ORTHOGONAL PROJECTION REGULARIZATION OPERATORS

S. MORIGI*, L. REICHEL†, AND F. SGALLARI‡

Abstract. Tikhonov regularization often is applied with a finite difference regularization operator that approximates a low-order derivative. This paper proposes the use of orthogonal projections as regularization operators, e.g., with the same null space as commonly used finite difference operators. Applications to iterative and SVD-based methods for Tikhonov regularization are described. Truncated iterative and SVD methods are also considered.

Key words. ill-posed problem, regularization operator, orthogonal projection, Tikhonov regularization, truncated iteration, SVD, GSVD.

1. Introduction. This paper considers the computation of an approximate solution of linear systems of equations of the form

$$A\boldsymbol{x} = \boldsymbol{b}, \qquad A \in \mathbb{R}^{m\times n}, \qquad \boldsymbol{x} \in \mathbb{R}^n, \qquad \boldsymbol{b} \in \mathbb{R}^m, \qquad m \geq n, \tag{1.1}$$

where the matrix A is assumed to be of ill-determined rank. In particular, A is severely ill conditioned and may be singular. Linear systems of equations with a matrix of ill-determined rank are often referred to as linear discrete ill-posed problems. They arise from the discretization of ill-posed problems, such as Fredholm integral equations of the first kind with a smooth kernel, as well as in image deblurring problems. The system (1.1) is not required to be consistent.

Ill-posed problems occur when one seeks to determine the cause of an observed effect. The latter is represented by the right-hand side $\boldsymbol{b}$, which in applications often is contaminated by an unknown measurement error $\boldsymbol{e} \in \mathbb{R}^m$, i.e.,

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

where $\hat{\boldsymbol{b}}$ denotes the unknown error-free right-hand side vector associated with $\boldsymbol{b}$. We will refer to the error $\boldsymbol{e}$ as "noise."

The present paper discusses methods for the solution of (1.1) that are applicable when the norm of the noise,

$$\varepsilon := \|\boldsymbol{e}\|, \tag{1.3}$$

or an accurate approximation thereof, are available. However, we remark that orthogonal projection regularization operators also can be applied in methods that do not require knowledge of the error norm (1.3). Throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm.

Introduce the linear system of equations with unknown error-free right-hand side,

$$A\boldsymbol{x} = \hat{\boldsymbol{b}}, \tag{1.4}$$

associated with (1.1). We assume that the system (1.4) is consistent, and let $\hat{\boldsymbol{x}}$ denote its solution of minimal Euclidean norm. We would like to determine an approximation

*Department of Mathematics, University of Bologna, Piazza Porta S. Donato 5, 40127 Bologna, Italy. E-mail: morigi@dm.unibo.it.

†Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. E-mail: reichel@math.kent.edu. Research supported in part by an OBR Research Challenge Grant.

‡CIRAM Department of Mathematics, University of Bologna, via Saragozza 8, 40123 Bologna, Italy. E-mail: sgallari@dm.unibo.it. Research supported in part by PRIN 2004 grant 2004014411-005.

of $\hat{\boldsymbol{x}}$ and seek to achieve this by computing an approximate solution of the available linear discrete ill-posed problem (1.1). We remark that due to the severe ill-conditioning of A and the error $\boldsymbol{e}$ in the right-hand side $\boldsymbol{b}$, the least-squares solution of minimal Euclidean norm of (1.1) typically is not a meaningful approximation of $\hat{\boldsymbol{x}}$. In order to be able to obtain an accurate approximation of $\hat{\boldsymbol{x}}$, the system (1.1) has to be replaced by a nearby system, whose solution is less sensitive to perturbations in the right-hand side $\boldsymbol{b}$. This replacement is known as regularization.

One of the most popular regularization methods, known as Tikhonov regularization, replaces the linear system (1.1) by the minimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \{\|A\boldsymbol{x}-\boldsymbol{b}\|^2 + \mu\|L\boldsymbol{x}\|^2\}, \tag{1.5}$$

where the matrix $L \in \mathbb{R}^{k\times n}$, $k \leq n$, is referred to as the regularization operator and the scalar $\mu \geq 0$ as the regularization parameter; see, e.g., Engl et al. [6], Hanke and Hansen [9], and Hansen [11] for discussions on Tikhonov regularization.

Let $\mathcal{N}(M)$ and $\mathcal{R}(M)$ denote the null space and range of the matrix M, respectively. We assume that the matrices A and L satisfy

$$\mathcal{N}(A) \cap \mathcal{N}(L) = \{\boldsymbol{0}\}. \tag{1.6}$$

Then the minimization problem (1.5) has the unique solution

$$\boldsymbol{x}_\mu := (A^T A + \mu L^T L)^{-1} A^T \boldsymbol{b}. \tag{1.7}$$

Common choices of regularization operators in applications are the identity matrix, as well as the following scaled discretizations of one-dimensional finite difference operators

$$L_1 := \begin{bmatrix} 1 & -1 & & & 0 \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ 0 & & & 1 & -1 \end{bmatrix} \in \mathbb{R}^{(n-1)\times n}, \tag{1.8}$$

$$L_2 := \begin{bmatrix} 1 & -2 & 1 & & & 0 \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ 0 & & & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{(n-2)\times n} \tag{1.9}$$

and

$$L_3 := \begin{bmatrix} -1 & 3 & -3 & 1 & & & 0 \\ & -1 & 3 & -3 & 1 & & \\ & & \ddots & \ddots & \ddots & \ddots & \\ 0 & & & -1 & 3 & -3 & 1 \end{bmatrix} \in \mathbb{R}^{(n-3)\times n}. \tag{1.10}$$

It is the purpose of the present paper to propose the use of orthogonal projections

$$L := I - WW^T, \qquad W \in \mathbb{R}^{n\times \ell}, \qquad W^T W = I, \tag{1.11}$$

as regularization operator. The advantages of these orthogonal projection regularization operators, compared with the finite difference-based operators (1.8)-(1.10), include

i) The A-weighted pseudoinverse of L, defined in Section 2, is of simple form. This makes the regularization operators (1.11) easy to use in Tikhonov regularization.
ii) The matrix W can be chosen in many different ways, some of which may yield regularization operators that give more accurate approximations of $\hat{\boldsymbol{x}}$ than finite difference-based regularization operators.
iii) The regularization operators (1.11) are represented by a square matrix and can be applied in conjunction with iterative methods, such as MR-II and RRGMRES, that do not allow regularization operators that are represented by a non-square matrix, such as (1.8)-(1.10); see Section 4 for further details.

Orthogonal projection regularization operators (1.11) with W defined as described in the following examples will be applied in the numerical experiments of Section 5.

Example 1.1. The orthogonal projection (1.11) with

$$W := \frac{1}{n^{1/2}}[1, 1, \ldots, 1]^T \in \mathbb{R}^n$$

has the same null space as the regularization operator (1.8). □

Example 1.2. Let W be determined by the QR-factorization

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n \end{bmatrix} = WR,$$

where $W \in \mathbb{R}^{n\times 2}$ has orthonormal columns and $R \in \mathbb{R}^{2\times 2}$ is upper triangular. Then the orthogonal projection (1.11) has the same null space as the regularization operator (1.9). □

Example 1.3. The choice $W := \boldsymbol{b}/\|\boldsymbol{b}\|$ in (1.11) gives an orthogonal projection regularization operator, whose null space generally is different from the null-spaces of of the finite difference regularization operators (1.8)-(1.10). □

We solve the Tikhonov minimization problem (1.5) by first transforming it to standard form. This is discussed in Section 2. An advantage of regularization operators of the form (1.11), when compared with the operators (1.8)-(1.10), is the simplicity of this transformation. Section 3 describes the solution of the transformed problem by direct methods based on the singular value decomposition (SVD) and generalized singular value decomposition (GSVD). Both Tikhonov regularization and regularization by truncated expansion is discussed. Section 4 is concerned with iterative solution methods suitable for large-scale problems. We discuss iterative methods for Tikhonov regularization, as well as regularization by truncated iteration. Section 5 presents a few computed examples, and concluding remarks can be found in Section 6.

2. Transformation to standard form. Let $L^\dagger \in \mathbb{R}^{n\times k}$ denote the Moore-Penrose pseudoinverse of the regularization operator L in (1.5). The A-weighted pseudoinverse of L is defined by

$$L_A^\dagger := \left(I - (A(I - L^\dagger L))^\dagger A\right) L^\dagger \in \mathbb{R}^{n\times k}; \tag{2.1}$$

see, e.g., Eldén [5] or Hansen [11, Section 2.3]. Introduce the matrix

$$\bar{A} := AL_A^\dagger \tag{2.2}$$

and vectors

$$\boldsymbol{x}^{(0)} := \left(A(I - L^\dagger L)\right)^\dagger \boldsymbol{b}, \tag{2.3}$$

$$\bar{\boldsymbol{b}} := \boldsymbol{b} - A\boldsymbol{x}^{(0)}. \tag{2.4}$$

Let $\bar{\boldsymbol{x}} := L\boldsymbol{x}$. Then the Tikhonov minimization problem (1.5) can be expressed in standard form

$$\min_{\bar{\boldsymbol{x}} \in \mathbb{R}^k} \{\|\bar{A}\bar{\boldsymbol{x}} - \bar{\boldsymbol{b}}\|^2 + \mu\|\bar{\boldsymbol{x}}\|^2\}. \tag{2.5}$$

The solution $\boldsymbol{x}_\mu$ of (1.5), given by (1.7), can be recovered from the solution $\bar{\boldsymbol{x}}_\mu$ of (2.5) according to

$$\boldsymbol{x}_\mu = L_A^\dagger \bar{\boldsymbol{x}}_\mu + \boldsymbol{x}^{(0)}, \tag{2.6}$$

see, e.g., [5] or [11, Section 2.3]. The following theorem shows that the A-weighted pseudoinverse (2.1) and the matrix (2.2) can be expressed in a simple manner when the regularization operator L is an orthogonal projection.

THEOREM 2.1. *Let the regularization operator L and the matrix $W \in \mathbb{R}^{n\times\ell}$ with orthonormal columns be given by (1.11), and assume that (1.6) holds. Introduce the QR-factorization*

$$AW = QR, \tag{2.7}$$

where $Q \in \mathbb{R}^{n\times\ell}$ satisfies $Q^TQ = I$ and $R \in \mathbb{R}^{\ell\times\ell}$ is upper triangular. Then

$$L_A^\dagger = I - WR^{-1}Q^TA, \tag{2.8}$$

$$AL_A^\dagger = (I - QQ^T)A. \tag{2.9}$$

Moreover,

$$L_A^\dagger \boldsymbol{z} = \boldsymbol{0}, \qquad \forall \boldsymbol{z} \in \mathcal{R}(W). \tag{2.10}$$

Proof. The columns of W span $\mathcal{N}(L)$. It follows from (1.6) that AW is of full rank or, equivalently, that the factor R in (2.7) is nonsingular.

Let the matrices $M \in \mathbb{R}^{m\times\ell}$ and $N \in \mathbb{R}^{\ell\times n}$ be of rank ℓ. Then

$$(MN)^\dagger = N^\dagger M^\dagger;$$

see, e.g., Björck [1, Theorem 1.2.13]. This property of the Moore-Penrose pseudoinverse and the following identities for L,

$$L^\dagger = L, \qquad I - L^\dagger L = WW^T$$

yield

$$(A(I - L^\dagger L))^\dagger = (AWW^T)^\dagger = W(AW)^\dagger = W(QR)^\dagger = WR^{-1}Q^T.$$

Substitution into (2.1) shows that

$$L_A^\dagger = (I - WR^{-1}Q^TA)L,$$

which simplifies to (2.8). Multiplying (2.8) by A and using (2.7) yields (2.9). Equation (2.10) follows from (2.8). □

We conclude this section with some properties of the solution $\boldsymbol{x}_\mu$ of the Tikhonov minimization problem (1.5).

THEOREM 2.2. *Let the regularization operator $L \in \mathbb{R}^{k\times n}$, $k \leq n$, satisfy (1.6) and assume that the right-hand side vector $\boldsymbol{b}$ in (1.1) lives in $A\mathcal{N}(L) = \{A\boldsymbol{z} : \boldsymbol{z} \in \mathcal{N}(L)\}$. Then the Tikhonov solution $\boldsymbol{x}_\mu$, defined by (1.7), is independent of μ.*

Proof. Let the orthonormal columns of the matrix $W \in \mathbb{R}^{n\times \ell}$ span $\mathcal{N}(L)$. Then, similarly as in the proof of Theorem 2.1,

$$(A(I - L^\dagger L))^\dagger = WR^{-1}Q^T, \tag{2.11}$$

where the matrices Q and R are defined by the QR-factorization (2.7). By assumption, $\boldsymbol{b} = AW\boldsymbol{y}$ for some $\boldsymbol{y} \in \mathbb{R}^\ell$. Substitution into (2.3), using (2.7) and (2.11), yields $\boldsymbol{x}^{(0)} = W\boldsymbol{y}$. It follows that $\bar{\boldsymbol{b}}$, defined by (2.4), vanishes. Thus, (2.5) has the unique solution $\bar{\boldsymbol{x}} = \boldsymbol{0}$ for any $\mu > 0$. Finally, by (2.6), $\boldsymbol{x}_\mu = \boldsymbol{x}^{(0)}$, which shows the independence of the solution on μ. □

THEOREM 2.3. *Let the orthonormal columns of the matrix $W \in \mathbb{R}^{n\times\ell}$ span the null space of the regularization operator $L \in \mathbb{R}^{k\times n}$ and assume that (1.6) holds. Let $\mu > 0$ and let $\boldsymbol{x}_\mu$ be the unique solution (1.7) of the Tikhonov minimization problem (1.5). Consider the discrete ill-posed problem with modified right-hand side*

$$A\boldsymbol{x} = \boldsymbol{b}', \qquad \boldsymbol{b}' := \boldsymbol{b} + AW\boldsymbol{y},$$

for some $\boldsymbol{y} \in \mathbb{R}^\ell$. Then the unique solution of the associated Tikhonov minimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \{\|A\boldsymbol{x} - \boldsymbol{b}'\|^2 + \mu\|L\boldsymbol{x}\|^2\} \tag{2.12}$$

is given by $\boldsymbol{x}'_\mu := \boldsymbol{x}_\mu + W\boldsymbol{y}$.

Proof. Let $\mu > 0$. The unique solution of (1.5) can be expressed as $\boldsymbol{x}_\mu = \bar{\boldsymbol{x}}_\mu + \boldsymbol{x}^{(0)}$, where $\bar{\boldsymbol{x}}_\mu$ is the unique solution of (2.5) and $\boldsymbol{x}^{(0)}$ is given by (2.3). Similarly, the unique solution $\boldsymbol{x}'_\mu$ of (2.12) can be expressed as

$$\boldsymbol{x}'_\mu = \bar{\boldsymbol{x}}'_\mu + \boldsymbol{x}^{(0)'}$$

with

$$\boldsymbol{x}^{(0)'} = (A(I - L^\dagger L))^\dagger(\boldsymbol{b} + AW\boldsymbol{y}) = \boldsymbol{x}^{(0)} + W\boldsymbol{y},$$

where the last equality follows as in the proof of Theorem 2.2.

Let $\bar{\boldsymbol{x}}'_\mu \in \mathbb{R}^k$ be the unique solution of the minimization problem

$$\min_{\bar{\boldsymbol{x}}\in\mathbb{R}^k} \{\|\bar{A}\bar{\boldsymbol{x}} - \bar{\boldsymbol{b}}'\|^2 + \mu\|\bar{\boldsymbol{x}}\|^2\},$$

which is analogous to (2.5), with

$$\bar{\boldsymbol{b}}' = \boldsymbol{b}' - A\boldsymbol{x}^{(0)'} = \boldsymbol{b} + AW\boldsymbol{y} - A(\boldsymbol{x}^{(0)} + W\boldsymbol{y}) = \boldsymbol{b} - A\boldsymbol{x}^{(0)} = \bar{\boldsymbol{b}}.$$

It follows that $\bar{\boldsymbol{x}}'_\mu = \bar{\boldsymbol{x}}_\mu$, which shows the theorem. □

Theorems 2.2 and 2.3 illustrate that some important properties of the computed solution $\boldsymbol{x}_\mu$ only depend on the null space of the regularization operator. We therefore can expect the computed solutions determined by finite difference-based regularization operators and by orthogonal projection regularization operators (1.11) to be quite similar if the regularization operators have the same null space. Differences in the computed solutions depend on the action of the regularization operators in the complement of the null space. For instance, since the operator (1.8) approximates the first derivative, its Moore-Penrose pseudoinverse $L^\dagger$ approximates an integral operator and, therefore, tends to be a low-pass filter. This, typically, is a desirable property when the sought solution $\hat{\boldsymbol{x}}$ is smooth. If $\hat{\boldsymbol{x}}$ is not smooth, then an orthogonal projection regularization operator may perform better than a finite difference-based regularization operator with the same null space; see Section 5 for a few computed examples.

Theorem 2.3 shows that if we would like the computed solution to be invariant under certain simple transformations, such as the choice of units in a physical model, then we can achieve this by choosing a regularization operator with a suitable null space. For instance, the orthogonal projection operator (1.11), with W chosen as in Example 1.2, and the finite difference-based operator (1.9) give computed solutions that are invariant under a change of units in the solution $\boldsymbol{x}_\mu$ that correspond to a linear transformation, e.g., the change from degrees Kelvin to degrees Fahrenheit.

3. Direct solution methods. This section discusses the computation of approximate solutions of (1.1) by methods based on computing the SVD of the matrix $AL_A^\dagger$ or computing the GSVD of the matrix pair $\{A, L\}$. The approximate solutions are determined with the aid of the *discrepancy principle*. Let $\eta \geq 1$ be a constant and let ε be determined by (1.3). A vector $\boldsymbol{x} \in \mathbb{R}^n$ is said to satisfy the discrepancy principle if

$$\|A\boldsymbol{x} - \boldsymbol{b}\| \leq \eta\varepsilon; \tag{3.1}$$

see, e.g., Hansen [11, Section 7.2] for a discussion. We seek to determine a value of the regularization parameter μ and an associated solution $\boldsymbol{x}_\mu$ of (1.5) that satisfies the discrepancy principle (3.1).

We note for future reference that for $\boldsymbol{x} = L_A^\dagger \bar{\boldsymbol{x}} + \boldsymbol{x}^{(0)}$, we have

$$\bar{A}\bar{\boldsymbol{x}} - \bar{\boldsymbol{b}} = A\boldsymbol{x} - \boldsymbol{b}, \tag{3.2}$$

where $\bar{A}$, $\boldsymbol{x}^{(0)}$, and $\bar{\boldsymbol{b}}$ are given by (2.2)-(2.4).

Orthogonal projection regularization operators (1.11) allow fairly efficient evaluation of the matrix $\bar{A}$; the computation of $\bar{A}$ from A requires the evaluation of ℓ matrix-vector products with A^T. Since the value of ℓ used in applications typically is small, e.g., $1 \leq \ell \leq 3$, and, in particular, $\ell \ll n$, the computational effort required to determine $\bar{A}$ from A generally is negligible compared with the computations needed for the determination of the singular value decomposition

$$\bar{A} = \bar{U}\bar{\Sigma}\bar{V}^T. \tag{3.3}$$

Here the matrices $\bar{U} = [\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_n] \in \mathbb{R}^{m \times n}$ and $\bar{V} = [\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n] \in \mathbb{R}^{n \times n}$ have orthonormal columns and

$$\bar{\Sigma} = \mathrm{diag}[\bar{\sigma}_1, \bar{\sigma}_2, \ldots, \bar{\sigma}_n], \qquad \bar{\sigma}_1 \geq \bar{\sigma}_2 \geq \ldots \geq \bar{\sigma}_n \geq 0.$$

It is quite straightforward to use the decomposition (3.3) to compute a value of the regularization parameter μ, such that the associated solution $\bar{\boldsymbol{x}}_\mu$ of (2.5) satisfies

$$\|\bar{A}\bar{\boldsymbol{x}}_\mu - \bar{\boldsymbol{b}}\| \leq \eta\varepsilon; \tag{3.4}$$

see, e.g., Hansen [11, Section 5.1] for a discussion on the organization of the computations. It follows from (3.2) that the solution $\boldsymbol{x}_\mu$ of (1.5) associated with $\bar{\boldsymbol{x}}_\mu$, cf. (2.6), satisfies the discrepancy principle (3.1).

The vectors (2.3) and (2.4) are also required. It follows from the proof of Theorem 2.1 that $\boldsymbol{x}^{(0)} = WR^{-1}Q^T\boldsymbol{b}$ and therefore

$$\bar{\boldsymbol{b}} = \boldsymbol{b} - QQ^T\boldsymbol{b}. \tag{3.5}$$

We remark that an alternative, mathematically equivalent, approach to the computation of $\boldsymbol{x}_\mu$ is to apply the generalized SVD (GSVD) of the matrix pair $\{A, L\}$; see Hansen [11, Section 2.3]. When orthogonal projection regularization operators (1.11) are used, the computation of the GSVD of $\{A, L\}$ typically is considerably more expensive than the formation of the matrix $\bar{A}$ and the computation of the SVD of $\bar{A}$. However, the situation is different when applying finite difference-based regularization operators, such as (1.8)-(1.10). It is more expensive to evaluate $\bar{A}$ for the latter operators, see, e.g., Eldén [4], and it is therefore attractive to use the GSVD of $\{A, L\}$. The latter approach is used in the computed examples of Section 5.

Instead of determining μ and solving (2.5), it can be attractive to compute an approximate solution of (1.1) by truncated singular value decomposition (TSVD). Specifically, we determine the smallest integer $j \geq 0$, such that

$$\bar{\boldsymbol{x}}_j := \sum_{k=1}^{j} \frac{\bar{\boldsymbol{u}}_k^T\bar{\boldsymbol{b}}}{\bar{\sigma}_k}\bar{\boldsymbol{v}}_k$$

satisfies

$$\|\bar{A}\bar{\boldsymbol{x}}_j - \bar{\boldsymbol{b}}\| \leq \eta\varepsilon. \tag{3.6}$$

It follows from (3.2) that the associated approximate solution

$$\boldsymbol{x}_j := L_A^\dagger\bar{\boldsymbol{x}}_j + \boldsymbol{x}^{(0)} \tag{3.7}$$

of (1.1) satisfies the discrepancy principle (3.1). The GSVD of $\{A, L\}$ can be applied in a similar manner to yield the truncated GSVD (TGSVD) method; see, e.g., Hansen [11, Section 3.2] for details.

4. Iterative solution methods. This section first discusses iterative methods for Tikhonov regularization and thereafter regularization by truncated iteration. An approximate solution of the Tikhonov equations (2.5) can be determined conveniently by partial Lanczos bidiagonalization of the matrix $\bar{A}$; see, e.g., [3, 7] and references therein. Each bidiagonalization step requires the evaluation of matrix-vector products with the matrices $\bar{A}$ and $\bar{A}^T$. We remark that these matrices are not explicitly formed. When L is given by (1.11), it follows from the special form of $\bar{A}$ and $\bar{A}^T$, and the orthogonality of $\bar{\boldsymbol{b}}$ to $\mathcal{R}(Q)$, that matrix-vector products with $\bar{A}^T$ can be replaced by matrix-vector products with A^T.

The computed examples of Tikhonov regularization via Lanczos bidiagonalization reported in Section 5 apply the scheme described in [3] to the approximate solution of (2.5). This scheme determines a suitable number of Lanczos bidiagonalization steps j and a value of the regularization parameter μ, such that the associated computed approximate solution $\bar{\boldsymbol{x}}_{\mu,j}$ of (2.5) satisfies (3.4). The corresponding approximate solution $\boldsymbol{x}_{\mu,j}$ of (1.5) then satisfies the discrepancy principle (3.1).

Regularization by truncated iteration is based on terminating the computations with an iterative method as soon as an iterate that satisfies the discrepancy principle (3.1) has been found. For instance, the jth iterate, $\bar{\boldsymbol{x}}_j$, generated by the LSQR iterative method applied to the solution of the, possibly overdetermined and inconsistent, system of equations

$$\bar{A}\bar{\boldsymbol{x}} = \bar{\boldsymbol{b}} \tag{4.1}$$

with initial iterate $\bar{\boldsymbol{x}}_0 = \boldsymbol{0}$ satisfies

$$\|\bar{A}\bar{\boldsymbol{x}}_j - \bar{\boldsymbol{b}}\| = \min_{\bar{\boldsymbol{x}} \in \mathcal{K}_j(\bar{A}^T\bar{A}, \bar{A}^T\bar{\boldsymbol{b}})} \|\bar{A}\bar{\boldsymbol{x}} - \bar{\boldsymbol{b}}\|, \qquad \bar{\boldsymbol{x}}_j \in \mathcal{K}_j(\bar{A}^T\bar{A}, \bar{A}^T\bar{\boldsymbol{b}}).$$

Here

$$\mathcal{K}_j(\bar{A}^T\bar{A}, \bar{A}^T\bar{\boldsymbol{b}}) := \mathrm{span}\{\bar{A}^T\bar{\boldsymbol{b}}, \bar{A}^T\bar{A}\bar{A}^T\bar{\boldsymbol{b}}, \ldots, (\bar{A}^T\bar{A})^{j-1}\bar{A}^T\bar{\boldsymbol{b}}\}$$

denotes a Krylov subspace. Iterations with LSQR are terminated as soon as an iterate, say $\bar{\boldsymbol{x}}_j$, satisfies the inequality (3.6). The associated vector $\boldsymbol{x}_j$ given by (3.7) then satisfies the discrepancy principle (3.1) and is our computed approximate solution of (1.1); see, e.g., [1, 12] for discussions of LSQR and its properties. Truncated iteration with LSQR is described in [11], where also further references can be found. Each iteration of LSQR requires the computation of one Lanczos bidiagonalization step. This demands the evaluation of matrix-vector products with $\bar{A}$ and $\bar{A}^T$, and the above comments on these evaluations apply.

When the matrix $\bar{A}$ is square, also other iterative methods, such as GMRES and RRGMRES, can be used to determine an approximate solution of (4.1). For instance, the jth iterate, $\bar{\boldsymbol{x}}_j$, determined by RRGMRES applied to (4.1) with initial approximate solution $\bar{\boldsymbol{x}}_0 = \boldsymbol{0}$ satisfies

$$\|\bar{A}\bar{\boldsymbol{x}}_j - \bar{\boldsymbol{b}}\| = \min_{\bar{\boldsymbol{x}} \in \mathcal{K}_j(\bar{A}, \bar{A}\bar{\boldsymbol{b}})} \|\bar{A}\bar{\boldsymbol{x}} - \bar{\boldsymbol{b}}\|, \qquad \bar{\boldsymbol{x}}_j \in \mathcal{K}_j(\bar{A}, \bar{A}\bar{\boldsymbol{b}}); \tag{4.2}$$

see [2] for details on the RRGMRES method and its application to the solution of ill-posed problems. The computations are terminated similarly as for LSQR, i.e., as soon as an iterate, say $\bar{\boldsymbol{x}}_j$, that satisfies (3.6) has been found. The associated vector $\boldsymbol{x}_j$ defined by (3.7) then satisfies (3.1) and is our computed approximate solution of (1.1). When the desired solution $\hat{\boldsymbol{x}}$ is the discretization of a smooth function, RRGMRES often yields a more accurate approximations of $\hat{\boldsymbol{x}}$ than GMRES.

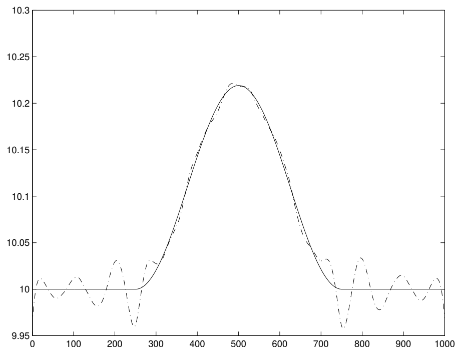
The matrix $\bar{A}$ is square, for instance, when $A \in \mathbb{R}^{n \times n}$ and L is given by (1.11). Note that $\bar{A}$ is singular in this situation. Modifications of RRGMRES and GMRES that do not break down prematurely for singular systems are discussed in [14]. If A is square, but L is one of the regularization operators (1.8)-(1.10), the matrix $\bar{A}$ is not square and RRGMRES and GMRES cannot be applied.

The minimal residual iterative method MR-II analyzed by Hanke [8, Chapter 6] is designed for the solution of linear discrete ill-posed problems with a symmetric, possibly indefinite, matrix. When a regularization operator of the form (1.11) is used, MR-II can be applied to the approximate solution of (4.1). The jth iterate, $\bar{\boldsymbol{x}}_j$, determined by MR-II with initial approximate solution $\bar{\boldsymbol{x}}_0 = \boldsymbol{0}$ satisfies equation (4.2). We remark that MR-II can be applied to the solution of (4.1) even though the matrix $\bar{A}$ given by (2.9) is not symmetric, because the Krylov subspace generated is the same as if $\bar{A}$ were replaced by the symmetric matrix $\bar{A}(I - QQ^T) = (I - QQ^T)A(I - QQ^T)$. Finite difference-based regularization operators of the form (1.8)-(1.10) cannot be applied with MR-II.

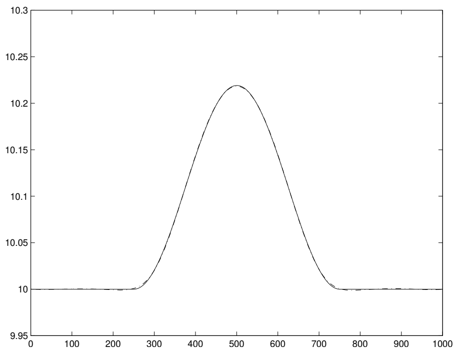
5. Computed examples. We illustrate the performance of the orthogonal projection regularization operators (1.11) and compare them to the finite difference-based operators (1.8)-(1.10). The matrices in all examples are of ill-determined rank. The computations are carried out in Matlab with about 16 significant decimal digits.

Method	c	Reg. Oper.	j	$\lVert \boldsymbol{x}_j - \hat{\boldsymbol{x}} \rVert / \lVert \hat{\boldsymbol{x}} \rVert$
LSQR	10	P_1	9	$5.84 \cdot 10^{-5}$
LSQR	10	L_1	11	$6.76 \cdot 10^{-5}$
LSQR	10	–	33	$1.26 \cdot 10^{-3}$
MR-II	10	P_1	9	$6.39 \cdot 10^{-5}$
MR-II	10	–	23	$1.34 \cdot 10^{-3}$
LSQR	100	P_1	9	$5.86 \cdot 10^{-6}$
LSQR	100	L_1	11	$6.80 \cdot 10^{-6}$
LSQR	100	–	58	$5.08 \cdot 10^{-4}$
MR-II	100	P_1	9	$6.42 \cdot 10^{-6}$
MR-II	100	–	43	$4.83 \cdot 10^{-4}$

TABLE 5.1
Example 5.1: Approximate solutions $\boldsymbol{x}_j$ determined by truncated iteration.



(a)



(b)

FIG. 5.1. *Example 5.1: Computed approximate solutions $\boldsymbol{x}_j$ determined by MR-II and the discrepancy principle with $c = 10$. The continuous curves show the vector $\hat{\boldsymbol{x}}$; the dash-dotted curves display in (a) the iterate $\boldsymbol{x}_{23}$ determined without regularization operator, and in (b) the iterate $\boldsymbol{x}_9$ determined with the regularization operator P_1. The continuous and dash-dotted curves in (b) are barely distinguishable.*

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

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

with kernel and solution given by

$$\begin{aligned} \kappa(\tau,\sigma) &:= x(\tau-\sigma) - c, \\ x(\sigma) &:= \begin{cases} c + 1 + \cos(\frac{\pi}{3}\sigma), & \text{if } |\sigma| < 3, \\ c, & \text{otherwise,} \end{cases} \end{aligned} \tag{5.2}$$

where $c \geq 0$ is a constant. The right-hand side $g(\tau)$ is defined by (5.1). Denote the solution for $c = 0$ by $x_0(\sigma)$. The solution for a general constant c is then given by $x_c(\sigma) = x_0(\sigma) + c$. Phillips [13] discusses the case when $c = 0$.

We discretize (5.1) by a Galerkin method with orthonormal box functions, using a slight modification of the Matlab code phillips available in [10] for the discretization of the equation with $c = 0$. This yields the matrix $A \in \mathbb{R}^{1000\times 1000}$ and the error-free right-hand side $\hat{\boldsymbol{b}} \in \mathbb{R}^{1000}$ of (1.4). An "error-vector" $\boldsymbol{e}$ of norm 0.01 with normally distributed components with zero mean is added to $\hat{\boldsymbol{b}}$ to yield the right-hand side $\boldsymbol{b}$ of the linear system (1.1); cf. (1.2).

Table 5.1 reports the performance of truncated iteration with the LSQR and MR-II methods (without reorthogonalization) applied to the solution of (1.1) when no regularization operator is used, or to the solution of (4.1) with the regularization operators L_1, defined by (1.8), or (1.11) with $W := [1, 1, \ldots, 1]^T/\sqrt{1000}$. The latter operator is denoted by P_1. Table 5.1 reports computed results for $c = 10$ and $c = 100$. The table shows the truncation index j determined by the discrepancy principle with $\eta = 1$, as well as the relative error in the associated computed approximate solutions $\boldsymbol{x}_j$ of (1.1). The regularization operators P_1 and L_1 can be seen to yield approximate solutions with errors of about the same size. The benefit from using these regularization operators increases with the value of c; in fact when $c = 0$ the regularization operators P_1 and L_1 do not improve the quality of the computed approximate solutions. LSQR and MR-II determine approximate solutions $\boldsymbol{x}_j$ with errors of about the same size. However, LSQR requires the evaluation of more matrix-vector products than MR-II; each iteration with LSQR requires the computation of two matrix-vector products while each iteration with MR-II only demands one matrix-vector product evaluation; see the discussion in Section 4. The regularization operator L_1 cannot be used with MR-II.

Figure 5.1(a) displays the approximate solution $\boldsymbol{x}_{23}$ determined by MR-II without regularization operator (dash-dotted curve) as well as the desired solution $\hat{\boldsymbol{x}}$ (continuous curve). Application of the regularization operator P_1 yields the computed solution $\boldsymbol{x}_9$, which is depicted by the dash-dotted graph of Figure 5.1(b). $\Box$

Method	c	Reg. Oper.	$\lVert\boldsymbol{x}_j - \hat{\boldsymbol{x}}\rVert/\lVert\hat{\boldsymbol{x}}\rVert$
TSVD	0	P_1	$9.93\cdot 10^{-3}$
TGSVD	0	L_1	$1.01\cdot 10^{-2}$
TSVD	0	–	$1.37\cdot 10^{-2}$
TSVD	10	P_1	$2.08\cdot 10^{-4}$
TGSVD	10	L_1	$2.12\cdot 10^{-4}$
TSVD	10	–	$2.42\cdot 10^{-3}$

TABLE 5.2

Example 5.2: Approximate solutions $\boldsymbol{x}_j$ determined by truncated singular value and generalized singular value decompositions.

Example 5.2. We discretize the integral equation (5.1) similarly as in Example 5.1 to obtain a linear system of equations (1.1) of order $n = 200$. Approximations $\boldsymbol{x}_j$ of $\hat{\boldsymbol{x}}$ are determined by TSVD and the discrepancy principle with $\eta = 1$ when no regularization operator or the operator P_1, defined in Example 5.1, are applied. We use the TGSVD method to apply the regularization operator L_1, given by (1.8). Table 5.2 shows the performance of the regularization operators for $c = 0$ and $c = 10$. Similarly as in Example 5.1, the use of regularization operators is most beneficial for the larger value of c. $\Box$

Example 5.3. We consider the reconstruction of a discrete wedge-shaped signal $\hat{\boldsymbol{x}} \in \mathbb{R}^{1000}$ defined on a uniform grid from an available noisy (due to the transmission

ε	Reg. Oper.	Reg. Param. μ	Bidiag. Steps j	$\|\boldsymbol{x}_{\mu,j} - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\|$
$1 \cdot 10^{-1}$	I	$1.77 \cdot 10^{1}$	2	$7.76 \cdot 10^{-2}$
	P_3	$2.88 \cdot 10^{0}$	2	$3.80 \cdot 10^{-2}$
$5 \cdot 10^{-2}$	I	$4.06 \cdot 10^{1}$	3	$5.1 \cdot 10^{-2}$
	P_3	$4.91 \cdot 10^{0}$	2	$2.37 \cdot 10^{-2}$
$1 \cdot 10^{-2}$	I	$1.95 \cdot 10^{2}$	6	$2.07 \cdot 10^{-2}$
	P_3	$2.32 \cdot 10^{1}$	4	$8.11 \cdot 10^{-3}$
$5 \cdot 10^{-3}$	I	$3.69 \cdot 10^{2}$	7	$1.41 \cdot 10^{-2}$
	P_3	$5.09 \cdot 10^{1}$	6	$6.13 \cdot 10^{-3}$

TABLE 5.3

Example 5.3: Reconstruction of wedge-shape signal by Tikhonov regularization based on partial Lanczos bidiagonalization with regularization operators P_3 or I for several noise levels.

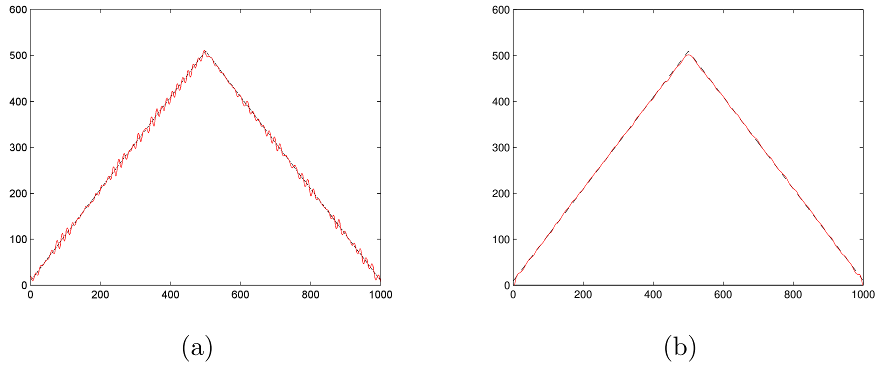


FIG. 5.2. *Example 5.3: Computed approximate solutions $\boldsymbol{x}_{\mu,j}$ determined by Tikhonov regularization based on Lanczos bidiagonalization and the discrepancy principle for noise level $1 \cdot 10^{-2}$. The dashed curves in (a) and (b) show the vector $\hat{\boldsymbol{x}}$; the continuous (red) curves display in (a) the approximate solution $\boldsymbol{x}_{\mu,6}$ determined with regularization operator I, and in (b) the approximate solution $\boldsymbol{x}_{\mu,4}$ determined with regularization operator P_3. The continuous and dashed curves in (b) are very close.*

process) and slightly smoothed (due to the signal capturing method) signal $\hat{\boldsymbol{b}} \in \mathbb{R}^{1000}$. The smoothing is modeled by convolution with a Gaussian with mean zero and variance 0.4. Let $A \in \mathbb{R}^{1000\times1000}$ be a Toeplitz matrix, such that $\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}$ models the convolution of $\hat{\boldsymbol{x}}$. The available signal $\boldsymbol{b}$ is obtained by adding noise $\boldsymbol{e} \in \mathbb{R}^{1000}$ to $\hat{\boldsymbol{b}}$ with relative error

$$\varepsilon = \frac{\|\boldsymbol{e}\|}{\|\hat{\boldsymbol{x}}\|}. \tag{5.3}$$

We refer to ε as the noise level.

Table 5.3 shows the error in computed reconstructions $\boldsymbol{x}_{\mu,j}$ of $\hat{\boldsymbol{x}}$ using Tikhonov regularization based on partial Lanczos bidiagonalization. The method determines the number of bidiagonalization steps j and a value of the regularization parameter μ, such that the computed approximate solution $\boldsymbol{x}_{\mu,j}$ satisfies the discrepancy principle (3.1) with $\eta = 1.1$. The finite difference-based regularization operators (1.8)-(1.10) give worse reconstructions than the regularization operator I and we therefore do not display these results. The regularization operator P_3, which is an orthogonal projection with the same null space as (1.10), yields improved reconstructions.

Figures 5.2(a) and (b) display the desired solution $\hat{\boldsymbol{x}}$ by a dashed curve. The noise level is $\varepsilon = 1 \cdot 10^{-2}$. The continuous curve of Figure 5.2(a) shows the computed solution $\boldsymbol{x}_{\mu,6}$ obtained with regularization operator I and the continuous curve of Figure 5.2(b) depicts the computed solution $\boldsymbol{x}_{\mu,4}$ obtained with regularization operator P_3. The latter clearly is a more accurate approximation of $\hat{\boldsymbol{x}}$ than the former.

We also apply the orthogonal projection regularization operator

$$P_4 := I - \frac{\boldsymbol{b}\boldsymbol{b}^T}{\|\boldsymbol{b}\|^2} \tag{5.4}$$

for the noise level $5 \cdot 10^{-3}$ and obtain an approximate solution with relative error $5.17 \cdot 10^{-3}$. This error is somewhat smaller than the error achieved with P_3 and illustrates that it can be beneficial to choose orthogonal projection regularization operators with a null space different from the null spaces of the operators (1.8)-(1.10). $\square$

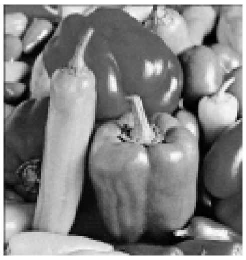
Reg. Oper.	Reg. Param. μ	Bidiag. Steps j	$\\|\boldsymbol{x}_{\mu,j} - \hat{\boldsymbol{x}}\\|/\\|\hat{\boldsymbol{x}}\\|$
I	$2.25 \cdot 10^{1}$	5	$9.48 \cdot 10^{-2}$
L_1	$4.57 \cdot 10^{0}$	67	$1.03 \cdot 10^{-1}$
P_1	$7.07 \cdot 10^{0}$	3	$6.70 \cdot 10^{-2}$
P_2	$7.04 \cdot 10^{0}$	3	$6.75 \cdot 10^{-2}$
P_3	$7.04 \cdot 10^{0}$	3	$6.75 \cdot 10^{-2}$
P_4	$1.00 \cdot 10^{-10}$	1	$7.30 \cdot 10^{-2}$
P_5	$1.00 \cdot 10^{-10}$	1	$7.35 \cdot 10^{-2}$

Table 5.4
Example 5.4: Restoration of pepper image with different regularization operators.

Example 5.4. We consider the restoration of a discrete image that has been contaminated by blur and noise. The "original" noise- and blur-free image, which is displayed by Figure 5.3(a), is assumed not to be available. It consists of 128×128 pixels, whose values are stored column-wise in the vector $\hat{\boldsymbol{x}} \in \mathbb{R}^{16384}$. The Matlab function blur from [10] with parameters $band = 3$ and $sigma = 0.7$ is applied to generate the blurring operator $A \in \mathbb{R}^{16384 \times 16384}$, which models spatially invariant Gaussian blur. Thus, the vector $\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}$ represents a blurred version of the original image $\hat{\boldsymbol{x}}$. The vector $\boldsymbol{e} \in \mathbb{R}^{16384}$, with normally distributed components with zero mean, models 5% noise; i.e., $\varepsilon = 5 \cdot 10^{-2}$ in (5.3). The right-hand side $\boldsymbol{b}$ of (1.1), obtained from (1.2), represents the available contaminated version of $\hat{\boldsymbol{x}}$ and is displayed by Figure 5.3(b).

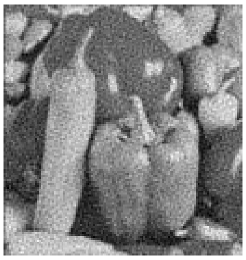
Table 5.4 shows the relative error in the restored images determined by Tikhonov regularization based on partial Lanczos bidiagonalization. The number of bidiagonalization steps is denoted by j, the computed value of the regularization parameter by μ, and the associated computed approximate solution of (1.5) by $\boldsymbol{x}_{\mu,j}$; the latter satisfy the discrepancy principle (3.1) with $\eta = 1.1$. The finite difference-based regularization operators (1.8)-(1.10) give worse reconstructions than not using a regularization operator; Table 5.4 shows the result for L_1. Let P_k, for $1 \leq k \leq 3$, denote the orthogonal projection (1.11) with the same null space as the finite difference-based regularization operator L_k. These projections give higher accuracy than the regularization operator I; the best restoration is achieved with P_1.

Table 5.4 also shows results for the regularization operators P_4, defined by (5.4), and P_5, given by (1.11) with $\mathcal{R}(W) = \text{span}\{\boldsymbol{b}, [1, 1, \ldots, 1]^T\}$. These operators yield

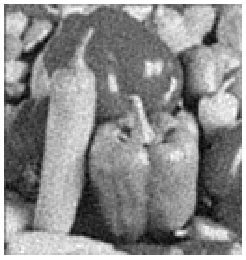


(a)

(b)



(c)



(d)

FIG. 5.3. *Example 5.4: (a) original noise- and blur-free image, (b) blurred and noisy image, (c) image determined by Tikhonov regularization with regularization operator I, and (d) image determined by Tikhonov regularization with regularization operator P_1.*

better restorations than the regularization operator I and require only one Lanczos bidiagonalization step.

Figure 5.3 displays (a) the original noise- and blur-free image, which is assumed not be available, (b) the available blurred and noisy image, and computed restorations determined with (c) the regularization operator I, and (d) the regularization operator P_1. The latter operator yields the best restoration of the noise- and blur-free image. $\square$

6. Conclusion. The numerical examples presented show that orthogonal projection regularization operators may yield approximations of the desired solution $\hat{\boldsymbol{x}}$ of higher accuracy than commonly used finite difference-based regularization operators. Moreover, orthogonal projection regularization operators are easy to compute with and can be applied with iterative methods that do not allow the use of the finite difference-based regularization operators (1.8)-(1.10).

REFERENCES

[1] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.

[2] D. Calvetti, B. Lewis, and L. Reichel, *On the choice of subspace for iterative methods for linear discrete ill-posed problems*, Int. J. Appl. Math. Comput. Sci., 11 (2001), pp. 1069–1092.

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

[4] L. Eldén, *Algorithms for the regularization of ill-conditioned least squares problems*, BIT, 17 (1977), pp. 134–145.

[5] L. Eldén, *A weighted pseudoinverse, generalized singular values, and constrained least squares problems*, BIT, 22 (1982), pp. 487–501.
[6] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
[7] G. H. Golub and U. von Matt, *Tikhonov regularization for large scale problems*, in Workshop on Scientific Computing, eds. G. H. Golub, S. H. Lui, F. Luk, and R. Plemmons, Springer, New York, 1997, pp. 3–26.
[8] M. Hanke, *Conjugate Gradient Type Methods for Ill-Posed Problems*, Longman, Essex, 1995.
[9] M. Hanke and P. C. Hansen, *Regularization methods for large-scale problems*, Surv. Math. Ind., 3 (1993), pp. 253–315.
[10] P. C. Hansen, *Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35. Software is available in Netlib at the web site `http://www.netlib.org`
[11] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.
[12] C. C. Paige and M. A. Saunders, *LSQR: An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software, 8 (1982), pp. 43–71.
[13] D. L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, J. ACM, 9 (1962), pp. 84–97.
[14] L. Reichel and Q. Ye, *Breakdown-free GMRES for singular systems*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 1001–1021.