SIMPLIFIED GSVD COMPUTATIONS FOR THE SOLUTION OF LINEAR DISCRETE ILL-POSED PROBLEMS

L. DYKES* AND L. REICHEL†

Abstract. The generalized singular value decomposition (GSVD) often is used to solve Tikhonov regularization problems with a regularization matrix without exploitable structure. This paper describes how the standard methods for the computation of the GSVD of a matrix pair can be simplified in the context of Tikhonov regularization. Also, other regularization methods, including truncated GSVD, are considered. We compare the computational efforts required by the simplified GSVD method and the A-weighted generalized inverse introduced by Eldén.

Key words. ill-posed problem, generalized singular value decomposition, Tikhonov regularization, TGSVD

1. Introduction. We consider the solution of least-squares problems

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

with a matrix that has many singular values of different orders of magnitude close to the origin. Minimization problems with this kind of matrix often are referred to as discrete ill-posed problems. They arise, for instance, when discretizing ill-posed problems, such as Fredholm integral equations of the first kind with a smooth kernel. The vector $\boldsymbol{b}$ is in typical applications in science and engineering contaminated by an unknown error $\boldsymbol{e}\in\mathbb{R}^m$ stemming from measurement inaccuracies. We will refer to the vector $\boldsymbol{e}$ as "noise."

The presence of singular values close to the origin makes the solution of the least-squares problem (1.1) sensitive to the noise $\boldsymbol{e}$, as well as to round-off errors introduced during the solution process. A common approach to remedy this difficulty is to replace the least-squares problem by a nearby problem that is less sensitive to perturbations. One of the most popular replacement approaches is known as Tikhonov regularization, which replaces (1.1) by a penalized least-squares problem of the form

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{\|A\boldsymbol{x}-\boldsymbol{b}\|^2+\mu^2\|L\boldsymbol{x}\|^2\right\}. \tag{1.2}$$

Throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm or the associated induced matrix norm. The matrix $L\in\mathbb{R}^{p\times n}$ is referred to as the regularization matrix and the scalar $\mu\geq 0$ as the regularization parameter; see, e.g., Engl et al. [10] and Hansen [15] for discussions on Tikhonov regularization. The number of rows, p, of L generally is smaller than or equal to n, but regularization matrices with $p>n$ also find applications. The matrix L is assumed to be chosen so that

$$\mathcal{N}(A)\cap\mathcal{N}(L)=\{\boldsymbol{0}\}, \tag{1.3}$$

where $\mathcal{N}(M)$ denotes the null space of the matrix M. Then the Tikhonov minimization problem (1.2) has the unique solution

$$\boldsymbol{x}_\mu := (A^TA+\mu^2L^TL)^{-1}A^T\boldsymbol{b} \tag{1.4}$$

[0]Version May 6, 2013

*Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA, and University School, Hunting Valley, OH 44022, USA. E-mail: ldykes@math.kent.edu.

†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.

for any $\mu > 0$. Here and below the superscript ^{T} denotes transposition.

Let $\breve{\boldsymbol{b}}$ denote the unknown error-free vector associated with $\boldsymbol{b}$, i.e.,

$$\boldsymbol{b} = \breve{\boldsymbol{b}} + \boldsymbol{e}. \tag{1.5}$$

We will throughout this paper assume that the linear system of equations with the unknown error-free right-hand side

$$A\boldsymbol{x} = \breve{\boldsymbol{b}} \tag{1.6}$$

is consistent and we denote its solution of minimal Euclidean norm by $\breve{\boldsymbol{x}}$.

It is usually part of the solution process for (1.2) to determine a value of the parameter $\mu > 0$ so that the Tikhonov solution (1.4) is an as accurate as possible approximation of $\breve{\boldsymbol{x}}$. There are many methods available for trying to determine such a value of μ, including the discrepancy principle, the L-curve criterion, generalized cross validation, and extrapolation; see [4, 5, 15, 16] for examples and discussions. All of these methods typically require the evaluation of $\|A\boldsymbol{x}_\mu - \boldsymbol{b}\|$ and possibly also of $\|\boldsymbol{x}_\mu\|$ for several values of μ. For instance, when a bound for the error

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

is known, the discrepancy principle can be applied and prescribes that the regularization parameter $\mu > 0$ be chosen so that the associated solution (1.4) of (1.2) satisfies

$$\|A\boldsymbol{x}_\mu - \boldsymbol{b}\| = \eta\varepsilon. \tag{1.7}$$

Here $\eta \geq 1$ is a user-supplied parameter, which usually is chosen to be fairly close to unity. The norm $\|A\boldsymbol{x}_\mu - \boldsymbol{b}\|$ is referred to as the discrepancy; see, e.g., [10, 15] for discussions on the discrepancy principle. The determination of $\mu > 0$ so that (1.7) holds demands the application of a zero-finder to the nonlinear function

$$\phi(\mu) := \|\boldsymbol{b} - A\boldsymbol{x}_\mu\|^2 - \eta^2\varepsilon^2. \tag{1.8}$$

The zero-finder generally requires the evaluation of $\phi(\mu)$ for several values of μ, and this makes repeated solution of (1.2) for different values of μ necessary.

To reduce the computational effort required for evaluating $\|A\boldsymbol{x}_\mu - \boldsymbol{b}\|$ for several values of μ and to gain insight into the minimization problem (1.2), one often first computes the generalized singular value decomposition (GSVD) of the matrix pair $\{A, L\}$. The availability of the GSVD also speeds up the computation of the norm $\|L\boldsymbol{x}_\mu\|$, which is required when μ is determined by the L-curve criterion. The GSVD of the matrix pair $\{A, L\}$ is given by

$$\begin{aligned} A &= U\mathrm{diag}\big[\sigma_1', \sigma_2', \ldots, \sigma_p', \underbrace{1, 1, \ldots, 1}_{n-p}\big]X^{-1}, \\ L &= V\Big[\mathrm{diag}\left[\rho_1', \rho_2', \ldots, \rho_p'\right], \underbrace{\mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}}_{n-p}\Big]X^{-1}, \end{aligned} \tag{1.9}$$

where the matrices $U \in \mathbb{R}^{m\times n}$ and $V \in \mathbb{R}^{p\times p}$ have orthonormal columns, $X^{-1} \in \mathbb{R}^{n\times n}$ is nonsingular, and the scalars σ_j' and ρ_j' are ordered and scaled so that

$$0 \leq \sigma_1' \leq \sigma_2' \leq \ldots \leq \sigma_p' \leq 1, \qquad 0 \leq \rho_p' \leq \rho_{p-1}' \leq \ldots \leq \rho_1' \leq 1$$

with

$$(\sigma'_j)^2 + (\rho'_j)^2 = 1, \qquad 1 \leq j \leq p;$$

see, e.g., [12, 13, 15] for details. Substituting the decompositions (1.9) into (1.2) gives an equivalent minimization problem with A and L replaced by diagonal matrices. This makes it possible to evaluate the discrepancy $\|A\boldsymbol{x}_\mu - \boldsymbol{b}\|$ and norm $\|L\boldsymbol{x}_\mu\|$ in only $\mathcal{O}(n)$ arithmetic floating point operations (flops) for each value of μ. The quotients σ'_j/ρ'_j are referred to as generalized singular values of the matrix pair $\{A, L\}$; they provide insight into the sensitivity of the solution $\boldsymbol{x}_\mu$ to perturbations. The generalized singular values are the singular values of AL^{-1} when the inverse L^{-1} exists.

The main drawback of the GSVD is that it is fairly expensive to compute for matrices of moderate or large sizes. Bai [1, Table 5.1] shows the leading term in the flop counts for several GSVD algorithms when $A, L \in \mathbb{R}^{n\times n}$. The counts for some of the algorithms depend on the number of sweeps, j, required. When this number is small, Paige's algorithm has the smallest leading coefficient, $(5.3 + 15j)n^3$, which generally is at least $35.3n^3$, but could be much larger. This count is for the computation of all matrices in (1.9). However, savings can be made by observing that the matrices U and V are not explicitly needed when solving (1.2); only the matrix-vector product $U^T\boldsymbol{b}$ is required and can be formed "on the fly," while the matrix V is not required at all. Further savings can be achieved by reorganizing the computations. It is the purpose the present paper to discuss ways to organize the computations to reduce the flop count when the GSVD is applied in Tikhonov regularization (1.2) or in other regularization methods, including the truncated GSVD (TGSVD) method.

The GSVD computations can be sidestepped completely by determining the A-weighted generalized inverse

$$L_A^\dagger := (I - (A(I - L^\dagger L))^\dagger A)L^\dagger \tag{1.10}$$

of L introduced by Eldén [9]. This matrix allows the transformation of (1.2) to the equivalent problem

$$\min_{\boldsymbol{y}\in\mathbb{R}^n} \left\{ \|AL_A^\dagger \boldsymbol{y} - \bar{\boldsymbol{b}}\|^2 + \mu^2\|\boldsymbol{y}\|^2 \right\}, \tag{1.11}$$

where $\bar{\boldsymbol{b}} = P^\perp_{A\mathcal{N}(L)}\boldsymbol{b}$ and $P^\perp_{A\mathcal{N}(L)}$ denotes the orthogonal projector onto the complement of $A\mathcal{N}(L)$; see Section 3 for details. The singular values of $AL_A^\dagger$ are the generalized singular values of the matrix pair $\{A, L\}$. Thus, the computation of the singular value decomposition (SVD) of $AL_A^\dagger$ can be considered equivalent to the computation of the GSVD of $\{A, L\}$. The application of $L_A^\dagger$ is attractive when L has a structure that allows inexpensive determination of $AL_A^\dagger$. This is the case when L is a banded matrix or an orthogonal projection operator; see [9, 18]. However, there are many regularization matrices of interest without such a structure; see, e.g., [6, 8], for examples. We are interested in comparing the use of the GSVD, our proposed simplified GSVD method, and the use of $L_A^\dagger$ for the latter kind of matrices.

This paper is organized as follows. Section 2 describes simplifications of the GSVD computations that can be made when we are interested in the generalized singular values of the matrix pair $\{A, L\}$ and in the solution of (1.2). The use of the A-weighted generalized inverse $L_A^\dagger$ is discussed in Section 3. We will compare the leading coefficient in the flop count for these methods. Section 4 discusses regularization methods different from Tikhonov regularization (1.2) for which our simplified GSVD also is well suited. This includes generalizations to Tikhonov minimization problems in general form of a modified Tikhonov regularization method for problems in standard form proposed in [11]. A few numerical examples are presented in Section 5 and concluding remarks can be found in Section 6. We note that the reduction of the matrix pair $\{A, L\}$ described in [8] has a lower flop count for large problems than the methods of the present paper, but does not determine the generalized singular values of $\{A, L\}$. Knowledge of the generalized singular values may provide important insight into the problem (1.2) and allows regularization by TGSVD.

2. Simplified GSVD computations. This section describes an approach for reducing the matrix pair $\{A, L\}$ to a pair of simpler matrices, that are well suited for use in Tikhonov regularization. Our initial computations are identical with those of several algorithms for computing the GSVD of $\{A, L\}$. A thorough discussion of such algorithms is provided by Bai [1]; see also [2]. All flop counts are for the case when $m = n = p$.

The first step of many GSVD algorithms is to compute the QR factorization

$$\begin{bmatrix} A \\ L \end{bmatrix} = QR, \quad Q = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}, \quad Q_1 \in \mathbb{R}^{m\times n}, \quad Q_2 \in \mathbb{R}^{p\times n}, \quad R \in \mathbb{R}^{n\times n}, \tag{2.1}$$

where Q has orthonormal columns and R is upper triangular. It follows from (1.3) that R is nonsingular. The computation of the QR factorization with Householder matrices requires about $3\frac{1}{3}n^3$ flops; see [12, Section 5.2.1]. This yields Q in factored form. We will need elements of the matrix Q_1, but not of Q_2. Given Q in factored form, the entries of Q_1 can be determined in $3\frac{1}{3}n^3$ flops by computing them from bottom up.

Many GSVD algorithms proceed by determining the singular value decomposition (SVD) of Q_1. The method of this subsection also does this. Let

$$Q_1 = U_1 \Sigma_1 V_1^T, \tag{2.2}$$

where the matrices

$$U_1 = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m] \in \mathbb{R}^{m\times m} \quad \text{and} \quad V_1 = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n] \in \mathbb{R}^{n\times n}$$

are orthogonal, and the (possibly rectangular) diagonal matrix

$$\Sigma_1 = \mathrm{diag}[\sigma_1, \sigma_2, \ldots, \sigma_n] \in \mathbb{R}^{m\times n}$$

contains the singular values of Q_1 ordered according to

$$1 \geq \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_n = 0. \tag{2.3}$$

Thus, the matrix Q_1 has rank r.

The matrix U_1 does not have to be computed explicitly, only the matrix-vector product $U_1^T \boldsymbol{b}$ is required. This matrix-vector product can be determined during the computation of the SVD (2.2) without explicitly forming U_1. The operation count for computing Σ_1, V_1, and $U_1^T \boldsymbol{b}$ is about $12n^3$ flops when $m = n = p$; see [12, p. 254].

We obtain the decomposition

$$Q = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} U_1 & 0 \\ 0 & I_p \end{bmatrix} \begin{bmatrix} \Sigma_1 \\ Q_2 V_1 \end{bmatrix} V_1^T, \tag{2.4}$$

where I_p denotes the $p \times p$ identity matrix. The fact that Q has orthonormal columns yields

$$I_n = \Sigma_1^T \Sigma_1 + V_1^T Q_2^T Q_2 V_1. \tag{2.5}$$

PROPOSITION 2.1. *Let the σ'_k and ρ'_k be the entries of the diagonal matrices of the GSVD (1.9) of the matrix pair $\{A, L\}$. Then the singular values σ_j of the matrix Σ_1 given by (2.2) satisfy*

$$\begin{array}{rcll} \sigma_j & = & 1, & j = 1, 2, \ldots, n-p, \\ \sigma_{n-p+j} & = & \sigma'_{p+1-j}, & j = 1, 2, \ldots, p. \end{array}$$

Moreover, $V_1^T Q_2^T Q_2 V_1 = \mathrm{diag}[\rho_1^2, \rho_2^2, \ldots, \rho_n^2]$ with

$$\begin{array}{rcll} \rho_j & = & 0, & j = 1, 2, \ldots, n-p, \\ \rho_{n-p+j} & = & \rho'_{p+1-j}, & j = 1, 2, \ldots, p. \end{array}$$

The property (2.3) is equivalent to $\sigma'_j = 0$ for $j = 1, 2, \ldots, n-r$. The quotients $\gamma_k = \sigma_k/\rho_k$ are generalized singular values of the matrix pair $\{A, L\}$. We define $\gamma_k = \infty$ when $\rho_k = 0$. Then $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_r > \gamma_{r+1} = \ldots = \gamma_n = 0$.

Proof. It follows from (2.5) that the matrix $Q_2 V_2$ has orthogonal columns. The proposition now is a consequence of (1.9). □

Substituting the decompositions (2.1) and (2.4) into (1.2) gives the equivalent least-squares problem

$$\min_{\boldsymbol{y} \in \mathbb{R}^n} \left\| \begin{bmatrix} \Sigma_1 \\ \mu Q_2 V_1 \end{bmatrix} \boldsymbol{y} - \begin{bmatrix} U_1^T \boldsymbol{b} \\ \boldsymbol{0} \end{bmatrix} \right\|, \quad \boldsymbol{y} = V_1^T R \boldsymbol{x}.$$

Using (2.5), the associated normal equations can be expressed as

$$\left(\mu^2 I_n + (1-\mu^2)\Sigma_1^T \Sigma_1\right) \boldsymbol{y} = \Sigma_1^T U_1^T \boldsymbol{b}. \tag{2.6}$$

We denote the solution by $\boldsymbol{y}_\mu$. The desired solution $\boldsymbol{x}_\mu$ of (1.2) is given by

$$\boldsymbol{x}_\mu = R^{-1} V_1 \boldsymbol{y}_\mu. \tag{2.7}$$

It is generally advantageous, because it may yield a computed approximate solution of higher accuracy, to solve a least-squares problem rather than the associated normal equations. We therefore calculate $\boldsymbol{y}_\mu$ by solving the least-squares problem

$$\min_{\boldsymbol{y} \in \mathbb{R}^n} \left\| \begin{bmatrix} \delta \Sigma_1 \\ \mu I_n \end{bmatrix} \boldsymbol{y} - \begin{bmatrix} \frac{1}{\delta} U_1^T \boldsymbol{b} \\ \boldsymbol{0} \end{bmatrix} \right\|, \quad \delta := \sqrt{1-\mu^2}, \tag{2.8}$$

associated with the normal equations (2.6). The solution is computed by first transforming the matrix to diagonal form with the aid of (at most) n Givens rotations. The evaluation of the solution for a fixed value of $\mu > 0$ requires only $\mathcal{O}(n)$ flops. In many applications, the regularization parameter μ is smaller than unity; otherwise the problem should be rescaled to achieve this to avoid having to use complex arithmetic.

In order to determine a suitable value of the regularization parameter such that the associated solution $\boldsymbol{x}_\mu$ satisfies the discrepancy principle, we have to compute the zero of the function (1.8). Using (2.1), (2.2), and (2.7) gives

$$\phi(\mu) = \|\Sigma_1 \boldsymbol{y}_\mu - U_1^T \boldsymbol{b}\|^2 - \eta^2 \varepsilon^2, \tag{2.9}$$

which yields

$$\phi(1/\mu) = \sum_{j=1}^{n} \frac{\left(1-\sigma_j^2\right)^2}{\left(\mu^2\sigma_j^2 + \left(1-\sigma_j^2\right)\right)^2}\, \tilde{b}_j^2 + \sum_{j=n+1}^{m} \tilde{b}_j^2 - \eta^2\varepsilon^2,$$

where $\tilde{\boldsymbol{b}} = [\tilde{b}_1, \tilde{b}_2, \ldots, \tilde{b}_m]^T := U_1^T\boldsymbol{b}$. It follows from this representation that the function $\psi(\mu) := \phi(1/\sqrt{\mu})$ is decreasing and convex for $\mu > 0$ with

$$\psi(0) = \|\boldsymbol{b}\|^2 - \eta^2\varepsilon^2, \qquad \lim_{\mu\to\infty} \psi(\mu) = \|P_{\mathcal{N}(Q_1^T)}\boldsymbol{b}\|^2 - \eta^2\varepsilon^2, \tag{2.10}$$

where $P_{\mathcal{N}(Q_1^T)}$ denotes the orthogonal projection of onto $\mathcal{N}(Q_1^T)$. Generally, it holds that $\|P_{\mathcal{N}(Q_1^T)}\boldsymbol{b}\| \ll \eta\varepsilon$. We will assume that $\|P_{\mathcal{N}(Q_1^T)}\boldsymbol{b}\| < \eta\varepsilon$. Then ψ has a unique finite zero. It can be computed in a variety of ways. Here we only note that Newton's method with initial approximation $\mu_0 = 0$ converges monotonically and quadratically to the zero of ψ.

A slight modification of the above approach is obtained by replacing the (full) singular value decomposition (2.2) by the reduced singular value decomposition

$$Q_1 = \hat{U}_1\hat{\Sigma}_1 V_1^T, \quad \hat{U}_1 \in \mathbb{R}^{m\times n}, \quad V_1 \in \mathbb{R}^{n\times n}, \tag{2.11}$$

where $\hat{U}_1$ has orthonormal columns, V_1 is orthogonal, and

$$\hat{\Sigma}_1 = \mathrm{diag}[\sigma_1, \sigma_2, \ldots, \sigma_n] \in \mathbb{R}^{n\times n}.$$

The least-squares problem (2.8) is replaced by

$$\min_{\boldsymbol{y}\in\mathbb{R}^n} \left\| \begin{bmatrix} \delta\hat{\Sigma}_1 \\ \mu I_n \end{bmatrix} \boldsymbol{y} - \begin{bmatrix} \frac{1}{\delta}\hat{U}_1^T\boldsymbol{b} \\ \boldsymbol{0} \end{bmatrix} \right\|, \quad \delta := \sqrt{1-\mu^2}, \tag{2.12}$$

which has the same solution. However, we cannot evaluate (2.9) and instead compute

$$\hat{\phi}(\mu) := \|\hat{\Sigma}_1\boldsymbol{y}_\mu - \hat{U}_1^T\boldsymbol{b}\|^2 - \eta^2\varepsilon^2. \tag{2.13}$$

We have

$$\hat{\phi}(1/\mu) = \sum_{j=1}^{n} \frac{\left(1-\sigma_j^2\right)^2}{\left(\mu^2\sigma_j^2 + \left(1-\sigma_j^2\right)\right)^2}\tilde{b}_j^2 - \eta^2\varepsilon^2.$$

It is clear that $\hat{\phi}(\mu) \leq \phi(\mu)$ for all $\mu > 0$. Therefore, the zero of $\hat{\phi}$ is larger than or equal to the zero of ϕ. Thus, letting the regularization parameter be the zero of $\hat{\phi}$ may give more, but not less, regularization than when the zero of ϕ is used.

The main arithmetic work for the method of this section is required for the QR factorization (2.1), the formation of Q_1, and the computation of Σ_1 and V_1. For large problems, the dominating term in the flop count therefore is $18\frac{2}{3}n^3$, which is smaller than for the GSVD; cf. Section 1.

We note that when the computations are carried out using the standard MATLAB function `svd`, then the full SVD of the matrix Q_1 has to be computed. This increases the flop count from about $12n^3$ to about $22n^3$, and increases the total flop count to about $28\frac{2}{3}n^3$. This is still less than the flop count for the standard GSVD method for large n.

Having determined the zero μ of the function (2.9) and computed the associated solution $\boldsymbol{y}_\mu$ of (2.8), or calculated the zero μ of the function (2.13) and the allied solution $\boldsymbol{y}_\mu$ of (2.12), we evaluate the solution $\boldsymbol{x}_\mu$ of (1.2) using (2.7).

We conclude this section by relating the matrices in the factorization

$$\begin{bmatrix} A \\ L \end{bmatrix} = \begin{bmatrix} U_1\Sigma_1 \\ Q_2V_1 \end{bmatrix} V_1^T R, \tag{2.14}$$

obtained from (2.1) and (2.4), and in the GSVD (1.9). First observe that the matrix $V_1^T Q_2^T Q_2 V_1 \in \mathbb{R}^{n\times n}$ in (2.5) is of rank $\ell \leq p$. It follows from (2.3) that

$$\left\{ \begin{array}{rrcll} & \sigma_j & = & 1, & j = 1, 2, \ldots, n-\ell, \\ 1 > & \sigma_j & > & 0, & j = n-\ell+1, n-\ell+2, \ldots, r, \\ & \sigma_j & = & 0, & j = r+1, r+2, \ldots, n; \end{array} \right. \tag{2.15}$$

see also Proposition 2.1. Substitution into the normal equations (2.6) shows that the components of the solution $\boldsymbol{y}_\mu = [y_1, y_2, \ldots, y_n]^T$ are given by

$$\left\{ \begin{array}{rcll} y_j & = & \boldsymbol{u}_j^T \boldsymbol{b}, & j = 1, 2, \ldots, n-\ell, \\ y_j & = & \dfrac{\sigma_j}{\mu^2 + (1-\mu^2)\sigma_j^2}\, \boldsymbol{u}_j^T \boldsymbol{b}, & j = n-\ell+1, n-\ell+2, \ldots, r, \\ y_j & = & 0, & j = r+1, r+2, \ldots, n, \end{array} \right. \tag{2.16}$$

where we have assumed that $\mu > 0$.

Let Π denote the reversal permutation matrix, i.e., the matrix $M\,\Pi$ is made up of the columns of the matrix M in reverse order. Comparing the GSVD (1.9) and the matrices in the factorization (2.14) shows that we may choose

$$X^{-1} = V_1^T R\,\Pi, \qquad \sigma_j' = \sigma_{n+1-j}, \quad j = 1, 2, \ldots, n,$$

and the first n columns of $U_1\,\Pi$ are the columns of the matrix U in (1.9). We remark that since the singular values of U_1 might not be distinct, the GSVD of the matrix pair $\{A, L\}$ is not guaranteed to be unique.

3. Solution by application of $L_A^\dagger$. The solution $\boldsymbol{x}_\mu$ of (1.2) can be determined from the solution $\boldsymbol{y}_\mu$ of (1.11) via

$$\boldsymbol{x}_\mu = L_A^\dagger \boldsymbol{y}_\mu + \boldsymbol{x}_0,$$

where $\boldsymbol{x}_0 = (A(I - L^\dagger L))^\dagger \boldsymbol{b}$. This section discusses the flop count required to solve (1.11) by using the SVD of the matrix $AL_A^\dagger$. The relation

$$\|A\boldsymbol{x}_\mu - \boldsymbol{b}\| = \|AL_A^\dagger \boldsymbol{y}_\mu - \bar{\boldsymbol{b}}\| \tag{3.1}$$

is helpful when determining μ by the discrepancy principle (1.7), because it shows that it suffices to determine the discrepancy of the transformed problem; see, e.g., [9] for a proof of (3.1).

The solution $\boldsymbol{y}_\mu$ of (1.11) can be conveniently computed for several values of μ by first determining the SVD of $AL_A^\dagger$. The singular values provide insight into the minimization problem (1.11). Hansen [13, Section 6] has shown that the singular values of $AL_A^\dagger$ are the generalized singular values of the matrix pair $\{A, L\}$.

We are interested in the computational effort required to compute the SVD of $AL_A^\dagger$ in the situation when the regularization matrix $L \in \mathbb{R}^{p\times n}$ is a general matrix. We assume for notational simplicity that $p \leq n$. Let

$$L = \widetilde{U}\widetilde{\Sigma}\widetilde{V}^T \tag{3.2}$$

be an SVD of L with orthogonal matrices $\widetilde{U} \in \mathbb{R}^{p\times p}$ and $\widetilde{V} \in \mathbb{R}^{n\times n}$, and the rectangular diagonal matrix

$$\widetilde{\Sigma} = \mathrm{diag}[\widetilde{\sigma}_1, \widetilde{\sigma}_2, \ldots, \widetilde{\sigma}_p] \in \mathbb{R}^{p\times n}.$$

The singular values are ordered according to

$$\widetilde{\sigma}_1 \geq \widetilde{\sigma}_2 \geq \ldots \geq \widetilde{\sigma}_{p-\ell} > \widetilde{\sigma}_{p-\ell+1} = \ldots = \widetilde{\sigma}_p = 0.$$

The number of vanishing singular values, ℓ, is assumed to be small and independent of n. It is clear from (1.11) that L can be replaced by

$$\widetilde{L} := \widetilde{U}^T L = \widetilde{\Sigma}\widetilde{V}^T$$

without changing the solution of the problem. Therefore, only the matrices $\widetilde{\Sigma}$ and $\widetilde{V}$ in the SVD of L have to be computed. Their computation requires about $12n^3$ flops when $p \approx n$; see [12, p. 254].

We turn to the computation of the matrix $AL_A^\dagger$. Split $\widetilde{V} = [\widetilde{\boldsymbol{v}}_1, \widetilde{\boldsymbol{v}}_2, \ldots, \widetilde{\boldsymbol{v}}_n]$ into

$$\widetilde{V}_1 = [\widetilde{\boldsymbol{v}}_1, \widetilde{\boldsymbol{v}}_2, \ldots, \widetilde{\boldsymbol{v}}_{p-\ell}], \qquad \widetilde{V}_0 = [\widetilde{\boldsymbol{v}}_{p-\ell+1}, \widetilde{\boldsymbol{v}}_{p-\ell+2}, \ldots, \widetilde{\boldsymbol{v}}_n].$$

Thus, the columns of $\widetilde{V}_0$ form an orthonormal basis for $\mathcal{N}(L)$. Compute the QR factorization

$$A\widetilde{V}_0 = \widetilde{Q}\widetilde{R},$$

where the matrix $\widetilde{Q} \in \mathbb{R}^{m\times(n+\ell-p)}$ has orthonormal columns and the matrix $\widetilde{R} \in \mathbb{R}^{(\ell+n-p)\times(\ell+n-p)}$ is upper triangular. It follows from (1.3) that $\widetilde{R}$ is nonsingular. Since

$$A(I - L^\dagger L) = A\widetilde{V}_0\widetilde{V}_0^T = \widetilde{Q}\widetilde{R}\widetilde{V}_0^T,$$

we have

$$(A(I - L^\dagger L))^\dagger = \widetilde{V}_0\widetilde{R}^{-1}\widetilde{Q}^T. \tag{3.3}$$

Using this formula, the computational effort to determine $\boldsymbol{x}_0 = \widetilde{V}_0\widetilde{R}^{-1}\widetilde{Q}^T\boldsymbol{b}$ and $\bar{\boldsymbol{b}} = (I - \widetilde{Q}\widetilde{Q}^T)\boldsymbol{b}$ is negligible. Substitution (3.3) into (1.10) and replacing the trailing $L^\dagger$ by $\widetilde{L}^\dagger$ yields

$$L_A^\dagger = (I - \widetilde{V}_0\widetilde{R}^{-1}\widetilde{Q}^T A)\Big[\widetilde{V}_1\mathrm{diag}[1/\widetilde{\sigma}_1, 1/\widetilde{\sigma}_2, \ldots, 1/\widetilde{\sigma}_{p-\ell}], \underbrace{\mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}}_{\ell}\Big]$$

and

$$AL_A^\dagger = (I - \widetilde{Q}\widetilde{Q}^T)A\Big[\widetilde{V}_1\mathrm{diag}[1/\widetilde{\sigma}_1, 1/\widetilde{\sigma}_2, \ldots, 1/\widetilde{\sigma}_{p-\ell}], \underbrace{\mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}}_{\ell}\Big],$$

where the trailing matrix is of size $(p-\ell)\times\ell$.

The dominating work for forming $AL_A^\dagger$ is, besides the computation of the matrices $\widetilde{\Sigma}$ and $\widetilde{V}$, the evaluation of the matrix-matrix product $A\widetilde{V}_1$, which requires about $2n^3$ flops when n is large, $p\approx n$, and ℓ is independent of n. Thus, the flop count for determining $AL_A^\dagger$ is about $14n^3$. When computing the SVD $AL_A^\dagger=\breve{U}\breve{\Sigma}\breve{V}^T$, we have to form $\breve{\Sigma}$ and $\breve{V}^T$, but can apply $\breve{U}^T$ to $\bar{\boldsymbol{b}}$ on the fly. The computation of $\breve{U}^T\bar{\boldsymbol{b}}$, $\breve{\Sigma}$, and $\breve{V}^T$ requires about $12n^3$ flops; see [12, p. 254]. The total flop count for computing $AL_A^\dagger$, $\breve{U}^T\bar{\boldsymbol{b}}$, $\breve{\Sigma}$, and $\breve{V}^T$ therefore is about $26n^3$. We conclude that this approach is faster than evaluating the GSVD of the matrix pair $\{A,L\}$, but slower than the approach of Section 2, when n is large.

We remark that when the computations are carried out in MATLAB using the function `svd`, all matrices in the SVD of L (3.2) and $AL_A^\dagger$ have to be computed. This increases the flop count to about $46n^3$.

4. Other regularization methods. This section first discusses the Tikhonov and TGSVD regularization methods in terms of filter factors using the simplified GSVD computations of Section 2. Subsequently, new regularization methods that combine features of Tikhonov and TGSVD regularization are described. These methods are generalizations of the modified Tikhonov regularization method for problems in standard form described in [11] to problems in general form. We will use the notation of Section 2.

It is convenient to express regularization methods with the aid of filter factors; see, e.g., [7, 15]. We obtain from (2.7) and (2.16) that the solution $\boldsymbol{x}_\mu$ of (1.2) can be written as

$$\boldsymbol{x}_{\text{filtered}}=R^{-1}\sum_{j=1}^{r}f_j\frac{\boldsymbol{u}_j^T\boldsymbol{b}}{\sigma_j}\boldsymbol{v}_j, \tag{4.1}$$

with the filter factors

$$f_j=\begin{cases}1, & 1\le j\le n-\ell,\\ \dfrac{\sigma_j^2}{\mu^2+(1-\mu^2)\sigma_j^2}, & n-\ell<j\le r.\end{cases} \tag{4.2}$$

We note for future reference that the nontrivial filter factors can be written as

$$f_j=\frac{\gamma_j^2}{\gamma_j^2+\mu^2},\quad n-\ell<j\le r, \tag{4.3}$$

where the γ_j are the generalized singular values of the matrix pair $\{A,L\}$; cf. Proposition 2.1.

The TGSVD method regularizes by determining approximate solutions $\boldsymbol{x}_k$ of (1.1) that do not contain components of generalized singular vectors $R^{-1}\boldsymbol{v}_j$ associated with the smallest singular values σ_j of Q_1. For instance, the TGSVD solution

$$\boldsymbol{x}_k=R^{-1}\sum_{j=1}^{k}\frac{\boldsymbol{u}_j^T\boldsymbol{b}}{\sigma_j}\boldsymbol{v}_j,\qquad k\le r, \tag{4.4}$$

does not contain terms with the $n-k$ smallest singular values $\sigma_{k+1},\sigma_{k+2},\dots,\sigma_n$. It can be expressed in the form (4.1) with the filter factors

$$f_j=\begin{cases}1, & 1\le j\le k,\\ 0, & k<j\le r.\end{cases} \tag{4.5}$$

Typically, k is chosen larger than $n-\ell$ so that solution components in $\mathcal{N}(L)$ are not damped. Differently from the situation when computing the Tikhonov solution (1.4), the computation of a TGSVD solution does not implicitly involve the singular values of the matrix Q_2 in (2.1). Therefore, the computations can be carried out independently of the scaling of the least-squares problem (1.1) and of the regularization matrix L. This differs from the situation for Tikhonov regularization, which requires appropriate scaling; see the discussion following (2.8).

The discrepancy principle prescribes that the number of terms, k, in (4.4) be chosen as small as possible such that the inequality

$$\|A\boldsymbol{x}_k - \boldsymbol{b}\| \leq \eta\epsilon \tag{4.6}$$

holds; cf. (1.7). We denote this value of k by $k_{\rm disc}$ and tacitly assume that $\eta \geq 1$ is chosen large enough so that $k_{\rm disc} \leq r$. This condition is easily verified during computations. A simple way to determine the index $k_{\rm disc}$ is to evaluate the right-hand side of

$$\|A\boldsymbol{x}_k - \boldsymbol{b}\|^2 = \sum_{j=k+1}^{m} (\boldsymbol{u}_j^T\boldsymbol{b})^2.$$

Discussions of the TGSVD and the related truncated singular value decomposition can be found in [10, 13, 15, 17].

We turn to a modification of Tikhonov regularization. It generalizes the modified Tikhonov regularization method for problems in standard form proposed in [11] to Tikhonov regularization problems in general form. Note that the TGSVD filter factors (4.5) with $k = k_{\rm disc}$ do not damp the first $k_{\rm disc}$ terms in (4.1). However, all terms in (4.1) with index $n-\ell < j \leq r$ are damped when Tikhonov regularization is applied with the filter factors (4.3). This suggests that we may reduce the damping in Tikhonov regularization and still obtain an accurate approximation of the desired solution $\breve{\boldsymbol{x}}$ of (1.6). Also note that when $0 \leq \gamma_j \ll \mu < 1$, the corresponding Tikhonov filter factor f_j given by (4.3) is about γ_j^2/μ^2. We therefore propose the modified Tikhonov filter factors

$$f_j = \begin{cases} 1, & 1 \leq j \leq k_{\mu_{\rm disc}}, \\ \dfrac{\gamma_j^2}{\mu_{\rm disc}^2}, & k_{\mu_{\rm disc}} < j \leq r, \end{cases} \tag{4.7}$$

where $k_{\mu_{\rm disc}}$ is the smallest index k such that $\gamma_k \geq \mu_{\rm disc}$.

To gain some insight into the properties of the regularization method determined by these filter factors, we introduce the matrix

$$\widetilde{\Gamma} = \mathrm{diag}[\gamma_{n-\ell+1}, \gamma_{n-\ell+2}, \ldots, \gamma_r].$$

It follows from (2.15) that $0 < \gamma_{n-\ell+1} \leq \gamma_{n-\ell+2} \leq \ldots \leq \gamma_r < 1$. Let the vector $\widetilde{\boldsymbol{y}} = [y_{n-\ell+1}, y_{n-\ell+2}, \ldots, y_r]^T$ be defined by (2.16) with $\mu = \mu_{\rm disc}$. Then $\widetilde{\boldsymbol{y}}$ satisfies

$$\left(\widetilde{\Gamma}^2 + \mu_{\rm disc}^2 I_{r-n+\ell}\right)\widetilde{\boldsymbol{y}} = \widetilde{\Gamma}\left[\frac{\boldsymbol{u}_{n-\ell+1}^T\boldsymbol{b}}{\sqrt{1-\sigma_{n-\ell+1}^2}}, \frac{\boldsymbol{u}_{n-\ell+2}^T\boldsymbol{b}}{\sqrt{1-\sigma_{n-\ell+2}^2}}, \ldots, \frac{\boldsymbol{u}_r^T\boldsymbol{b}}{\sqrt{1-\sigma_r^2}}\right]^T. \tag{4.8}$$

This is a Tikhonov regularization problem in standard form. In order for the solution $\widetilde{\boldsymbol{y}}$ to be meaningful in the presence of noise in $\boldsymbol{b}$, it is important that the smallest eigenvalue of the matrix $\widetilde{\Gamma}^2+\mu_{\rm disc}^2 I_{r-n+\ell}$ not be too small. This eigenvalue is bounded below by $\mu_{\rm disc}^2$.

PROPOSITION 4.1. *Replacing the filter factors (4.3) by the corresponding factors (4.7) is equivalent to replacing the matrix* $\widetilde{\Gamma}^2+\mu_{\rm disc}^2 I_{r-n+\ell}$ *in (4.8) by*

$$\widetilde{\Gamma}^2 + D_{\mu_{\rm disc}}, \tag{4.9}$$

where $D_{\mu_{\rm disc}} = {\rm diag}[d_{n-\ell+1}, d_{n-\ell+2}, \dots, d_r]$ *with* $d_j = \max\{0, \mu_{\rm disc}^2 - \gamma_j^2\}$. *The matrix (4.9) is the closest symmetric matrix in the Frobenius norm to the matrix* $\widetilde{\Gamma}^2$ *with smallest eigenvalue* $\mu_{\rm disc}^2$.

Proof. The relation between the filter factors associated with Tikhonov regularization (4.8) and the regularization method obtained by replacing the matrix $\widetilde{\Gamma}^2+\mu_{\rm disc}^2 I_{r-n+\ell}$ in (4.8) by (4.9) follows by direct computations. The closeness result for the matrix (4.9) is proved in [11, Theorem 2.1]. This result can be shown by application of the Wieland-Hoffman theorem. □

The filter factors (4.7) are meaningful independently of the size of $\mu_{\rm disk}$; the latter parameter does not have to be smaller than unity. We will illustrate the performance of the filter factors (4.7) in Section 5.

Application of the filter factors (4.7) requires that the value $\mu_{\rm disc}$ of the regularization parameter μ in (1.2) be computed. We can circumvent this by first determining the index $k_{\rm disc}$ by the discrepancy principle applied to TGSVD, see the discussion following (4.6), and then replacing $\mu_{\rm disc}$ by $\gamma_{k_{\rm disc}}$ in (4.7). This yields the filter factors

$$f_j = \begin{cases} 1, & 1 \le j \le k_{\rm disc}, \\ \dfrac{\gamma_j^2}{\gamma_{k_{\rm disc}}^2}, & k_{\rm disc} < j \le r. \end{cases} \tag{4.10}$$

We used the representation (4.3) of the Tikhonov filter factors as our starting point for deriving regularization methods different from Tikhonov regularization and TGSVD. Similarly, one can derive new regularization methods by starting with the filter factor representation (4.2). When using the latter representation, one has to take into consideration that the σ_j may not scale properly when rescaling A or L. We therefore feel it is easier to work with the representation (4.3).

The discussion in this section has focused on determining regularization parameters with the aid of the discrepancy principle. However, the parameters $\mu_{\rm disc}$ and $k_{\rm disc}$ also can be determined by other methods, including extrapolation, the L-curve and generalized cross validation; see [4, 5, 15, 16] for discussions of these and other methods.

5. Computed examples. We illustrate the computation of approximations of the solution $\breve{\boldsymbol{x}}$ of (1.6) by Tikhonov regularization (1.2) or by one of the regularization methods of Section 4. The regularization parameters are determined by the discrepancy principle. Thus, we let the regularization parameter μ in (1.2) be the zero $\mu_{\rm disc}$ of the function (2.9) and the truncation index $k_{\rm disc}$ for TGSVD is the smallest integer such that (4.6) holds.

The examples show that the use of a suitable regularization matrix L can improve the quality of the computed solution significantly. The choice of the regularization matrix should depend on known properties of the desired solution. We illustrate how suitable regularization matrices can be constructed. These "designer regularization matrices" do not have a structure that allows efficient computations with the A-weighted generalized inverse, $L_A^\dagger$, described in Section 3. We therefore apply the simplified GSVD computations of Section 2.

Several regularization matrices will be used in the computed examples, including the tridiagonal matrix

$$L' = \frac{1}{4}\begin{bmatrix} -1 & 2 & -1 & & & & 0 \\ & -1 & 2 & -1 & & & \\ & & \ddots & \ddots & \ddots & & \\ 0 & & & & -1 & 2 & -1 \end{bmatrix} \in \mathbb{R}^{(n-2)\times n}, \tag{5.1}$$

which is an approximation of a multiple of the second derivative operator in one space-dimension. Its null space is the span of the vectors

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

Let the desired solution $\breve{\boldsymbol{x}}$ of (1.6) be the discretization of a function $\breve{x}(t)$ at equidistant nodes $t_1 < t_2 < \ldots < t_n$. The matrix L' may be a suitable regularization matrix when the function $\breve{x}(t)$ has a significant linear component, because this component is in the the null space of L' and, therefore, is not damped by Tikhonov regularization (1.2) with $L = L'$.

If the desired solution $\breve{\boldsymbol{x}}$ has other known important features, in addition to having a significant linear component, then a regularization matrix of the form

$$L := L'(I - WW^T), \tag{5.2}$$

where the matrix $W = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_\ell] \in \mathbb{R}^{n\times\ell}$, $1 \leq \ell < n$, has orthonormal columns, may be appropriate for a suitable choice of W. We seek to choose W so that functions with the known features are in the null space of $I - WW^T$. This avoids these features being damped by regularization. We have

$$\mathcal{N}(L) = \mathcal{R}(W) \cup \mathcal{N}(L'_{\mathcal{R}(W)^\perp}),$$

where $\mathcal{R}(W)$ denotes the range of W and $L'_{\mathcal{R}(W)^\perp}$ is the restriction of L' to the orthogonal complement of $\mathcal{R}(W)$; see [8, Proposition 3.1].

For instance, let $\breve{\boldsymbol{x}}$ have a significant component of the vector obtained by discretizing the function $\cos(2t)$ at equidistant nodes in the interval $[0, \pi]$. Then it can be advantageous to choose $W \in \mathbb{R}^n$ to be a discretization of this function; see below for illustrations. The flop count of Section 3 applies to the regularization matrix (5.2) and, therefore, the regularization approaches of Sections 2 and 4 require fewer flops for large problem.

Consider the situation when the desired solution $\breve{\boldsymbol{x}} = [\breve{x}(t_1), \breve{x}(t_2), \ldots, \breve{x}(t_n)]^T$ of (1.6) is piecewise smooth with kinks at $\breve{x}(t_\ell)$ and $\breve{x}(t_{\ell+1})$ for some $3 \leq \ell < n$. Then it may be meaningful to apply the regularization matrix

$$L'' = \frac{1}{4}\begin{bmatrix} -1 & 2 & -1 & & & & & & & & 0 \\ & \ddots & \ddots & \ddots & & & & & & & \\ & & -1 & 2 & -1 & & & & & & \\ & & & & & -1 & 2 & -1 & & & \\ & & & & & & \ddots & \ddots & \ddots & & \\ 0 & & & & & & & -1 & 2 & -1 & \end{bmatrix} \in \mathbb{R}^{(n-4)\times n}, \tag{5.3}$$

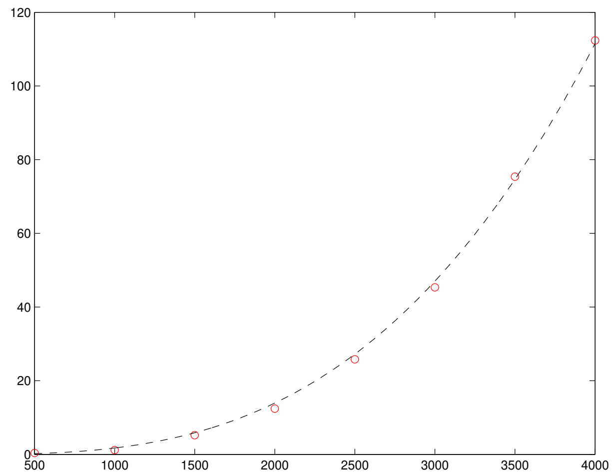

FIG. 5.1. *Example 5.1: Matrix size n versus CPU time in seconds required by the MATLAB function* gsvd*. The measured times are displayed with (red) circles. The continuous graph depicts the cubic monomial cn^3 with $c \approx 1.72 \cdot 10^{-9}$.*

which is obtained by removing rows $\ell - 1$ and ℓ from L'. The null space of L'' is spanned by the vectors

$$\begin{array}{rclrcl} \boldsymbol{n}_1 & = & [\underbrace{1,\ldots,1}_{\ell},\underbrace{0,\ldots,0}_{n-\ell}]^T, & \boldsymbol{n}_2 & = & [\underbrace{0,\ldots,0}_{\ell},\underbrace{1,\ldots,1}_{n-\ell}]^T, \\ \boldsymbol{n}_3 & = & [1,2,\ldots,\ell,\underbrace{0,\ldots,0}_{n-\ell}]^T, & \boldsymbol{n}_4 & = & [\underbrace{0,\ldots,0}_{n-\ell},1,2,\ldots,\ell]^T. \end{array}$$

Analogously to (5.2), we also will apply the regularization matrix

$$L := L''(I - WW^T), \tag{5.4}$$

where $W \in \mathbb{R}^n$ similarly as above is defined by discretizing the function $\cos(2t)$ at equidistant nodes in the interval $[0, \pi]$. For discussions on the importance of choosing suitable regularization matrices and on the design of many such matrices; see, e.g., [3, 6, 19, 21].

All computations were carried out in MATLAB with about 16 significant decimal digits on a standard desktop computer. Our first computed example presents timings that show that the flop counts based on the leading term in the count of arithmetic floating point operations are relevant for problems that can be solved on a standard desktop computer.

Example 5.1. The flop counts reported in Sections 2 and 3 for the different approaches to solve (1.2) show the method of Section 2 to be competitive for large problems for which $L^\dagger$ is expensive to compute. However, flop counts only partially determine the execution time of a computer program. The execution time also is affected by the programming style, whether the code is interpreted or a compiled version is executed, the computer architecture, including communication speed and the amount of fast computer memory available, as well as the number of CPUs. Because the execution time depends on so many variables, we refrain from comparing timings for our interpreted MATLAB codes with the (probably compiled) internal MATLAB function gsvd for computing the GSVD of a matrix pair. Instead, we report

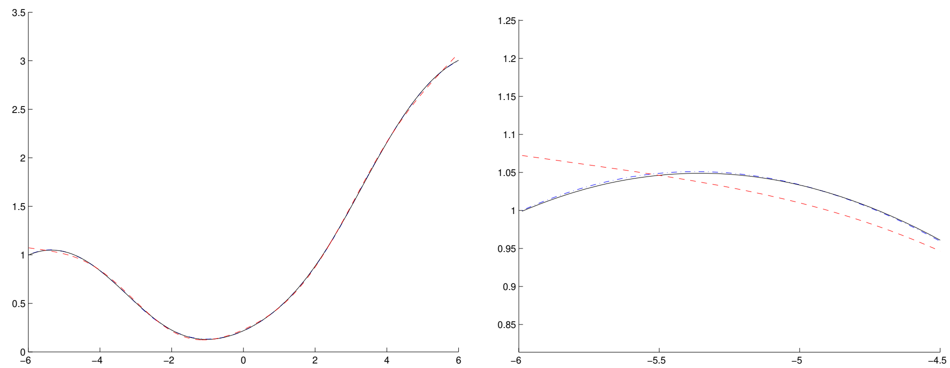

FIG. 5.2. *Example 5.2: Approximate solutions determined by TGSVD with regularization matrices (5.2) (black continuous graph) and (5.1) (red dashed graph), as well as the desired solution* $\breve{\boldsymbol{x}}$ *(blue dashed-dotted graph) for noise-level* $\alpha = 1 \cdot 10^{-2}$. *The graphs on the right-hand side are blow-ups of the ones on the left-hand side.*

timings for the MATLAB gsvd functions when applied to matrix pairs of different sizes. The timings illustrate that it is feasible to compute the GSVD of matrix pairs of sizes up to a few thousand on a standard desktop computer and that the execution time grows cubically with the order, n, of the square matrices in each pair. This indicates that the leading term in the flop counts as a function of n (with $m = p = n$) dominates. Our complexity analysis, which focuses on the leading term, therefore is relevant.

The circles of Figure 5.1 display CPU times for computing the GSVD of matrix pairs $\{A, B\}$ of $n \times n$ matrices for $n = 500, 1000, 1500, \ldots, 4000$ when using the internal MATLAB function gsvd. The timings were determined with the MATLAB commands tic and toc. The figure displays averages over five runs for each n. Fitting a monomial cn^3 to these data in the least-squares sense gives the coefficient $c \approx 1.72 \cdot 10^{-9}$. The continuous graph of Figure 5.1 displays this monomial and shows good agreement between the function $n \to cn^3$ and the measured execution times. This leads us to conclude that the cubic term dominates the flop count. □

The following examples compare the computed approximate solutions of (1.1) determined by the regularization methods discussed in this paper using the regularization matrices (5.1), (5.2), (5.3), (5.4), and $L = I$. The noise $\boldsymbol{e}$ in $\boldsymbol{b}$ has normally distributed entries with mean zero and is normalized to correspond to a chosen noise-level

$$\alpha = \frac{\|\boldsymbol{e}\|}{\|\breve{\boldsymbol{b}}\|}.$$

We let $\eta = 1$ in (1.7) and (4.6).

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

$$\int_{-6}^{6} \kappa(t,s)x(s)ds = g(t), \qquad -6 \le t \le 6, \tag{5.5}$$

with the kernel and solution given by

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

noise-level	regularization matrix	method	$\Vert \boldsymbol{x}_{\text{computed}} - \breve{\boldsymbol{x}} \Vert / \Vert \breve{\boldsymbol{x}} \Vert$
$1 \cdot 10^{-1}$	I	Tikhonov	$6.86 \cdot 10^{-2}$
$1 \cdot 10^{-1}$	I	modified Tikhonov	$6.98 \cdot 10^{-2}$
$1 \cdot 10^{-1}$	I	TGSVD	$6.31 \cdot 10^{-2}$
$1 \cdot 10^{-1}$	I	modified TGSVD	$6.42 \cdot 10^{-2}$
$1 \cdot 10^{-1}$	(5.1)	Tikhonov	$3.35 \cdot 10^{-2}$
$1 \cdot 10^{-1}$	(5.1)	modified Tikhonov	$2.87 \cdot 10^{-2}$
$1 \cdot 10^{-1}$	(5.1)	TGSVD	$1.71 \cdot 10^{-2}$
$1 \cdot 10^{-1}$	(5.1)	modified TGSVD	$1.72 \cdot 10^{-2}$
$1 \cdot 10^{-1}$	(5.2)	Tikhonov	$1.54 \cdot 10^{-2}$
$1 \cdot 10^{-1}$	(5.2)	modified Tikhonov	$1.48 \cdot 10^{-2}$
$1 \cdot 10^{-1}$	(5.2)	TGSVD	$1.45 \cdot 10^{-2}$
$1 \cdot 10^{-1}$	(5.2)	modified TGSVD	$1.44 \cdot 10^{-2}$
$1 \cdot 10^{-2}$	I	Tikhonov	$3.47 \cdot 10^{-2}$
$1 \cdot 10^{-2}$	I	modified Tikhonov	$3.03 \cdot 10^{-2}$
$1 \cdot 10^{-2}$	I	TGSVD	$3.00 \cdot 10^{-2}$
$1 \cdot 10^{-2}$	I	modified TGSVD	$3.00 \cdot 10^{-2}$
$1 \cdot 10^{-2}$	(5.1)	Tikhonov	$1.23 \cdot 10^{-2}$
$1 \cdot 10^{-2}$	(5.1)	modified Tikhonov	$1.15 \cdot 10^{-2}$
$1 \cdot 10^{-2}$	(5.1)	TGSVD	$1.17 \cdot 10^{-2}$
$1 \cdot 10^{-2}$	(5.1)	modified TGSVD	$1.17 \cdot 10^{-2}$
$1 \cdot 10^{-2}$	(5.2)	Tikhonov	$3.52 \cdot 10^{-3}$
$1 \cdot 10^{-2}$	(5.2)	modified Tikhonov	$1.97 \cdot 10^{-3}$
$1 \cdot 10^{-2}$	(5.2)	TGSVD	$1.83 \cdot 10^{-3}$
$1 \cdot 10^{-2}$	(5.2)	modified TGSVD	$1.85 \cdot 10^{-3}$

TABLE 5.1

Example 5.2. Errors in approximate computed solutions $\boldsymbol{x}_{\text{computed}}$ for several regularization methods with different regularization matrices and for two noise-levels.

The right-hand side $g(t)$ is defined by (5.5). This integral equation is discussed by Phillips [20]. We use the MATLAB code phillips from [14] to determine a discretization $A \in \mathbb{R}^{1000 \times 1000}$ by a Galerkin method with orthonormal box functions as test and trial basis functions. The code phillips also provides a discretization of a scaled solution, $\boldsymbol{x}_0 \in \mathbb{R}^{1000}$. We add a discretization of the function $1 + s/6 + \cos(2\pi(1 + s/6))$, $-6 < s < 6$, to the vector $\boldsymbol{x}_0$ to obtain a slowly oscillatory and increasing solution $\breve{\boldsymbol{x}}$. The noise-free right-hand side is given by $\breve{\boldsymbol{b}} = A\breve{\boldsymbol{x}}$. We generate a noise-vector $\boldsymbol{e}$ corresponding to the noise-level α and obtain the contaminated vector $\boldsymbol{b}$ in (1.1) from (1.5).

Table 5.1 reports the relative error in the computed solutions obtained for different regularization matrices and for several regularization methods. The table illustrate that the choice of regularization matrix can be important for the quality of the computed approximate solution. For instance, the tridiagonal regularization matrix (5.1) yields better approximations of the desired solution $\breve{\boldsymbol{x}}$ than $L = I$. The modification (5.2) of the regularization matrix (5.1) takes into account that $\breve{\boldsymbol{x}}$ is slowly oscillating, and improves the quality of the computed approximate solution further. Figure 5.2 shows computed approximate solutions determined with the regularization matrices (5.1) and (5.2) for the noise-level $\alpha = 1 \cdot 10^{-2}$. The relative error in these computed solutions is largest at the end points of the interval of integration and differs by more

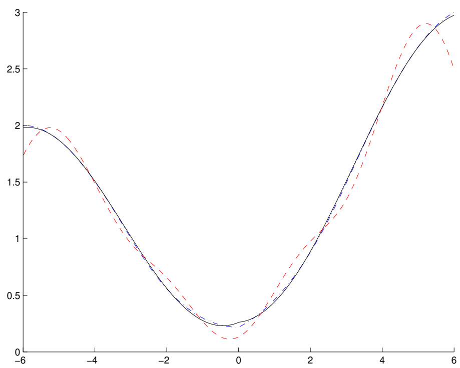

FIG. 5.3. *Example 5.3: Approximate solutions determined by TGSVD with regularization matrices (5.2) (black continuous graph) and $L = I$ (red dashed graph), as well as the desired solution $\breve{\boldsymbol{x}}$ (blue dashed-dotted graph) for noise-level $\alpha = 0.1$.*

than a factor 6. The error in the computed solution determined with the regularization matrix (5.1) is clearly visible. Table 5.1 shows modified Tikhonov (4.7) regularization often to yield a smaller error than the standard Tikhonov method. Moreover, TGSVD and its modification, defined by (4.5) and (4.10), respectively, often furnish the most accurate approximations of $\breve{\boldsymbol{x}}$. We will return to the latter below. Finally, we remark that the size of the regularization parameters for Tikhonov regularization differs significantly for the different regularization matrices. For instance, when $\alpha = 0.1$, the regularization parameter for $L = I$ is smaller than unity, while we scaled the matrix (5.1) by 10^4 and the matrix (5.2) by 10^5 to achieve that $\mu_{\text{disc}} < 1$. □

regularization matrix	method	$\|\boldsymbol{x}_{\text{computed}} - \breve{\boldsymbol{x}}\|/\|\breve{\boldsymbol{x}}\|$
I	Tikhonov	$6.93 \cdot 10^{-2}$
I	modified Tikhonov	$7.13 \cdot 10^{-2}$
I	TGSVD	$7.06 \cdot 10^{-2}$
I	modified TGSVD	$6.82 \cdot 10^{-2}$
(5.1)	Tikhonov	$3.09 \cdot 10^{-2}$
(5.1)	modified Tikhonov	$2.97 \cdot 10^{-2}$
(5.1)	TGSVD	$3.42 \cdot 10^{-2}$
(5.1)	modified TGSVD	$3.16 \cdot 10^{-2}$
(5.3)	Tikhonov	$3.11 \cdot 10^{-2}$
(5.3)	modified Tikhonov	$1.44 \cdot 10^{-2}$
(5.3)	TGSVD	$1.56 \cdot 10^{-2}$
(5.3)	modified TGSVD	$1.46 \cdot 10^{-2}$
(5.4)	Tikhonov	$1.34 \cdot 10^{-2}$
(5.4)	modified Tikhonov	$1.12 \cdot 10^{-2}$
(5.4)	TGSVD	$8.35 \cdot 10^{-3}$
(5.4)	modified TGSVD	$8.95 \cdot 10^{-3}$

TABLE 5.2

Example 5.3. Errors in approximate computed solutions $\boldsymbol{x}_{\text{computed}}$ for several regularization methods with different regularization matrices and noise-level $\alpha = 0.1$.

Example 5.3. We modify the problem considered in the previous example by adding a discretization of the piecewise linear function

$$z(t) = \begin{cases} -\frac{t}{6}, & -6 \leq t \leq 0, \\ 0, & 0 \leq t \leq 6, \end{cases}$$

to the desired solution $\breve{\boldsymbol{x}}$. The right-hand side is adjusted accordingly and contaminated by 10% noise.

Table 5.2 reports the relative error in the computed solutions obtained for different regularization matrices and for several regularization methods. Similarly as Table 5.1, this table shows the importance of the choice of regularization matrix. The quality of the computed approximations is the worst for $L = I$ and the best for L given by (5.4). The improvement in quality of the TGSVD solutions is by more than a factor 8. Figure 5.3 shows these computed approximate solutions. When Tikhonov regularization was used, the regularization matrices (5.1) and (5.3) were scaled by 10^4 and the matrix (5.4) by 10^5 to secure that $\mu_{\text{disc}} < 1$. □

The examples shown, as well as numerous other computed examples, illustrate the importance of choosing a suitable regularization matrix. The more information is available about the desired solution, the better regularization matrix can be constructed. In particular, the examples show that regularization matrices of the form (5.2) and (5.4) may yield more accurate approximations of the desired solution $\breve{\boldsymbol{x}}$ than the standard regularization matrices $L = I$ and (5.1).

We found TGSVD and the modified TGSVD method defined by (4.10) to give the best or near-best approximations of $\breve{\boldsymbol{x}}$ in many examples. These methods do not require the initial least-squares problem (1.2) to be scaled so that the Tikhonov regularization parameter μ is smaller than unity. This scaling is beneficial when solving the least-squares problem (2.8), but is not required when solving the associated normal equations (2.6).

6. Conclusion. Simplifications of the computations of the GSVD of a matrix pair $\{A, L\}$ have been described. The computations are particularly simple when regularization is carried out by the TGSVD or modified TGSVD methods, because then the constraint that the Tikhonov regularization parameter be bounded by one can be ignored. This fact and the competitive performance of these methods, when compared with Tikhonov and modified Tikhonov regularization, suggests that, generally, they be used with the simplified GSVD computations of the present paper.

REFERENCES

[1] Z. Bai, *The CSD, GSVD, their applications and computation*, IMA preprint 958, Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, MN, 1992.

[2] Z. Bai and J. W. Demmel, *Computing the generalized singular value decomposition*, SIAM J. Sci. Comput., 14 (1993), pp. 1464–1486.

[3] C. Brezinski, M. Redivo-Zaglia, G. Rodriguez, and S. Seatzu, *Multi-parameter regularization techniques for ill-conditioned linear systems*, Numer. Math., 94 (2003), pp. 203–228.

[4] C. Brezinski, G. Rodriguez, and S. Seatzu, *Error estimates for linear systems with application to regularization*, Numer. Algorithms, 49 (2008), pp. 85–104.

[5] C. Brezinski, G. Rodriguez, and S. Seatzu, *Error estimates for the regularization of least squares problems*, Numer. Algorithms, 51 (2009), pp. 61–76.

[6] 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.

[7] M. Donatelli and S. Serra-Capizzano, *Filter factor analysis of an iterative multilevel regularization method*, Electron. Trans. Numer. Anal., 29 (2008), pp. 163–177.

[8] L. Dykes and L. Reichel, *On the reduction of Tikhonov minimization problems and the construction of regularization matrices*, Numer. Algorithms, 60 (2012), pp. 683–696.

[9] L. Eldén, *A weighted pseudoinverse, generalized singular values, and constrained least squares problems*, BIT, 22 (1982), pp. 487–501.
[10] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
[11] M. Fuhry and L. Reichel, *A new Tikhonov regularization method*, Numer. Algorithms, 59 (2012), pp. 433–445.
[12] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, 1996.
[13] P. C. Hansen, *Regularization, GSVD and truncated GSVD*, BIT, 29 (1989), pp. 491–504.
[14] P. C. Hansen, *Regularization tools version 4.0 for Matlab 7.3*, Numer. Algorithms, 46 (2007), pp. 189–194.
[15] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.
[16] S. Kindermann, *Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems*, Electron. Trans. Numer. Anal., 38 (2011), pp. 233–257.
[17] S. Morigi, L. Reichel, and F. Sgallari, *A truncated projected SVD method for linear discrete ill-posed problems*, Numer. Algorithms, 43 (2006), pp. 197–213.
[18] S. Morigi, L. Reichel, and F. Sgallari, *Orthogonal projection regularization operators*, Numer. Algorithms, 44 (2007), pp. 99–114.
[19] S. Noschese and L. Reichel, *Inverse problems for regularization matrices*, Numer. Algorithms, 60 (2012), pp. 531–544.
[20] D. L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, J. ACM, 9 (1962), pp. 84–97.
[21] L. Reichel and Q. Ye, *Simple square smoothing regularization operators*, Electron. Trans. Numer. Anal., 33 (2009), pp. 63–83.