Old and new parameter choice rules for discrete ill-posed problems

Lothar Reichel* Giuseppe Rodriguez†

Dedicated to Claude Brezinski and Sebastiano Seatzu on the Occasion of Their 70th Birthdays.

Abstract

Linear discrete ill-posed problems are difficult to solve numerically because their solution is very sensitive to perturbations, which may stem from errors in the data and from round-off errors introduced during the solution process. The computation of a meaningful approximate solution requires that the given problem be replaced by a nearby problem that is less sensitive to disturbances. This replacement is known as regularization. A regularization parameter determines how much the regularized problem differs from the original one. The proper choice of this parameter is important for the quality of the computed solution. This paper studies the performance of known and new approaches to choosing a suitable value of the regularization parameter for the truncated singular value decomposition method and for the LSQR iterative Krylov subspace method in the situation when no accurate estimate of the norm of the error in the data is available. The regularization parameter choice rules considered include several L-curve methods, Regińska's method and a modification thereof, extrapolation methods, the quasi-optimality criterion, rules designed for use with LSQR, as well as hybrid methods.

Keywords Ill-posed problem, regularization, regularization parameter, TSVD, LSQR

1 Introduction

We would like to determine an approximate solution of least-squares problems of the form

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|A\boldsymbol{x}-\boldsymbol{b}\|, \qquad A\in\mathbb{R}^{m\times n}, \qquad \boldsymbol{x}\in\mathbb{R}^n, \qquad \boldsymbol{b}\in\mathbb{R}^m, \tag{1}$$

where the matrix A is of ill-determined rank, i.e., the singular values of A "cluster" at the origin. In particular, A is very ill-conditioned and may be

*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.

†Dipartimento di Matematica e Informatica, Università di Cagliari, viale Merello 92, 09123 Cagliari, Italy. E-mail: `rodriguez@unica.it`. Research supported in part by MIUR-PRIN grant no. 20083KLJEZ-003.

singular. Throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm or the associated induced matrix norm. Least-squares problems of the form (1) with a matrix of ill-determined rank often are referred to as discrete ill-posed problems. They arise, for instance, from the discretization of ill-posed problems, such as Fredholm integral equations of the first kind with a smooth kernel. The vector $\boldsymbol{b}$ in discrete ill-posed problems that arise in science and engineering represents available data and typically is contaminated by a measurement error $\boldsymbol{e} \in \mathbb{R}^m$, which we will refer to as "noise." For ease of notation, we will assume that $m \geq n$; however, only minor modifications are required when $m < n$.

Let $\hat{\boldsymbol{b}}$ be the unknown noise-free vector associated with $\boldsymbol{b}$, i.e., $\boldsymbol{b} = \hat{\boldsymbol{b}} + \boldsymbol{e}$, and consider the unavailable noise-free least-squares problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|A\boldsymbol{x} - \hat{\boldsymbol{b}}\| \tag{2}$$

associated with (1). Its solution of minimal Euclidean norm is given by $\hat{\boldsymbol{x}} = A^\dagger \hat{\boldsymbol{b}}$, where $A^\dagger$ denotes the Moore-Penrose pseudoinverse of A. We would like to determine an accurate approximation of $\hat{\boldsymbol{x}}$ by computing a suitable approximate solution of the available least-squares problem (1) with error-contaminated data $\boldsymbol{b}$. Note that, generally, the solution of (1) of minimal Euclidean norm, given by

$$A^\dagger \boldsymbol{b} = \hat{\boldsymbol{x}} + A^\dagger \boldsymbol{e},$$

is not a useful approximation of $\hat{\boldsymbol{x}}$, because the term $A^\dagger \boldsymbol{e}$ typically dominates $\hat{\boldsymbol{x}}$. Moreover, when the computations are carried out in finite-precision arithmetic, amplified propagated round-off errors generally yield a computed solution of poor quality also when there is no noise in $\boldsymbol{b}$.

A common approach to remedy these difficulties is to replace the least-squares problem (1) by a nearby problem, whose solution is less sensitive to the error $\boldsymbol{e}$ in $\boldsymbol{b}$ and to round-off errors introduced during the solution process. This replacement is known as regularization. We will focus on two well-known regularization methods: truncated singular value decomposition (TSVD) for small to medium-sized problems [15, 25] and truncated iteration with the iterative Krylov subspace method LSQR [20, 36]. For TSVD the regularization parameter is the truncation index, and for LSQR it is the index of the iterate chosen to approximate $\hat{\boldsymbol{x}}$. Thus, for both methods, the regularization parameter is integer-valued.

We conclude this section with a description of TSVD and outline the issues to be discussed in the remainder of this paper. The singular value decomposition (SVD) of the matrix A in (1) is given by

$$A = U \begin{bmatrix} \Sigma \\ O \end{bmatrix} V^T, \tag{3}$$

where the matrices $U = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m] \in \mathbb{R}^{m\times m}$ and $V = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n] \in \mathbb{R}^{n\times n}$ are orthogonal, $O \in \mathbb{R}^{(m-n)\times n}$ is a matrix with all entries zero, and the entries of the diagonal matrix

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

are the singular values of A ordered according to

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_\ell > \sigma_{\ell+1} = \cdots = \sigma_n = 0.$$

The integer ℓ is the rank of A. We remark that when computing the SVD of a matrix of ill-determined rank in finite-precision arithmetic, ℓ typically is n and several of the smallest singular values are "tiny." The columns $\boldsymbol{u}_j$ of U are referred to as left singular vectors of A and the columns $\boldsymbol{v}_j$ of V as right singular vectors. The superscript ^{T} in (3) stands for transposition.

The TSVD is a regularization method which replaces the matrix A in (1) by a closest approximation of rank $k \leq \ell$, and determines the solution of minimal Euclidean norm of the least-squares problem so obtained. Let the matrix

$$\Sigma_k = \mathrm{diag}[\sigma_1, \sigma_2, \ldots, \sigma_k, \underbrace{0, \ldots, 0}_{n-k}] \in \mathbb{R}^{n \times n}$$

be defined by setting the last $n-k$ diagonal entries of Σ to zero for some $1 \leq k \leq \ell$. The Moore-Penrose pseudoinverse of Σ_k is

$$\Sigma_k^\dagger = \mathrm{diag}[\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_k^{-1}, \underbrace{0, \ldots, 0}_{n-k}] \in \mathbb{R}^{n \times n}.$$

A closest rank-k approximation of A with regard to the spectral norm $\|\cdot\|$ is given by

$$A_k = U \begin{bmatrix} \Sigma_k \\ O \end{bmatrix} V^T.$$

It has the Moore-Penrose pseudoinverse

$$A_k^\dagger = V[\Sigma_k^\dagger, O^T]U^T.$$

The minimal-norm solution of the least-squares problem obtained by replacing A by A_k in (1) can be expressed as

$$\boldsymbol{x}_k = A_k^\dagger \boldsymbol{b} = \sum_{j=1}^{k} \frac{\boldsymbol{u}_j^T \boldsymbol{b}}{\sigma_j} \boldsymbol{v}_j. \tag{4}$$

The truncation index k is the regularization parameter. The value of k determines the quality of the approximation $\boldsymbol{x}_k$ of $\hat{\boldsymbol{x}}$ as well as how sensitive $\boldsymbol{x}_k$ is to the error $\boldsymbol{e}$ and to round-off errors introduced during the computations. We would like to choose k so that the difference $\boldsymbol{x}_k - \hat{\boldsymbol{x}}$ is small. Generally, $\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\|$ decreases as k increases and is fairly small, but increases rapidly with k when k is large. Therefore, it is important to choose a suitable truncation index k.

When an accurate estimate of the norm of the error $\boldsymbol{e}$ in $\boldsymbol{b}$ is known, an appropriate value of k often can be determined with the aid of the discrepancy principle, which prescribes that the truncation index k be chosen as the smallest index $k = k_{\mathrm{discr}}$ such that

$$\|A\boldsymbol{x}_{k_{\mathrm{discr}}} - \boldsymbol{b}\| \leq \tau \|\boldsymbol{e}\|. \tag{5}$$

Here $\tau > 1$ is a user-supplied constant independent of $\|\boldsymbol{e}\|$. Properties of the discrepancy principle are discussed, e.g., by Engl et al. [15, Section 4.3], where a proof in Hilbert space setting that, under suitable conditions, $\boldsymbol{x}_{k_{\mathrm{discr}}}$ converges to $\hat{\boldsymbol{x}}$ as $\|\boldsymbol{e}\| \searrow 0$, can be found. Note that k_{discr} increases as $\|\boldsymbol{e}\|$ decreases. The discrepancy principle typically yields a suitable truncation index when an accurate bound for $\|\boldsymbol{e}\|$ is available. We are interested in the situation when no such bound is known and, therefore, the discrepancy principle cannot be applied.

It has been shown by Bakushinski [1] that regularization parameter choice rules that do not use a bound for $\|\boldsymbol{e}\|$ will fail to determine a suitable value of k for some problems. Such rules are therefore sometimes referred to as "heuristic." Nevertheless, there is a need to use heuristic parameter choice rules, because for many discrete ill-posed problems (1) that arise in applications in science and engineering no accurate bound for $\|\boldsymbol{e}\|$ is available. For this reason, many heuristic parameter choice rules have been proposed in the literature, including the L-curve criterion, the residual L-curve criterion, generalized cross validation, error estimation methods based on extrapolation, the quasi-optimality criterion, and Regińska's method; see, e.g., [6, 7, 24, 25, 28, 29, 38, 39, 41, 43] and references therein.

Recently, considerable progress has been made in understanding the performance of the quasi-optimality criterion and of other heuristic methods; see [2, 31]. These results are very valuable; however, they involve constants that describe the performance of the parameter choice rules, and the sizes of these constants can be difficult to assess. It is therefore of interest to investigate the performance of heuristic parameter choice rules when applied to a variety of linear discrete ill-posed problems with different amounts of noise $\boldsymbol{e}$ in the data $\boldsymbol{b}$.

A thorough comparison of many heuristic parameter choice rules is reported in [3]. All the linear discrete ill-posed problems in this comparison have a diagonal matrix. This implies that the singular vectors of the matrices are the axis vectors. We compare several old and new parameter choice rules when applied to discretized Fredholm integral equations of the first kind with a smooth kernel. The singular vectors for matrices obtained in this way typically are discretizations of oscillatory functions, with the frequency of the oscillations increasing with the index of the vector. We include several methods in our comparison that are not considered in [3]. Another recent comparison of heuristic parameter choice rules is reported in [29].

Besides the TSVD regularization method, we consider regularization by truncated iteration with the LSQR iterative Krylov subspace method [20, 36]. This minimal residual iterative method is frequently used to solve large-scale linear discrete ill-posed problems; see, e.g., [20, 22, 25] for illustrations. The number of iterations carried out is the regularization parameter. We compare several well-known and new methods that do not require an accurate bound for $\|\boldsymbol{e}\|$ for determining a suitable number of iterations.

This paper is organized as follows. Section 2 discusses parameter choice rules for small to medium-sized problems for which it is feasible to compute the SVD of A. The solution of large-scale minimization problems (1) by the LSQR iterative method is discussed in Section 3. Both available and new parameter choice rules for determining a suitable number of iterations are described. Section 4 presents computed examples, some of which illustrate how the application of a user-specified regularization matrix affects the performance of the parameter choice rules. Concluding remarks can be found in Section 5.

Claude Brezinski and Sebastiano Seatzu have over many years made a large number of very significant contributions to our understanding of linear discrete ill-posed problems and they have developed powerful numerical methods for their solution. In particular, they have developed several novel techniques for determining a suitable value of the regularization parameter. It is a pleasure to dedicate this paper to them.

2 Parameter choice rules for small to medium-sized problems

The parameter choice rules of this section are designed to determine the truncation index in the TSVD method. We describe both known and new rules.

2.1 The L-curve criterion

The L-curve when applied to determine a suitable truncation index k for TSVD is a plot of the points

$$\{\log \|\boldsymbol{x}_j\|, \log \|\boldsymbol{r}_j\|\}, \quad j = 1, 2, \ldots, \ell, \tag{6}$$

where the $\boldsymbol{x}_j$ are given by (4) and $\boldsymbol{r}_j = \boldsymbol{b} - A\boldsymbol{x}_j$ is the associated residual error.[1] Note that $\|\boldsymbol{x}_j\|$ and $\|\boldsymbol{r}_j\|$, $j = 1, 2, \ldots, \ell$, can be evaluated easily and quickly when the SVD of A (3) is available. It is straightforward to show that $\|\boldsymbol{x}_j\| \leq \|\boldsymbol{x}_{j+1}\|$ and $\|\boldsymbol{r}_j\| \geq \|\boldsymbol{r}_{j+1}\|$ for all j.

Ideally, the plot of the points (6) (or part thereof) looks like a discretization of the letter "L". Therefore, the plot is referred to as the L-curve or discrete L-curve. We seek to determine a truncation index $k = k_{\text{L-curve}}$ that corresponds to a point close to the "vertex" of the "L." This choice is designed to determine an approximate solution $\boldsymbol{x}_{k_{\text{L-curve}}}$ of (1) of moderate norm with an associated residual error $\boldsymbol{r}_{k_{\text{L-curve}}}$ of fairly small norm. The latter property indicates that $\boldsymbol{x}_{k_{\text{L-curve}}}$ is a quite accurate solution of (1), and the former property suggests that $\boldsymbol{x}_{k_{\text{L-curve}}}$ is not severely contaminated by propagated noise; see, e.g., [24, 25, 28] for discussions on the L-curve criterion.

It is often quite easy to visually determine the index $k_{\text{L-curve}}$ from the plot, but it is a fairly difficult task to write a computer program that finds this index. A reason for this is that when consecutive points of the plot are connected by line segments, the resulting graph is not guaranteed to be convex and may have several kinks that should not be considered the "vertex" of the "L." Castellanos et al. [12] proposed a scheme for determining a truncation index close to the "vertex" of the "L" by forming a sequence of triangles with vertices at the points (6) and then determining the desired "vertex" of the "L" from the shape of these triangles. We will refer to this scheme as the *L-triangle method* in the numerical examples reported in Section 4.

Hansen et al. [27] proposed an alternative approach to determining the "vertex" of the "L" from the point set (6). They construct a sequence of *pruned* L-curves, removing an increasing number of points, and consider a list of candidate "vertices" produced by two different selection algorithms. The "vertex" of the "L" is selected from this list by taking the last point before reaching the part of the L-curve, where the norm of the computed approximate solution starts to increase rapidly and the norm of the associated residual vectors stagnates. We will refer to this scheme as the *L-corner method*.

[1] In some references the plot of $\{\log \|\boldsymbol{r}_j\|, \log \|\boldsymbol{x}_j\|\}$, $j = 1, 2, \ldots, \ell$, is referred to as the L-curve.

It is known that the L-curve method may fail to determine a suitable truncation index; see [21, 44] for analyses. However, in applications to problems in science and engineering, the L-curve method often performs well when there is a fair amount of noise in the data. The L-curve method therefore continues to be a popular approach to determine a suitable truncation index for the TSVD.

2.2 The residual L-curve criterion

Sometimes several of the points (6) "cluster" near the "vertex" of the "L." This may make it difficult to distinguish the "vertex." The residual L-curve seeks to circumvent this difficulty by instead considering the set

$$\{\log j, \log \|\boldsymbol{r}_j\|\}, \quad j = 1, 2, \ldots, \ell. \tag{7}$$

The "vertex" of this point set is obtained by applying the L-corner method [27] to the set (7). The index k of the vertex is the truncation index for the TSVD method. We found the residual L-curve based on the point set (7) to give better approximations $\boldsymbol{x}_k$ of $\hat{\boldsymbol{x}}$ than the residual L-curve determined by the points $\{j, \log \|\boldsymbol{r}_j\|\}_{j=1}^{\ell}$, which were used in [39].

2.3 The condition L-curve criterion

The condition L-curve plots $\|\boldsymbol{r}_j\|$ versus the condition number $\kappa(A_j)$ of A_j in log-log scale. The latter is defined as σ_1/σ_j and therefore is easy to compute when the SVD of A is available. Since the singular values are a decreasing function of their index, we have $\kappa(A_j) \leq \kappa(A_{j+1})$ for all j. Thus, the condition L-curve is the plot of the point set

$$\{\log \kappa(A_j), \log \|\boldsymbol{r}_j\|\}, \quad j = 1, 2, \ldots, \ell. \tag{8}$$

We determine the "vertex" of the graph of this point set with the aid of the L-corner method [27]. The index of the vertex is the truncation index for the TSVD method. The condition L-curve was introduced in [10] for use with the GMRES iterative method, because the norm of iterates generated by GMRES is not guaranteed to increase with the iteration number.

2.4 Regińska's method

The L-curve criterion of Subsection 2.1 also can be applied to determine the regularization parameter in Tikhonov regularization; see, e.g., [24, 25]. The discrete set (6) then is replaced by a continuous set. Regińska [41] proposed a parameter choice rule for Tikhonov regularization and showed how it relates to the L-curve criterion for Tikhonov regularization. This rule easily can be adapted to the situation when the regularization parameter is discrete. Regińska's parameter choice rule applied to the set $\{\|\boldsymbol{x}_j\|, \|\boldsymbol{r}_j\|\}_{j=1}^{\ell}$ prescribes that the truncation index k be the minimizer of the function

$$\phi_\nu(j) = \|\boldsymbol{r}_j\| \, \|\boldsymbol{x}_j\|^\nu, \quad j = 1, 2, \ldots, \ell - 1, \tag{9}$$

where $\nu > 0$ is a user-specified parameter. We will use $\nu = 2$ in the computed examples of Section 4.

2.5 The restricted Regińska method

The performance of the Regińska method can be improved by two preprocessing steps that select a subset of consecutive elements of the set $\{\|\boldsymbol{x}_j\|, \|\boldsymbol{r}_j\|\}_{j=1}^{\ell-1}$. Regińska's method is then applied to this subset. We refer to this scheme as the restricted Regińska method. It consists of the following steps:

- *Eliminate points* $\{\|\boldsymbol{x}_j\|, \|\boldsymbol{r}_j\|\}$ *with* $\|\boldsymbol{x}_j\|$ *very large:* When working in double precision arithmetic, let $M = 10^{10}$ and determine the smallest index j_M such that
$$\|\boldsymbol{x}_{j_M}\| > M \min_{1 \le j < \ell} \|\boldsymbol{x}_j\|, \tag{10}$$
where we recall that $\|\boldsymbol{x}_j\| \le \|\boldsymbol{x}_{j+1}\|$ for all j. If the inequality (10) is violated for all indices j_M, then set $j_M = \ell - 1$. Discard the points $\{\boldsymbol{x}_j, \boldsymbol{r}_j\}_{j=j_M+1}^{\ell-1}$. The minimization in (10) is carried out over nonvanishing $\|\boldsymbol{x}_j\|$ only.

- *Select the largest convex subset:* We say that a subset of consecutive points $\{\|\boldsymbol{x}_j\|, \phi_\nu(j)\}_{j=j_i}^{j_{i+1}}$ with $1 \le j_i < j_{i+1} \le j_M$ is convex if the piecewise linear function obtained by connecting consecutive points in the set by line segments is convex. Subdivide the point set $\{\|\boldsymbol{x}_j\|, \phi_\nu(j)\}_{j=1}^{j_M}$ into the smallest number of convex subsets of consecutive points $\{\|\boldsymbol{x}_j\|, \phi_\nu(j)\}_{j=j_i}^{j_{i+1}}$, $i = 1, 2, \ldots, t-1$, where $1 \le j_1 < j_2 < \cdots < j_t \le j_M$. Let ι be the smallest index of the convex point set with the most elements, i.e.,
$$j_{\iota+1} - j_\iota = \max_{1 \le i < t} (j_{i+1} - j_i). \tag{11}$$
If there are several convex point sets with the same largest number of elements, then we choose ι to be the smallest possible index such that (11) holds.

- *Apply Regińska's method:* Let the truncation index k for TSVD be the minimizer of
$$\phi_\nu(j) = \|\boldsymbol{r}_j\| \, \|\boldsymbol{x}_j\|^\nu, \quad j = j_\iota, j_\iota + 1, \ldots, j_{\iota+1}. \tag{12}$$
Similarly as in (9), we use $\nu = 2$ in the computed examples of Section 4.

2.6 The quasi-optimality criterion

This criterion lets the truncation index k be the smallest index j_q, such that

$$\|\boldsymbol{x}_{j_q+1} - \boldsymbol{x}_{j_q}\| = \min_{1 \le j < \ell} \|\boldsymbol{x}_{j+1} - \boldsymbol{x}_j\|.$$

Recent theoretical and numerical investigations are reported in [2, 3, 31]. Subsampling can improve the performance of the quasi-optimal criterion, but choosing an appropriate subsample can be difficult; see [4] for an analysis and discussion. The present paper will not discuss this modification.

2.7 Generalized cross validation (GCV)

This method is derived within a statistical framework. Roughly, one would like the regularization parameter to be such that when an entry of the data vector $\boldsymbol{b}$ were missing, the regularized solution should be able to predict this entry accurately. The error $\boldsymbol{e}$ in $\boldsymbol{b}$ is assumed to be random with zero mean and covariance matrix $\sigma^2 I$ with σ^2 unknown. When regularization is carried out by TSVD, the GCV method prescribes that the truncation index k be the integer j that minimizes the function

$$G(j) = \frac{\|\boldsymbol{r}_j\|^2}{(m-j)^2};$$

see, e.g., [5, 25] for further details on this method. For many problems, the function G is quite flat around the minimum. The GCV method is known to be unreliable for small to medium-sized problems and robust versions have been developed; see [3, 32]. The latter methods require the choice of an additional parameter and, therefore, will not be considered in the present paper.

The application of the GCV method to large linear discrete ill-posed problems for which it is unattractive to compute the SVD is fairly complicated; see [19].

2.8 Error estimation methods based on extrapolation

Let as usual $\boldsymbol{x}_j$ be defined by (4) and $\boldsymbol{r}_j = \boldsymbol{b} - A\boldsymbol{x}_j$. Brezinski et al. [7] show that

$$\eta_j^{(\nu)} = d_0^{\nu-1} d_1^{5-2\nu} d_2^{\nu-3}, \tag{13}$$

where $d_0 = \|\boldsymbol{r}_j\|^2$, $d_1 = \|A^T\boldsymbol{r}_j\|^2$, and $d_2 = \|AA^T\boldsymbol{r}_j\|^2$, is an estimate for the squared norm of the error in $\boldsymbol{x}_j$ for each $\nu \in \mathbb{R}$ under the assumption that $\boldsymbol{r}_j$ is the error in $\boldsymbol{b}$. In fact, using the SVD of A (3) and letting $\alpha_i = \boldsymbol{u}_i^T\boldsymbol{b}$, we can write

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}_j\|^2 = \sum_{i=j+1}^{\ell} \frac{\alpha_i^2}{\sigma_i^2} \simeq \alpha^2\sigma^{-2}, \tag{14}$$

where the final approximation is obtained by keeping just one term in the summation. We obtain similarly

$$d_0 \simeq \alpha^2, \quad d_1 \simeq \sigma^2\alpha^2, \quad d_2 \simeq \sigma^4\alpha^2.$$

Substituting these expressions into (13) and comparing with (14) yields

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}_j\|^2 \simeq \eta_j^{(\nu)}, \quad \forall \nu \in \mathbb{R}.$$

When the matrix A is square, one can show similarly that

$$e_j^{(\nu)} = d_0^{\nu-1} c_1^{3-\nu} d_1^{\nu-4} \tag{15}$$

also is an estimate of $\|\hat{\boldsymbol{x}} - \boldsymbol{x}_j\|^2$ for all $\nu \in \mathbb{R}$, where $c_1 = (\boldsymbol{r}_j^T A \boldsymbol{r}_j)^2$; see [6] for details. The quantities in the error estimates (13) and (15) can be computed inexpensively for several values of j when the SVD of A is available. Some particular error estimates belonging to the above families are

$$\eta_j^{(2)} = \frac{\|\boldsymbol{r}_j\|^2 \, \|A^T \boldsymbol{r}_j\|^2}{\|AA^T \boldsymbol{r}_j\|^2}, \qquad \eta_j^{(3)} = e_j^{(3)} = \frac{\|\boldsymbol{r}_j\|^4}{\|A^T \boldsymbol{r}_j\|^2}. \tag{16}$$

In the computed examples reported in Section 4, we let the truncation index k be the index $k = j$ that minimizes $\eta_j^{(3)}$. We will refer to this parameter choice rule as the *extrapolation method*. Numerous computed examples reported in [6, 7] show this choice of truncation index often to be appropriate, even though $\eta_k^{(2)}$, $\eta_k^{(3)}$, and $e_k^{(3)}$ typically are poor approximations of $\|\hat{\boldsymbol{x}} - \boldsymbol{x}_k\|^2$ when the matrix A is ill-conditioned.

2.9 Hybrid quasi-optimal methods

The function

$$j \to \|\boldsymbol{x}_{j+1} - \boldsymbol{x}_j\| \tag{17}$$

typically has many local minima; see Figure 1 of Section 4. When several of the local minima are close to being global minima, the index j_q chosen by the quasi-optimality criterion might not be appropriate. We therefore investigate the performance of hybrid methods that determine an initial index j_{init} by some method different from the quasi-optimality criterion, say the L-curve criterion or the restricted Regińska method, and then apply the quasi-optimality criterion in a neighborhood of j_{init}. Specifically, we let the truncation index k be a local minimum, j_{loc}, of the function (17) closest to j_{init}, and as small as possible in case of nonunicity.

An interactive hybrid method that combines the L-curve criterion with quasi-optimality to determine which iterate generated by a Krylov subspace method to use as approximation of $\hat{\boldsymbol{x}}$ has been described in [33]. We comment on this method further below.

3 Parameter choice rules for large problems

Popular solution methods for large linear discrete ill-posed problems include iterative Krylov subspace methods, such as LSQR [36] or RRGMRES [34, 35]. Regularization is achieved by terminating the iterations sufficiently early. The difficulty is to determine which iterate to choose as an approximation of $\hat{\boldsymbol{x}}$. This section focuses on the LSQR iterative method; however, the discussion, suitably modified, also applies to other Krylov subspace methods.

The LSQR method is an implementation of the conjugate gradient method applied to the normal equations associated with (1). Let the initial iterate be $\boldsymbol{x}_0 = \boldsymbol{0}$. Then the iterates $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots$, generated by LSQR are characterized by

$$\|A\boldsymbol{x}_j - \boldsymbol{b}\| = \min_{\boldsymbol{x} \in \mathcal{K}_j(A^T A, A^T \boldsymbol{b})} \|A\boldsymbol{x} - \boldsymbol{b}\|, \quad \boldsymbol{x}_j \in \mathcal{K}_j(A^T A, A^T \boldsymbol{b}), \tag{18}$$

where

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

is a Krylov subspace. The LSQR method is implemented with the aid of Lanczos bidiagonalization of A:

$$AV_j = U_{j+1}\bar{C}_j, \qquad A^T U_j = V_j C_j^T, \tag{19}$$

where the columns of $U_{j+1} \in \mathbb{R}^{m\times(j+1)}$ form an orthonormal basis for the Krylov subspace $\mathcal{K}_{j+1}(AA^T, \boldsymbol{b})$ with $U_{j+1}\boldsymbol{e}_1 = \boldsymbol{b}/\|\boldsymbol{b}\|$, and $U_j \in \mathbb{R}^{m\times j}$ consists of the first j columns of U_{j+1}. The columns of $V_j \in \mathbb{R}^{n\times j}$ form an orthonormal basis for $\mathcal{K}_j(A^TA, A^T\boldsymbol{b})$. The rectangular matrix $\bar{C}_j \in \mathbb{R}^{(j+1)\times j}$ is lower bidiagonal with leading submatrix $C_j \in \mathbb{R}^{j\times j}$. Iteration $j+1$ appends one column to U_{j+1} and V_j to obtain U_{j+2} and V_{j+1}, respectively, as well as a row and a column to $\bar{C}_j$ to obtain $\bar{C}_{j+1}$. We assume that j is small enough so that the matrices U_{j+1}, V_j, and $\bar{C}_j$ with the stated properties exist.

Substituting (19) into (18) yields the equivalent reduced minimization problem

$$\left\|\bar{C}_j\boldsymbol{y}_j - \boldsymbol{e}_1\|\boldsymbol{b}\|\right\| = \min_{\boldsymbol{y}\in\mathbb{R}^j}\left\|\bar{C}_j\boldsymbol{y} - \boldsymbol{e}_1\|\boldsymbol{b}\|\right\|, \quad \boldsymbol{y}_j \in \mathbb{R}^j, \tag{20}$$

where

$$\boldsymbol{x}_j = V_j\boldsymbol{y}_j, \quad \boldsymbol{r}_j = A\boldsymbol{x}_j - \boldsymbol{b} = U_{j+1}\left(\bar{C}_j\boldsymbol{y}_j - \boldsymbol{e}_1\|\boldsymbol{b}\|\right). \tag{21}$$

We are in a position to discuss several parameter choice rules and first describe how the rules of Section 2 can be applied.

3.1 Parameter choice rules based on the reduced problem (20)

It follows from (21) that

$$\|\boldsymbol{x}_j\| = \|\boldsymbol{y}_j\|, \quad \|\boldsymbol{x}_{j+1} - \boldsymbol{x}_j\| = \left\|\boldsymbol{y}_{j+1} - \begin{bmatrix} \boldsymbol{y}_j \\ 0 \end{bmatrix}\right\|, \quad \|\boldsymbol{r}_j\| = \|\bar{C}_j\boldsymbol{y}_j - \boldsymbol{e}_1\|\boldsymbol{b}\|\|.$$

These relations show that the quantities required by the L-curve and residual L-curve criteria, Regińska's and the restricted Regińska methods, as well as by the quasi-optimality criterion discussed in Section 2 can be determined inexpensively from the reduced problem (20). In order for several of these parameter choice rules to perform well, one has to carry out a few more iteration steps than the index of the iterate that best approximates $\hat{\boldsymbol{x}}$. For instance, the L-corner algorithm requires that enough iterates be carried out so that the discrete L-curve includes a portion where the norm of the residual vector does not change much during the last iterations.

The condition L-curve criterion can be applied by computing the condition number of a sequence of small bidiagonal matrices $\bar{C}_1, \bar{C}_2, \bar{C}_3, \ldots$ and by using the associated residual vector norms $\|\bar{C}_1\boldsymbol{y}_1 - \boldsymbol{e}_1\|\boldsymbol{b}\|\|, \|\bar{C}_2\boldsymbol{y}_2 - \boldsymbol{e}_1\|\boldsymbol{b}\|\|, \|\bar{C}_3\boldsymbol{y}_3 - \boldsymbol{e}_1\|\boldsymbol{b}\|\|, \ldots$. Similarly, all quantities required when applying the extrapolation method of Subsection 2.8 can be evaluated from the reduced problem (20). For instance, we have

$$\left\|A^T\boldsymbol{r}_j\right\| = \left\|C_{j+1}^T\left(\bar{C}_j\boldsymbol{y}_j - \boldsymbol{e}_j\|\boldsymbol{b}\|\right)\right\|, \quad \left\|AA^T\boldsymbol{r}_j\right\| = \left\|\bar{C}_{j+1}C_{j+1}^T\left(\bar{C}_j\boldsymbol{y}_j - \boldsymbol{e}_j\|\boldsymbol{b}\|\right)\right\|;$$

see [7] for details on the organization of the computations. We illustrate the performance of some of these methods in Section 4.

The parameter choice rules of this subsection can, suitably modified, also be applied to other iterative Krylov subspace methods, including RRGMRES. The parameter choice rules described in the following subsections are specific for LSQR.

3.2 The quadrature method

Hnětynková et al. [30] recently proposed to use the connection between Lanczos bidiagonalization (referred to as Golub-Kahan bidiagonalization) and Gauss quadrature to estimate the norm of the error $\boldsymbol{e}$ in $\boldsymbol{b}$. We refer to this approach as the quadrature method. The connection between Lanczos bi- or tridiagonalization, and Gauss quadrature is developed and applied in the recent book by Golub and Meurant [17]; see also [16]. Applications to the solution of linear discrete ill-posed problems also are described in, e.g., [8, 9, 11, 18, 38].

Introduce the spectral factorization

$$AA^T = W\Lambda W^T, \quad \Lambda = \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_m] \in \mathbb{R}^{m\times m}$$

with $W \in \mathbb{R}^{m\times m}$ orthogonal. Then the sum

$$\boldsymbol{b}^T f(AA^T)\boldsymbol{b} = \sum_{j=1}^{m} f(\lambda_j)\omega_j^2, \quad \omega_j = \boldsymbol{b}^T W \boldsymbol{e}_j,$$

can be viewed as a Stieltjes integral with a measure induced by the matrix A and vector $\boldsymbol{b}$. It can be shown that

$$G_j f := \|\boldsymbol{b}\|^2 \boldsymbol{e}_1^T f(C_j C_j^T)\boldsymbol{e}_1 \tag{22}$$

is a j-point Gauss quadrature rule associated with this measure; see, e.g., [8, 17] for details. Hnětynková et al. [30] use this connection between Lanczos bidiagonalization and Gauss quadrature to derive a method for detecting the amount of "noise" in the vector $\boldsymbol{b}$. Introduce the singular value decomposition

$$C_j = W_j S_j \tilde{W}_j^T, \tag{23}$$

where the matrices $W_j, \tilde{W}_j \in \mathbb{R}^{j\times j}$ are orthogonal and the diagonal entries of $S_j = \mathrm{diag}[s_1^{(j)}, s_2^{(j)}, \ldots, s_j^{(j)}] \in \mathbb{R}^{j\times j}$ are the singular values of C_j ordered according to $s_1^{(j)} \geq s_2^{(j)} \geq \ldots \geq s_j^{(j)} \geq 0$.

Substituting (23) into (22) yields the expression

$$G_j f = \sum_{i=1}^{j} f((s_i^{(j)})^2)(\omega_i^{(j)})^2,$$

where

$$\omega_i^{(j)} = \|\boldsymbol{b}\|\boldsymbol{e}_1^T W_j \boldsymbol{e}_i. \tag{24}$$

Thus, the squared singular values are Gaussian nodes and the squared normalized first component of the left singular vectors are Gaussian weights.

Hnětynková et al. [30] show that the Gaussian weight $(\omega_j^{(j)})^2$, which is associated with the smallest singular value, is a decreasing function of j, and that the decrease stagnates when j is large. Then $(\omega_j^{(j)})^2 \approx \|\boldsymbol{e}\|^2$. Let $k \geq 1$ be the smallest integer such that

$$\frac{|\boldsymbol{e}_1^T \boldsymbol{w}^{(k+1)}|}{|\boldsymbol{e}_1^T \boldsymbol{w}^{(k+1+\text{step})}|} < \left(\frac{|\boldsymbol{e}_1^T \boldsymbol{w}^{(k)}|}{|\boldsymbol{e}_1^T \boldsymbol{w}^{(k+1)}|} \right)^{\zeta}$$

where $\boldsymbol{w}^{(k)} = W_k \boldsymbol{e}_k$, $\zeta = 0.5$, and step $= 3$. Hnětynková et al. [30] say that stagnation occurs at step k, and use the iterate $\boldsymbol{x}_k$ as an approximation of $\hat{\boldsymbol{x}}$.

It is important that the columns of the matrices $U_1, U_2, U_3, \ldots$ in (19) be numerically orthonormal for the good performance of this regularization parameter choice rule. To secure numerical orthogonality, Hnětynková et al. [30] carry out double reorthogonalization; the software they provide for bidiagonalization is used in the numerical illustrations presented in Section 4.

3.3 The ratio method

This parameter choice rule is a modification of the quadrature method. Let $\omega_j^{(j)}$, $j = 1, 2, \ldots, i$, be defined by (24), where the number of iterations i is suitably large. We consider the ratios

$$\rho_j = \frac{\omega_j^{(j)}}{s_j^{(j)}}, \qquad j = 1, 2, \ldots, i,$$

where $s_j^{(j)}$ is the smallest singular value of the bidiagonal matrix C_j in (19). We choose the iterate $\boldsymbol{x}_k$ as an approximation of $\hat{\boldsymbol{x}}$, where k is the smallest index such that

$$\rho_k = \min_{1 \leq j \leq i} \rho_j.$$

Here we assume that i is large enough so that $k < i$. This approach yields a suitable iterate when the function $j \to \rho_j$ is not flat around $j = k$; otherwise we define a plateau containing the minimal value of ρ_j, and let k be the index of the last point of the plateau. A plateau is a set of consecutive ratios of near-minimal value. Specifically, consecutive ratios that contain $\rho_{\min} = \min_{1 \leq j \leq i} \rho_j$ and satisfy

$$\log \rho_j \leq \log \rho_{\min} + \frac{1}{10} \left(\log \max_{1 \leq j \leq i} \rho_j - \log \rho_{\min} \right)$$

form a plateau.

4 Computed examples

We present several computed examples that illustrate the performance of the parameter choice rules described in Sections 2 and 3. These rules are applied to a set of linear discrete ill-posed problems from Regularization Tools [26] (Baart, Deriv2 (with parameter 2), Foxgood, Gravity, Heat (with parameter 1), Ilaplace (with parameter 3), Phillips [37], and Shaw [42]) as well as to the problems Hilbert and Lotkin from the MATLAB `gallery`. The desired solutions $\hat{\boldsymbol{x}}$ are provided by the test problems from [26]; for the problems Hilbert and Lotkin, we use the solution $\hat{\boldsymbol{x}}$ of the Shaw test problem. The error-free data vector is

Table 1: Parameter choice rules for TSVD.

Method	$n \times n$	$2n \times n$
L-corner	6%(0%)	4%(0%)
Res. L-curve	3%(1%)	1%(0%)
Cond. L-curve	2%(1%)	8%(6%)
Regińska	4%(0%)	2%(0%)
Restr. Regińska	3%(0%)	2%(0%)
Quasi-optimality	9%(1%)	7%(1%)
GCV	24%(20%)	8%(4%)
Extrapolation	5%(1%)	9%(0%)
Discrepancy	0%(0%)	0%(0%)

for all problems generated according to $\hat{\boldsymbol{b}} := A\hat{\boldsymbol{x}}$. All computations are carried out in MATLAB with $\epsilon_{\text{machine}} = 2.22 \cdot 10^{-16}$.

The perturbed data vectors $\boldsymbol{b}$ are determined by adding a "noise-vector" $\boldsymbol{e}$ to $\hat{\boldsymbol{b}}$. Specifically, we let the vector $\boldsymbol{w}$ have normally distributed entries with mean zero and variance one, and compute

$$\boldsymbol{b} = \hat{\boldsymbol{b}} + \boldsymbol{w}\,\|\hat{\boldsymbol{b}}\|\frac{\delta}{\sqrt{n}}.$$

By the Central Limit Theorem, the relative error in $\boldsymbol{b}$ satisfies

$$\frac{\|\boldsymbol{b} - \hat{\boldsymbol{b}}\|}{\|\hat{\boldsymbol{b}}\|} \approx \delta.$$

We use the noise-levels $\delta = 1\cdot 10^{-1}$, $\delta = 1\cdot 10^{-2}$, and $\delta = 1\cdot 10^{-3}$ in the computed examples. These noise-levels are compatible with real world applications. The value of δ is used in the discrepancy principle (5), where we substitute $\delta\|\hat{\boldsymbol{b}}\|$ for $\|\boldsymbol{e}\|$; the other parameter choice rules do not explicitly use the value of δ.

Table 1 shows the performance of the parameter choice rules listed in the first column when applied to determining the truncation index in the TSVD method for all the test problems. We consider matrices A of sizes $n \times n$ and $2n \times n$ for $n = 40$ and $n = 100$. The rectangular matrices are obtained by first generating $2n \times 2n$ matrices with the available software and then retaining the first n columns. We solve each of the 10 square and 10 rectangular problems 10 times with different "noise-vectors." Since three noise-levels are used, this amounts to 1200 experiments.

Let k_{best} denote the truncation index that yields the smallest error, i.e.,

$$\|\boldsymbol{x}_{k_{\text{best}}} - \hat{\boldsymbol{x}}\| = \min_j \|\boldsymbol{x}_j - \hat{\boldsymbol{x}}\|.$$

The first (green) entries in the second and third columns of Table 1 show the percentage of experiments that produced an approximate solution $\boldsymbol{x}_k$ with

$$\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\| > 10\,\|\boldsymbol{x}_{k_{\text{best}}} - \hat{\boldsymbol{x}}\|; \tag{25}$$

the second entries of the second and third columns (red and in parentheses) display the percentage of experiments that produced an approximate solution $\boldsymbol{x}_k$ with

$$\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\| > 100\,\|\boldsymbol{x}_{k_{\text{best}}} - \hat{\boldsymbol{x}}\|. \tag{26}$$

Table 2: Parameter choice rules for LSQR.

Method	$n \times n$	$2n \times n$
L-corner	4%(0%)	2%(0%)
L-triangle	4%(3%)	4%(4%)
Res. L-curve	13%(12%)	10%(10%)
Cond. L-curve	4%(4%)	13%(9%)
Regińska	3%(0%)	1%(0%)
Restr. Regińska	1%(0%)	1%(0%)
Quasi-optimality	2%(0%)	1%(0%)
Extrapolation	8%(0%)	12%(1%)
Quadrature	3%(1%)	2%(0%)
Ratio	0%(0%)	0%(0%)
Discrepancy	0%(0%)	0%(0%)

We used the parameter $\tau = 1.3$ for the discrepancy principle (5). Table 1 shows most of the heuristic methods, except for GCV and the quasi-optimality criterion, to perform quite well, with Regińska's and the restricted Regińska methods, as well as the residual and condition L-curve criteria to perform the best for square matrices. The former three methods perform even better when applied to discrete ill-posed problems with rectangular matrices. The table illustrates that it is beneficial to use the discrepancy principle when possible and, when this is not the case, fairly good results can be expected by using certain heuristic parameter choice rules. Table 1 displays the reliability of the parameter choice rules compared. Since the factors 10 and 100 in (25) and (26), respectively, are quite large, a large entry in Table 1 indicates that the corresponding method should not be used. For instance, the GCV method is seen to give an error that is 10 times larger than the smallest possible in about 1/4 of the experiments. Therefore GCV is an unreliable parameter choice rule and should not be applied to the problems at hand.

Table 2 shows the performance of the parameter choice rules of Section 3 when applied to determining which iterate $\boldsymbol{x}_k$ generated by LSQR to use as an approximation of $\hat{\boldsymbol{x}}$. The table is analogous to Table 1. The matrices used were of sizes $n \times n$ or $2n \times n$ with $n = 100$ and $n = 400$. The table shows the restricted Regińska and the ratio methods to be the best heuristic parameter choice rules. In fact, the latter performed as well as the discrepancy principle. LSQR is applied with reorthogonalization. The effect of loss of orthogonality of the columns of the matrices U_j and V_j in (19) on the quasi-optimality criterion is illustrated in Figure 3.

The performance of heuristic parameter choice rules may depend on the noise-level. Many of the rules perform better for large noise-levels than for small ones. This is illustrated by Table 3 for TSVD. The L-corner, residual L-curve, Regińska, restricted Regińska, and extrapolation methods all perform quite well for noise-levels $1 \cdot 10^{-2}$ and larger. It is remarkable that the condition L-curve gives good results for the noise-level $1 \cdot 10^{-3}$ as well.

The quasi-optimality criterion does not perform well in Table 1. We therefore applied this criterion in conjunction with the other parameter choice rules as described in Subsection 2.9. Table 4 displays the performance of these hybrid methods when applied to determine the truncation index for TSVD. Application

Table 3: Parameter choice rules for TSVD for $n \times n$ matrices and several noise-levels δ.

Method	$\delta = 1 \cdot 10^{-3}$	$\delta = 1 \cdot 10^{-2}$	$\delta = 1 \cdot 10^{-1}$
L-corner	5%(0%)	1%(0%)	0%(0%)
Res. L-curve	2%(1%)	1%(0%)	1%(0%)
Cond. L-curve	0%(0%)	0%(0%)	2%(1%)
Regińska	3%(0%)	1%(0%)	0%(0%)
Restr. Regińska	3%(0%)	0%(0%)	0%(0%)
Quasi-optimality	4%(0%)	4%(0%)	1%(0%)
GCV	8%(7%)	8%(7%)	9%(7%)
Extrapolation	3%(0%)	1%(0%)	0%(0%)
Discrepancy	0%(0%)	0%(0%)	0%(0%)

Table 4: Hybrid parameter choice rules using the quasi-optimality criterion applied to TSVD.

Method	$n \times n$	$2n \times n$
Q-L-corner	6%(0%)	4%(0%)
Q-Res. L-curve	2%(1%)	1%(0%)
Q-Cond. L-curve	1%(1%)	5%(2%)
Q-Regińska	4%(0%)	2%(0%)
Q-Restr. Regińska	3%(0%)	2%(0%)
Q-GCV	20%(17%)	4%(1%)
Q-Extrapolation	1%(1%)	0%(0%)

Table 5: Hybrid parameter choice rules using the quasi-optimality criterion applied to LSQR.

Method	$n \times n$	$2n \times n$
Q-L-corner	4%(0%)	1%(0%)
Q-L-triangle	4%(3%)	4%(4%)
Q-Res. L-curve	12%(12%)	10%(10%)
Q-Cond. L-curve	4%(4%)	11%(7%)
Q-Regińska	3%(0%)	1%(0%)
Q-Restr. Regińska	2%(0%)	1%(0%)
Q-Extrapolation	3%(0%)	1%(1%)
Q-Quadrature	2%(1%)	0%(0%)
Q-Ratio	0%(0%)	0%(0%)

Table 6: Parameter choice rules for TGSVD with regularization matrix (28).

Method	$n \times n$	$2n \times n$
Regińska	5%(0%)	6%(0%)
Restr. Regińska	2%(0%)	2%(0%)
Quasi-optimality	12%(0%)	6%(0%)
Extrapolation	8%(2%)	3%(0%)
Discrepancy	3%(0%)	2%(0%)

to choosing a suitable iterate computed with LSQR is illustrated in Table 5. The hybrid methods perform much better than the quasi-optimality criterion in Table 1, and several of the parameter choice rules of Table 1 are enhanced by combining them with the quasi-optimality criterion. Some parameter choice methods in Table 2 are improved by combining them with the quasi-optimality criterion. LSQR is used with reorthogonalization.

For some discrete ill-posed problems a more accurate approximation of the desired solution $\hat{\boldsymbol{x}}$ can be determined by using a suitable regularization matrix L. Common choices of regularization matrices are scaled finite difference approximations of derivative operators such as

$$L = \frac{1}{2}\begin{bmatrix} 1 & -1 & & & & 0 \\ & 1 & -1 & & & \\ & & 1 & -1 & & \\ & & & \ddots & \ddots & \\ 0 & & & & 1 & -1 \end{bmatrix} \in \mathbb{R}^{(n-1)\times n} \tag{27}$$

and

$$L = \frac{1}{4}\begin{bmatrix} -1 & 2 & -1 & & & 0 \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ 0 & & & -1 & 2 & -1 \end{bmatrix} \in \mathbb{R}^{(n-2)\times n}; \tag{28}$$

see, e.g., [13, 25, 40] for discussions on properties of these and other regularization matrices.

For small to medium-sized problems, the regularization matrices are applied by first reducing the matrix pair $\{A, L\}$ by the generalized SVD (GSVD) to a pair of diagonal matrices. Regularization then can be carried out by only including terms with sufficiently large generalized singular values in the representation of the computed approximation of $\hat{\boldsymbol{x}}$; see [25] for details. The regularization method so obtained is referred to as the truncated GSVD (TGSVD) method. A parameter choice rule is applied to determine a suitable truncation index, just as for TSVD. The TGSVD regularization method is equivalent to applying the TSVD of the matrix $AL_A^\dagger$ to regularize the least-squares problem (29) defined below; see [23] for details. Table 6 describes results obtained with this approach for the best performing parameter choice rules. The restricted Regińska method yields particularly good results; they are better than those achieved with the discrepancy principle.

It is not attractive to compute the GSVD of a matrix pair $\{A, L\}$ with large matrices due to the high computational cost. Eldén [14] introduced the

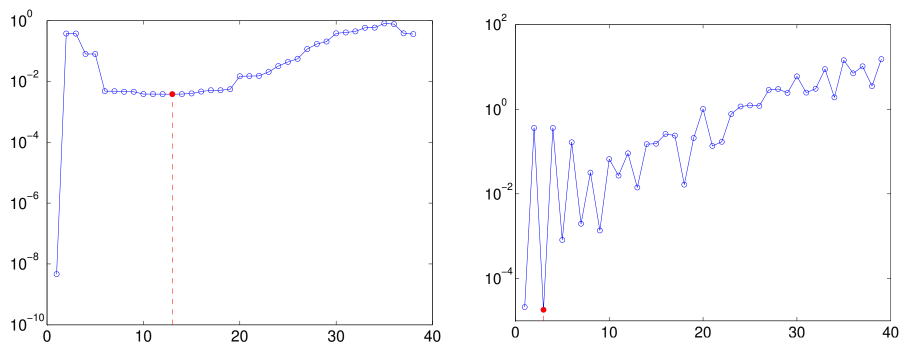
Table 7: Parameter choice rules for LSQR with regularization matrix (28).

Method	$n \times n$	$2n \times n$
Regińska	8%(0%)	7%(0%)
Restr. Regińska	1%(0%)	2%(0%)
Quasi-optimality	6%(0%)	4%(0%)
Ratio	11%(0%)	11%(1%)
Discrepancy	3%(0%)	3%(0%)

A-weighted generalized inverse of L, denoted by $L_A^\dagger$, with the property that the minimization problem

$$\min_{\boldsymbol{y}\in\mathbb{R}^n} \|AL_A^\dagger \boldsymbol{y} - \bar{\boldsymbol{b}}\| \tag{29}$$

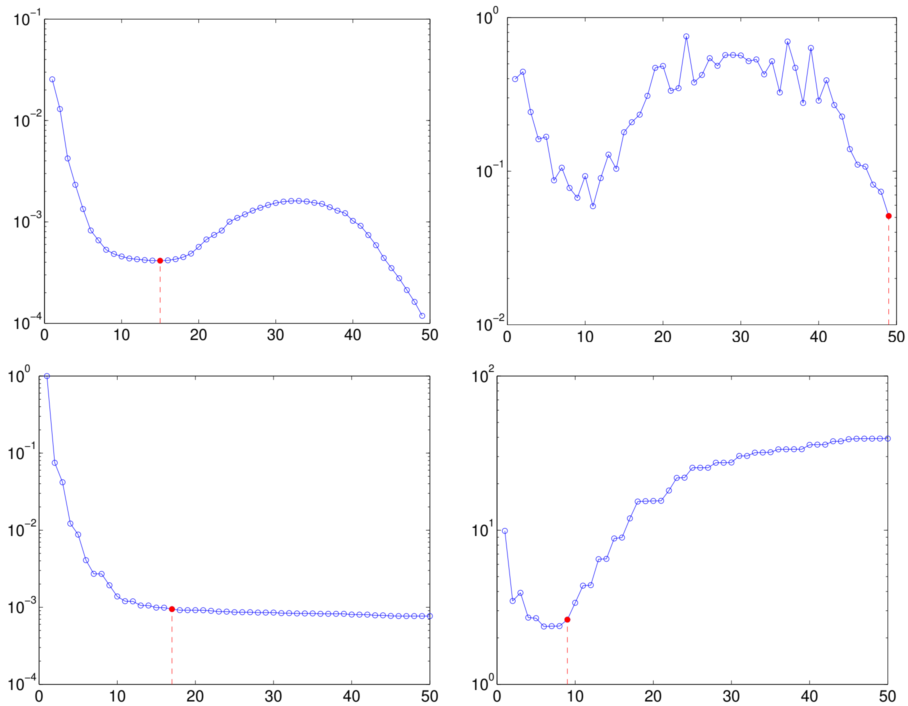
is equivalent to (1). Here $\bar{\boldsymbol{b}}$ is a suitable modification of the vector $\boldsymbol{b}$ in (1). We apply LSQR to the solution of (29) and use a parameter choice rule to determine which iterate $\boldsymbol{y}_k$ to select. This iterate is transformed to an approximate solution $\boldsymbol{x}_k$ of (1). When L is a banded matrix with an explicitly known null space of small dimension, it is not much more expensive to compute matrix-vector products with $AL_A^\dagger$ and $(AL_A^\dagger)^T$ than with A and A^T; see [14, 25] for details. Table 7 displays results achieved with this approach for the parameter choice rules that gave the best results. The restricted Regińska method performs especially well.



Method	Best	L-corner	Regińska	Restr. Regińska	Quasi-optimality
k	8	14	1	13	2
error	0.053	0.20	2.46	0.11	1.20

Figure 1: TGSVD, Phillips, $m = n = 40$, L defined by (27), noise-level $\delta = 1 \cdot 10^{-3}$: left Restr. Regińska, right Quasi-optimality.

We conclude this section with a few graphs that shed light on the performance of the parameter choice rules in our comparison. Each figure contains, for a subset of methods, a plot of the index associated to each method, and a table reporting the value of the truncation parameter k chosen and the corresponding error norm $\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\|$. Figure 1 shows an example for which the preprocessing in the restricted Regińska method is beneficial when determining the truncation index for the TGSVD method. The index chosen is marked by a red dot and by a red vertical dashed line. The quasi-optimality criterion is seen to pick a poor index. The table shows that none of the heuristic parameter choice rules is able to pick the index that yields the smallest error ($k = 8$), and that the restricted Regińska method determines the best approximation of $\hat{\boldsymbol{x}}$.



Method	Best	Restr. Regińska	Quasi-optimality	Quadrature	Ratio
k	11	15	49	17	9
error	0.29	0.44	8.42	0.73	0.32

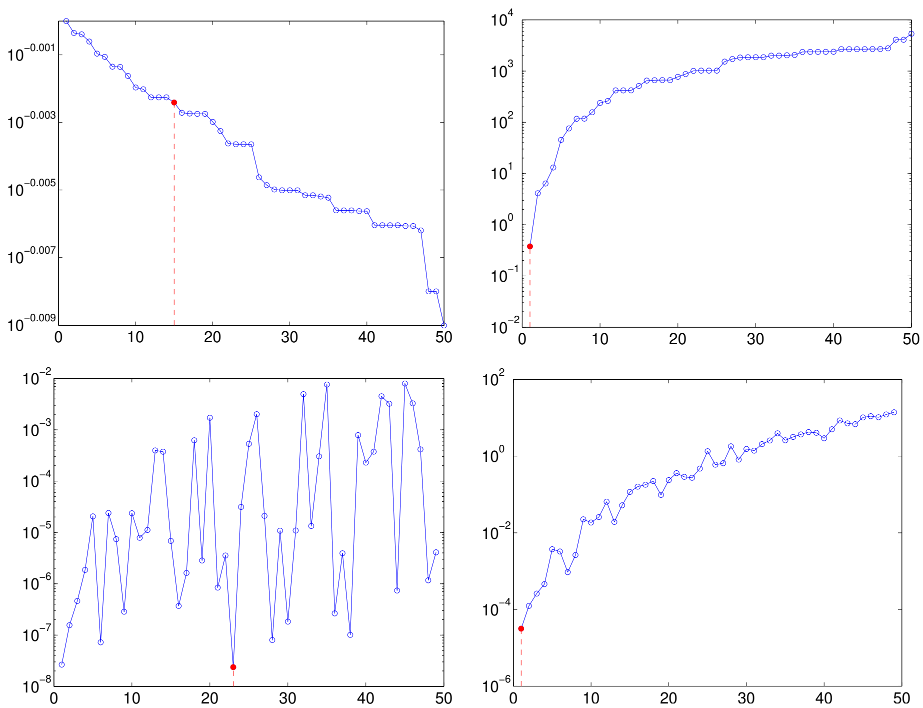
Figure 2: LSQR, Deriv2, $m = n = 100$, noise-level $\delta = 1 \cdot 10^{-3}$: NW restricted Regińska, NE Quasi-optimality, SW Quadrature, SE Ratio.

Figure 2 considers LSQR applied to Deriv2 with noise-level $1 \cdot 10^{-3}$. None of the heuristic parameter choice rules is able to determine the iteration number of the iterate $\boldsymbol{x}_{11}$ closest to $\hat{\boldsymbol{x}}$. The ratio parameter choice rule is seen to perform the best, but the restricted Regińska method also produces an acceptable error.

Finally, Figure 3 shows all parameter choice rules to perform well, except for the quasi-optimality criterion when LSQR is implemented without reorthogonalization. It is clear that reorthogonalization is mandatory when the quasi-optimality criterion is used to determine the LSQR iterate.

5 Conclusion

For many linear discrete ill-posed problems that arise in science and engineering, the noise-level δ is not explicitly known. It is therefore important to investigate



Method	Best	Restr. Regińska	Quad.	Ratio	Quasi-opt.	Q-opt./reorth.
k	1	1	15	1	23	1
error	0.021	0.021	9.21	0.021	27.21	0.021

Figure 3: LSQR, Deriv2, $m = n = 400$, L defined by (28), noise 10^{-1}: NW Quadrature, NE Ratio, SW Quasi-optimality applied to LSQR without reorthogonalization, SE Quasi-optimality applied to LSQR with reorthogonalization.

parameter choice rules that do not require knowledge about the noise-level. This paper compares the performance of several new and known parameter choice rules of this kind. For small to medium-sized problems that can be solved by TSVD, we find the residual L-curve method implemented with the L-corner algorithm and the restricted Regińska method to perform the best of the nonhybrid parameter choice rules. Some hybrid methods that use the quasi-optimality criterion as a postprocessing step perform even better. The hybrid quasi-optimality-residual L-curve and quasi-optimality-extrapolation parameter choice rules are particularly well suited to determine a suitable truncation index for the TSVD method.

For large scale problems that are solved by LSQR, the restricted Regińska and the ratio methods yield the best approximations of the desired solution; in fact, the latter parameter choice rule performs as well as the discrepancy principle. There does not seem to be a significant advantage using hybrid parameter choice rules that apply the quasi-optimality criterion when determining a suitable iterate.

We also applied heuristic parameter choice rules when the regularization matrices (27) and (28) were used. The restricted Regińska rule performs well in this setting, but many parameter choice rules do not.

Acknowledgments

We would like to thank Valia Guerra for making MATLAB code for the triangle method available and a referee for comments. LR would like to thank Giuseppe Rodriguez and Sebastiano Seatzu for an enjoyable visit to Cagliari, during which part of the work for this paper was carried out. This visit was supported by the Visiting Professor program, University of Cagliari, funded by the Regione Autonoma della Sardegna.

References

[1] A. B. Bakushinskii, Remarks on choosing a regularization parameter using quasi-optimality and ratio criterion, USSR Comp. Math. Math. Phys., 24,4 (1984), pp. 181–182.

[2] F. Bauer and S. Kindermann, Recent results on the quasi-optimality principle, J. Inv. Ill-Posed Probl., 17 (2009), pp. 5–18.

[3] F. Bauer and M. A. Lukas, Comparing parameter choice methods for regularization of ill-posed problems, Math. Comput. Simulation, 81 (2011), pp. 1795–1841.

[4] F. Bauer and M. Reiß, Regularization independent of the noise-level: an analysis of quasi-optimality, Inverse Problems, 24 (2008), 055009 (16 pp).

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

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

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

[8] D. Calvetti, G. H. Golub, and L. Reichel, Estimation of the L-curve via Lanczos bidiagonalization, BIT, 39 (1999), pp. 603–619.

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

[10] D. Calvetti, B. Lewis, and L. Reichel, GMRES, L-curves, and discrete ill-posed problems, BIT, 42 (2002), pp. 44–65.

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

[12] J. L. Castellanos, S. Gómez, and V. Guerra, The triangle method for finding the corner of the L-curve, Appl. Numer. Math., 43 (2002), pp. 359–373.

[13] M. Donatelli, Fast transforms for high order boundary conditions in deconvolution problems, BIT, 50 (2010), pp. 559–576.

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

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

[16] G. H. Golub and G. Meurant, Matrices, moments and quadrature, in Numerical Analysis 1993, eds. D. F. Griffiths and G. A. Watson, Longman, Essex, England, 1994, pp. 105–156.

[17] G. H. Golub and G. Meurant, Matrices, Moments and Quadrature with Applications, Princeton University Press, Princeton, 2010.

[18] 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.

[19] G. H. Golub and U. von Matt, Generalized cross-validation for large scale problems, J. Comput. Graph. Stat., 6 (1997), pp. 1–34.

[20] M. Hanke, Conjugate Gradient Type Methods for Ill-Posed Problems, Longman, Essex, 1995.

[21] M. Hanke, Limitations of the L-curve method in ill-posed problems, BIT, 36 (1996), pp. 287–301.

[22] M. Hanke and P. C. Hansen, Regularization methods for large-scale problems, Surv. Math. Ind., 3 (1993), pp. 253–315.

[23] P. C. Hansen, Regularization, GSVD and truncated GSVD, BIT, 29 (1989), pp. 491–504.

[24] P. C. Hansen, Analysis of the discrete ill-posed problems by means of the L-curve, SIAM Review, 34 (1992), pp. 561–580.

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

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

[27] P. C. Hansen, T. K. Jensen, and G. Rodriguez, An adaptive pruning algorithm for the discrete L-curve criterion, J. Comput. Appl. Math., 198 (2006), pp. 483–492.

[28] P. C. Hansen and D. P. O'Leary, The use of the L-curve in the regularization of discrete ill-posed problems, SIAM J. Sci. Comput., 14 (1993), pp. 1487–1503.

[29] U. Hämarik, R. Palm, and T. Raus, A family of rules for parameter choice in Tikhonov regularization of ill-posed problems with inexact noise level, J. Comput. Appl. Math., 236 (2012), pp. 2146–2157.

[30] I. Hnětynková, M. Plešinger, and Z. Strakoš, The regularizing effect of the Golub-Kahan iterative bidiagonalization and revealing the noise level in the data, BIT, 49 (2009), pp. 669–696.

[31] 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.

[32] M. A. Lukas, Robust generalized cross-validation for choosing the regularization parameter, Inverse Problems, 22 (2006), pp. 1883–1902.

[33] S. Morigi, L. Reichel, F. Sgallari, and F. Zama, Iterative methods for ill-posed problems and semiconvergent sequences, J. Comput. Appl. Math., 193 (2006), pp. 157–167.

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

[35] A. Neuman, L. Reichel, and H. Sadok, Algorithms for range restricted iterative methods for linear discrete ill-posed problems, Numer. Algorithms, 59 (2012), pp. 325–331.

[36] C. C. Paige and M. A. Saunders, LSQR: an algorithm for sparse linear equations and sparse least squares, Trans. Math. Software, 8 (1982), pp. 43–71.

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

[38] L. Reichel, G. Rodriguez, and S. Seatzu, Error estimates for large-scale ill-posed problems, Numer. Algorithms, 51 (2009), pp. 341–361.

[39] L. Reichel and H. Sadok, A new L-curve for ill-posed problems, J. Comput. Appl. Math., 219 (2008), pp. 493–508.

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

[41] T. Regińska, A regularization parameter in discrete ill-posed problems, SIAM J. Sci. Comput., 17 (1996), pp. 740–749.

[42] C. B. Shaw, Jr., Improvements of the resolution of an instrument by numerical solution of an integral equation, J. Math. Anal. Appl., 37 (1972), pp. 83–112.

[43] A. Tikhonov and V. Arsenin, Solutions of Ill-Posed Problems, Wiley, New York, 1977.

[44] C. R. Vogel, Non-convergence of the L-curve regularization parameter selection method, Inverse Problems, 12 (1996), pp. 535–547.