Numer. Algorithms, 60 (2012), pp. 531–544.

[19] D. L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, J. ACM, 9 (1962), pp. 84–97.

[20] L. Reichel and G. Rodriguez, *Old and new parameter choice rules for discrete ill-posed problems*, Numer. Algorithms, 63 (2013), pp. 65–87.

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