



Golub–Kahan vs. Monte Carlo: a comparison of bidiagonalization and a randomized SVD method for the solution of linear discrete ill-posed problems

Xianglan Bai¹ · Alessandro Buccini² · Lothar Reichel¹

Received: 24 February 2020 / Accepted: 5 March 2021
© The Author(s) 2021

Abstract

Randomized methods can be competitive for the solution of problems with a large matrix of low rank. They also have been applied successfully to the solution of large-scale linear discrete ill-posed problems by Tikhonov regularization (Xiang and Zou in *Inverse Probl* 29:085008, 2013). This entails the computation of an approximation of a partial singular value decomposition of a large matrix A that is of numerical low rank. The present paper compares a randomized method to a Krylov subspace method based on Golub–Kahan bidiagonalization with respect to accuracy and computing time and discusses characteristics of linear discrete ill-posed problems that make them well suited for solution by a randomized method.

Keywords Discrete ill-posed problem · Golub–Kahan bidiagonalization · Randomized SVD · Tikhonov regularization · Discrepancy principle

Mathematics Subject Classification 65F10 · 65F22

Communicated by Gunilla Kreiss.

✉ Alessandro Buccini
alessandro.buccini@unica.it

Xianglan Bai
xbai@kent.edu

Lothar Reichel
reichel@math.kent.edu

¹ Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA

² Dipartimento di Matematica e Informatica, Università di Cagliari, Cagliari 09123, Italy

1 Introduction

We are concerned with the solution of linear least-squares problems

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a large matrix, whose singular values “cluster” at the origin, and $\mathbf{b} \in \mathbb{R}^m$. In particular, \mathbf{A} has many “tiny” singular values of different orders of magnitude. Least-squares problem with a matrix of this kind commonly are referred to as linear discrete ill-posed problems. They arise, for instance, from the discretization of Fredholm integral equations of the first kind; see, e.g., [10, 19]. Applications of this kind of least-squares problems include image reconstruction and remote sensing. Throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm or the spectral matrix norm. Both the situations when $m \geq n$ and when $m < n$ will be considered.

The vector \mathbf{b} in linear discrete ill-posed problems that arise in applications typically represents measured data and is contaminated by a measurement error $\mathbf{e} \in \mathbb{R}^m$. Thus,

$$\mathbf{b} = \mathbf{b}_{\text{exact}} + \mathbf{e}, \quad (2)$$

where $\mathbf{b}_{\text{exact}}$ denotes the unknown error-free vector associated with \mathbf{b} . We are interested in determining the solution, $\mathbf{x}_{\text{exact}}$, of minimal Euclidean norm of the unavailable least-squares problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}_{\text{exact}}\|.$$

Due to the error \mathbf{e} in \mathbf{b} and the presence of small positive singular values of \mathbf{A} , the solution of (1) of minimal Euclidean norm typically is not a useful approximation of $\mathbf{x}_{\text{exact}}$. To determine a meaningful approximation of $\mathbf{x}_{\text{exact}}$, one generally replaces the minimization problem (1) by a nearby problem, whose solution is less sensitive to the error \mathbf{e} . This replacement is known as *regularization*. One of the most popular regularization methods is due to Tikhonov. It replaces (1) by a penalized least-squares problem of the form

$$\min_{\mathbf{x} \in \mathbb{R}^n} \{\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \mu \|\mathbf{L}\mathbf{x}\|^2\}, \quad (3)$$

where $\mathbf{L} \in \mathbb{R}^{p \times n}$ is a regularization matrix and $\mu \geq 0$ a regularization parameter. We require that $\mathcal{N}(\mathbf{A}) \cap \mathcal{N}(\mathbf{L}) = \{\mathbf{0}\}$, where $\mathcal{N}(\mathbf{M})$ denotes the null space of the matrix \mathbf{M} . Then the penalized least-squares problem (3) has a unique solution

$$\mathbf{x}_\mu = (\mathbf{A}^T \mathbf{A} + \mu \mathbf{L}^T \mathbf{L})^{-1} \mathbf{A}^T \mathbf{b} \quad (4)$$

for any $\mu > 0$; the superscript T denotes transposition. Common choices of regularization matrices \mathbf{L} include the identity matrix, denoted by \mathbf{I} , and discrete approximations

of differential operators; see, e.g., [5,7,19,23]. The value of the regularization parameter $\mu > 0$ determines how sensitive the vector (4) is to the error in \mathbf{b} and how close it is to the desired vector $\mathbf{x}_{\text{exact}}$.

We will determine μ with the aid of the discrepancy principle; see below. This requires that a bound ϵ for the error \mathbf{e} be known, i.e.,

$$\|\mathbf{e}\| \leq \epsilon, \quad (5)$$

and that $\mathbf{b}_{\text{exact}} \in \mathcal{R}(A)$, where $\mathcal{R}(A)$ denotes the range of A . If these requirements are not satisfied, then other methods, including the L-curve criterion and generalized cross validation can be used to determine a suitable value of μ ; see, e.g., [11,25,26,29] for discussions and illustrations.

The Tikhonov solution (4) is said to satisfy the *discrepancy principle* if

$$\|\mathbf{b} - A\mathbf{x}_\mu\| = \eta\epsilon, \quad (6)$$

where $\eta > 1$ is a user-chosen parameter that is independent of ϵ . When ϵ in (5) is known to be an accurate estimate of $\|\mathbf{e}\|$ and \mathbf{e} represents white Gaussian noise, then generally η is chosen to be close to unity. Equation (6) has a unique solution $\mu_{\text{discr}} = \mu > 0$ for many reasonable values of $\eta\epsilon > 0$; see, e.g., [2]. Several zero-finders for determining μ_{discr} are described in [30]. A proof in a Hilbert space setting that $\mathbf{x}_\mu \rightarrow \mathbf{x}_{\text{exact}}$ as $\epsilon \searrow 0$ can be found in [10].

It is the purpose of the present paper to compare a solution method for large-scale Tikhonov minimization problems (3) based on partial Golub–Kahan bidiagonalization of A to a randomized solution method. Partial Golub–Kahan bidiagonalization is the basis for the possibly most popular Krylov subspace methods for the solution of large-scale problems (3) with a nonsymmetric or symmetric indefinite matrix A ; see, e.g., [2,15,18,19,22,24] for discussions and illustrations of this solution approach. Iterative solution methods that are based on the Arnoldi process instead of Golub–Kahan bidiagonalization are competitive for certain problems, but may fail to determine accurate approximations of $\mathbf{x}_{\text{exact}}$ for some problems; see [6,14,15,27] for discussions and applications of the Arnoldi process to large-scale linear discrete ill-posed problems. We therefore focus on Golub–Kahan bidiagonalization in the present paper.

When solving (3) by application of ℓ steps of Golub–Kahan bidiagonalization, the matrix A is replaced by an approximation of rank at most ℓ . Typically, $1 \leq \ell \ll \max\{m, n\}$ in applications. Thus, Golub–Kahan bidiagonalization applied to the solution of (3) replaces A by a low-rank approximation of the matrix A , and then solves the low-rank problem instead of (3).

Randomized solution methods for the solution of large-scale problems have received considerable attention; see Halko et al. [17] for a survey. When applied to the solution of (3), these methods also determine a low-rank approximation of A . They compute an approximate solution of the original problem by replacing the given matrix by its low-rank approximation, and then solve the low-rank problem so obtained. Xiang and Zou [33,34] describe applications of this approach to the solution of large-scale Tikhonov minimization problems (3). To the best of our knowledge, very few comparisons of randomized and Krylov subspace-based solution methods

for linear discrete ill-posed problems are available in the literature. It is quite natural to compare these approaches to the solution of large-scale linear discrete ill-posed problems, because they both determine low-rank approximations of the large matrix A . Vatankeh et al. [32] show that randomized methods may be faster than Krylov subspace methods based on Golub–Kahan bidiagonalization for certain problems. We illustrate that, for some linear discrete ill-posed problems, methods based on Golub–Kahan bidiagonalization are competitive and we seek to shed light on for which kinds of linear discrete ill-posed problems Golub–Kahan bidiagonalization may be preferable.

This paper is organized as follows. Section 2 reviews methods based on partial Golub–Kahan bidiagonalization of the matrix A described in [2,22] for the solution of large-scale Tikhonov regularization problems (3). Section 3 outlines the randomized method proposed by Xiang and Zou [33]. The randomized method discussed in [34] also is commented on. Section 4 presents computed results, and Sect. 5 contains concluding remarks.

2 Solution methods based on Golub–Kahan bidiagonalization

This section reviews solution methods for the Tikhonov minimization problem described in [2,22]. They are based on reducing the matrix A to a small bidiagonal matrix by the application of $1 \leq \ell \ll \min\{m, n\}$ steps of Golub–Kahan bidiagonalization to A . The number of steps is chosen as small as possible so that the computed solution can satisfy the discrepancy principle. Thus, application of ℓ steps of Golub–Kahan bidiagonalization to A with initial vector \mathbf{b} gives the decompositions

$$AV_\ell = U_{\ell+1}\tilde{C}_\ell, \quad A^T U_\ell = V_\ell C_\ell^T, \quad (7)$$

where the matrices $U_{\ell+1} \in \mathbb{R}^{m \times (\ell+1)}$ and $V_\ell \in \mathbb{R}^{n \times \ell}$ have orthonormal columns, $U_\ell \in \mathbb{R}^{m \times \ell}$ consists of the first ℓ columns of $U_{\ell+1}$, and

$$U_{\ell+1}\mathbf{e}_1 = \mathbf{b}/\|\mathbf{b}\|. \quad (8)$$

Here and throughout this paper $\mathbf{e}_1 = [1, 0, \dots, 0]^T$ denotes the first axis vector of appropriate dimension. The range of V_ℓ is the Krylov subspace

$$\mathbb{K}_\ell(A^T A, A^T \mathbf{b}) = \text{span}\{A^T \mathbf{b}, (A^T A)A^T \mathbf{b}, \dots, (A^T A)^{\ell-1} A^T \mathbf{b}\}. \quad (9)$$

Further, the matrix

$$\tilde{C}_\ell = \begin{bmatrix} \rho_1 & & & & 0 \\ \sigma_2 & \rho_2 & & & \\ & \ddots & \ddots & & \\ & & \sigma_{\ell-1} & \rho_{\ell-1} & \\ & & & \sigma_\ell & \rho_\ell \\ 0 & & & & \sigma_{\ell+1} \end{bmatrix} \in \mathbb{R}^{(\ell+1) \times \ell}$$

is lower bidiagonal with positive entries σ_k and ρ_k , and $C_\ell \in \mathbb{R}^{\ell \times \ell}$ is obtained by removing the last row of \bar{C}_ℓ . We assume that ℓ is small enough so that the decompositions (7) with the described properties exist. This is the generic situation. The dominating computational effort required to determine the decompositions (7) is the sequential evaluation of ℓ matrix–vector products with each one the matrices A and A^T ; see, e.g., [16, Sect. 10.4.4] for an algorithm.

Following [22], we compute an approximate solution of (3) by minimizing over the Krylov subspace (9) instead of over \mathbb{R}^n . Thus, we solve

$$\min_{\mathbf{x} \in \mathbb{K}_\ell(A^T A, A^T \mathbf{b})} \{\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \mu \|\mathbf{L}\mathbf{x}\|^2\}, \quad (10)$$

which, by using the representation $\mathbf{x} = V_\ell \mathbf{y}$, and the relations (7) and (8), can be expressed as

$$\min_{\mathbf{y} \in \mathbb{R}^\ell} \{\|\bar{C}_\ell \mathbf{y} - \mathbf{e}_1 \|\mathbf{b}\| \|^2 + \mu \|L V_\ell \mathbf{y}\|^2\}. \quad (11)$$

Denote the solution of (11) by $\mathbf{y}_{\mu, \ell}$. Then $\mathbf{x}_{\mu, \ell} = V_\ell \mathbf{y}_{\mu, \ell}$ is an approximate solution of (3). We point out that the problems (3) and (10) only differ in the spaces over which they are minimized. The randomized method of Sect. 3 gives a minimization problem that differs in several ways from the problem (3).

First consider the situation when $L = I$. Then it is shown in [1, Theorem 5.1] that

$$\|\mathbf{A}\mathbf{x}_{\mu, \ell} - \mathbf{b}\| = \|\bar{C}_\ell \mathbf{y}_{\mu, \ell} - \mathbf{e}_1 \|\mathbf{b}\| \|. \quad (12)$$

It therefore suffices to choose $\mu > 0$ so that the reduced problem on the right-hand side satisfies the discrepancy principle; see [2] for details. It follows that it is quite inexpensive to determine a value of $\mu > 0$ such that the approximate solution $\mathbf{x}_{\mu, \ell} = V_\ell \mathbf{y}_{\mu, \ell}$ of (3) with $L = I$ satisfies (6). A discussion on how this can be done by using Newton's method can be found in [2]; other zero-finders are discussed in [30].

The expressions (12) decrease as ℓ increases. This follows from the fact that the dimension of the Krylov subspace in (10) increases with ℓ and that the subspaces are nested, i.e., $\mathbb{K}_\ell(A^T A, A^T \mathbf{b}) \subset \mathbb{K}_{\ell+1}(A^T A, A^T \mathbf{b})$ for $\ell = 1, 2, \dots$. We choose the number of bidiagonalization steps, ℓ , as small as possible to satisfy the discrepancy principle for some $0 < \mu < \infty$, i.e., we choose ℓ so that

$$\|\bar{C}_\ell \mathbf{y}_{\mu, \ell} - \mathbf{e}_1 \|\mathbf{b}\| \| < \eta \epsilon \leq \|\bar{C}_{\ell-1} \mathbf{y}_{\mu, \ell-1} - \mathbf{e}_1 \|\mathbf{b}\| \|.$$

Further details on the choice of ℓ are described in [2].

We turn to the case when $L \neq I$. This situation can be handled by several approaches; see, e.g., [3, 9, 22]. In the numerical examples of Sect. 4, we will apply the method described in [22]. Let $L \in \mathbb{R}^{p \times n}$ and assume that ℓ in (7) satisfies $1 \leq \ell \leq \min\{p, n\}$. Compute the QR factorization $Q_\ell R_\ell = L V_\ell$, where $Q_\ell \in \mathbb{R}^{n \times \ell}$ has orthonormal columns and $R_\ell \in \mathbb{R}^{\ell \times \ell}$ is upper triangular. Then (11) becomes

$$\min_{\mathbf{y} \in \mathbb{R}^\ell} \{\|\bar{C}_\ell \mathbf{y} - \mathbf{e}_1 \|\mathbf{b}\| \|^2 + \mu \|R_\ell \mathbf{y}\|^2\}.$$

Typically, the matrix R_ℓ is nonsingular and not very ill-conditioned. Then the change of variables $\mathbf{z} = R_\ell \mathbf{y}$ results in the minimization problem

$$\min_{\mathbf{z} \in \mathbb{R}^\ell} \{ \|\bar{C}_\ell R_\ell^{-1} \mathbf{z} - \mathbf{e}_1\| \|\mathbf{b}\|^2 + \mu \|\mathbf{z}\|^2 \}.$$

It can be shown that if μ is determined so that the solution $\mathbf{z}_{\mu,\ell}$ satisfies $\|\bar{C}_\ell R_\ell^{-1} \mathbf{z} - \mathbf{e}_1\| \|\mathbf{b}\| = \eta\epsilon$, then the associated approximate solution $\mathbf{x}_{\mu,\ell} = V_\ell R_\ell^{-1} \mathbf{z}_{\mu,\ell}$ satisfies the discrepancy principle (6); see [22] for details. Hence, it is quite cheap to determine an approximate solution of (3) that satisfies (6) also when $L \neq I$.

3 Randomized solution methods

The reduced singular value decomposition (SVD) of the matrix $A \in \mathbb{R}^{m \times n}$, with $m \geq n$, is of the form

$$A = U \Sigma V^T, \quad (13)$$

where the matrices $U \in \mathbb{R}^{m \times n}$ and $V \in \mathbb{R}^{n \times n}$ have orthonormal columns, and

$$\Sigma = \text{diag}[\sigma_1, \sigma_2, \dots, \sigma_n] \in \mathbb{R}^{m \times n}, \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0.$$

The diagonal entries σ_j are known as the singular values of A . An analogous decomposition is available when $m < n$; see, e.g., [16, Sect. 2.4] for details on the SVD of a matrix.

The computation of the SVD (13) requires $\mathcal{O}(\max\{m, n\} \min\{m, n\}^2)$ flops and, therefore, is expensive when m and n are large. Randomized SVD methods determine an approximation of the factorization (13) and are less expensive; see, e.g., Halko et al. [17]. Xiang and Zou [33] describe how randomized SVD methods can be applied to the solution of large-scale Tikhonov minimization problems (3) when $L = I$. We will outline their approaches and compare the performance of these randomized methods to the Golub–Kahan bidiagonalization method of Sect. 2 in Sect. 4.

Xiang and Zou [34] also describe several randomized approaches to the solution of Tikhonov regularization problems in general form (3). Some of these methods are based on first transforming (3) to an equivalent problem with $L = I$, similarly as outlined at the end of Sect. 2, while others apply a randomized generalized SVD. We will not discuss the latter methods in the present paper.

We first describe the method proposed by Xiang and Zou [33] for the approximate solution of (3) when $m \geq n$ and $L = I$. Let the entries of the matrix $\Omega_\ell \in \mathbb{R}^{n \times \ell}$, where $1 \leq \ell \ll n$, be identically and normally distributed random numbers with zero mean, and compute the QR factorization

$$Q_\ell R_\ell = A \Omega_\ell,$$

where $Q_\ell \in \mathbb{R}^{m \times \ell}$ has orthonormal columns and $R_\ell \in \mathbb{R}^{\ell \times \ell}$ is upper triangular. The matrix R_ℓ is assumed to be nonsingular in [33] and we will, for now, assume the same. Then the columns of Q_ℓ form an orthonormal basis for $\mathcal{R}(A\Omega_\ell)$. Let $B = Q_\ell^T A \in \mathbb{R}^{\ell \times n}$ and compute the reduced SVD,

$$B = \widehat{W} \widehat{\Sigma} \widehat{V}^T, \quad (14)$$

where the matrices $\widehat{W} \in \mathbb{R}^{\ell \times \ell}$ and $\widehat{V} \in \mathbb{R}^{n \times \ell}$ have orthonormal columns, and $\widehat{\Sigma} \in \mathbb{R}^{\ell \times \ell}$ is a diagonal matrix with nonnegative diagonal entries arranged in decreasing order. The right-hand side of

$$Q_\ell B = Q_\ell Q_\ell^T A = Q_\ell \widehat{W} \widehat{\Sigma} \widehat{V}^T \quad (15)$$

is an approximation of the SVD of A (13). The decomposition (15) is much cheaper to compute than (13) when m and n are large and $1 \leq \ell \ll n \leq m$. The following proposition provides bounds for the closeness of A and $Q_\ell Q_\ell^T A$.

Proposition 1 *Suppose that $A \in \mathbb{R}^{m \times n}$ has the singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\min\{m,n\}} \geq 0$. Let $\Omega_\ell \in \mathbb{R}^{n \times \ell}$ be a Gaussian matrix with $\ell := k + p \leq \min\{m, n\}$ and $p \geq 4$. Let the columns of Q_ℓ form an orthonormal basis for $\mathcal{R}(A\Omega_\ell)$. Then*

$$\sigma_{\ell+1} \leq \|A - Q_\ell Q_\ell^T A\| \leq (1 + 6\sqrt{\ell p \log p})\sigma_{k+1} + 3(\ell \sum_{j>k} \sigma_j^2)^{1/2} \quad (16)$$

with probability not less than $1 - 3p^{-p}$.

Proof The left-hand side inequality is a consequence of the Eckart and Young theorem [8], and the right-hand side inequality is shown by Halko et al. [17, Corollary 10.9]. \square

We will comment below on the significance of the upper bound (16). In order for this bound to be small, we have to choose k large enough so that σ_{k+1} is small, and p large enough so that the right-hand side inequality (16) holds with high probability. Common choices of p are 5 or 10.

The computational cost for determining the matrices Q_ℓ , \widehat{W} , $\widehat{\Sigma}$, and \widehat{V} in (15) is comprised of $\mathcal{O}(m\ell^2)$ flops for determining Q_ℓ from Ω_ℓ , ℓ matrix–vector product evaluations with A (to form $A\Omega_\ell$) and ℓ matrix–vector product evaluations with A^T (to form $B = Q_\ell^T A$), as well as $\mathcal{O}(n\ell^2)$ flops for the computation of the SVD of B (14). Some of these computations can be implemented efficiently by using high-level BLAS; see, e.g., [12] for a discussion on implementation issues.

We will use the decomposition (15) to determine an approximate solution of (3). Replacing A by this decomposition in (3) gives

$$\min_{\mathbf{x} \in \mathbb{R}^n} \{ \|Q_\ell Q_\ell^T A \mathbf{x} - Q_\ell Q_\ell^T \mathbf{b}\|^2 + \mu \|L \mathbf{x}\|^2 \} + \|(I - Q_\ell Q_\ell^T) \mathbf{b}\|^2, \quad (17)$$

which can be expressed as

$$\min_{\mathbf{x} \in \mathbb{R}^n} \{ \|\widehat{W} \widehat{\Sigma} \widehat{V}^T \mathbf{x} - Q_\ell^T \mathbf{b}\|^2 + \mu \|L \mathbf{x}\|^2 \} + \|(I - Q_\ell Q_\ell^T) \mathbf{b}\|^2. \quad (18)$$

Since we would like a solution \mathbf{x} of minimal Euclidean norm, it is natural to require the solution to be of the form $\mathbf{x} = \widehat{V}\mathbf{y}$ for some $\mathbf{y} \in \mathbb{R}^\ell$. Substitution into (18) gives the minimization problem

$$\min_{\mathbf{y} \in \mathbb{R}^\ell} \{ \|\widehat{\Sigma}\mathbf{y} - \widehat{W}^T Q_\ell^T \mathbf{b}\|^2 + \mu \|L\widehat{V}\mathbf{y}\|^2 \}. \quad (19)$$

This problem differs from (3) in several ways: i) A is replaced by $Q_\ell Q_\ell^T A$, ii) the term $\|(I - Q_\ell Q_\ell^T)\mathbf{b}\|^2$ in (18) is ignored, and iii) the space \mathbb{R}^n is replaced by $\mathcal{R}(\widehat{V})$. The differences between the minimization problems (3) and (19) are small if p and k , and therefore $\ell = p + k$, in Proposition 1 are sufficiently large. In particular, the choice of k has to be large enough in relation to how quickly the singular values of A decay to zero with increasing index.

Let \mathbf{x} denote the solution of (17) with $\mu > 0$ chosen so that \mathbf{x} satisfies the discrepancy principle (6). The discrepancy principle suggests that ℓ be chosen large enough so that

$$\|(A - Q_\ell Q_\ell^T A)\mathbf{x}\| \leq \|A - Q_\ell Q_\ell^T A\| \|\mathbf{x}\| \leq \eta' \epsilon,$$

where the parameter ϵ is the same as in (6), and $0 < \eta' \leq \eta$ is a user-chosen parameter with η the same as in (6). An accurate upper bound for $\|A - Q_\ell Q_\ell^T A\|$ can be determined with high probability by evaluating $\|(A - Q_\ell Q_\ell^T A)\mathbf{w}\|$ for sufficiently many random vectors with normally distributed entries with zero mean; see [17, Eq. (4.3) and Lemma 4.1]. The evaluation of such a bound increases the computational effort required by the randomized method. Moreover, we would like ℓ to be large enough so that

$$\|(I - Q_\ell Q_\ell^T)\mathbf{b}\| \leq \eta' \epsilon.$$

We illustrate in Sect. 4 that for some problems the parameters p and k in (16) have to be chosen too large to make the randomized solution method of this section competitive with the Krylov subspace method of Sect. 2.

There is a small probability that in a particular application of the randomized method, columns of the matrix Q_ℓ are singular vectors associated with “tiny” singular values of A . These singular vectors typically “oscillate” a lot, i.e., the vector entries as a function of their index number can be thought of as the discretization of a highly oscillatory function. The presence of such vectors in the solution subspace $\mathcal{R}(Q_\ell)$ typically would result in an a highly oscillatory, and therefore undesired, approximate solution of (3). This phenomenon may be considered an instability of the randomized method. However, we hasten to add that we have not observed this instability in any one of numerous computed examples that we have carried out. The occurrence of this instability, indeed, is rare.

When $L = I$, we have $\|L\widehat{V}\mathbf{y}\| = \|\mathbf{y}\|$. Otherwise, we compute the QR factorization $\widehat{Q}\widehat{R} = L\widehat{V}$, where $\widehat{Q} \in \mathbb{R}^{p \times \ell}$ has orthonormal columns and $\widehat{R} \in \mathbb{R}^{\ell \times \ell}$ is upper triangular. When the matrix \widehat{R} is of full rank and fairly well-conditioned, we proceed similarly as described at the end of Sect. 2.

Now consider the case when, numerically, $\text{rank}(A\Omega_\ell) < \ell$. This situation may arise when the matrix A is of numerical rank less than ℓ . Then we compute the SVD

$$R_\ell = U_\ell \Sigma_\ell V_\ell^T,$$

where the matrices $U_\ell = [\mathbf{u}_{\ell,1}, \mathbf{u}_{\ell,2}, \dots, \mathbf{u}_{\ell,\ell}] \in \mathbb{R}^{\ell \times \ell}$ and $V_\ell \in \mathbb{R}^{\ell \times \ell}$ are orthogonal and $\Sigma_\ell \in \mathbb{R}^{\ell \times \ell} = \text{diag}[\sigma_{\ell,1}, \sigma_{\ell,2}, \dots, \sigma_{\ell,\ell}]$ with $\sigma_{\ell,1} \geq \sigma_{\ell,2} \geq \dots \geq \sigma_{\ell,\ell} \geq 0$. Let $\sigma_{\ell,j}$ be the smallest numerically nonvanishing diagonal entry. Then, numerically, the columns of $Q_{\ell,j} := Q_\ell[\mathbf{u}_{\ell,1}, \mathbf{u}_{\ell,2}, \dots, \mathbf{u}_{\ell,j}] \in \mathbb{R}^{m \times j}$ form an orthonormal basis for $\mathcal{R}(A\Omega_\ell)$, and we replace the matrix Q_ℓ in (15), (17), and (19) by $Q_{\ell,j}$.

We turn to the situation when $m < n$. Following Xiang and Zou [33], let $\Omega_\ell \in \mathbb{R}^{\ell \times m}$, with $1 \leq \ell \ll m$, be a random matrix with the same kind of entries as above, and let the columns of $Q_\ell \in \mathbb{R}^{n \times \ell}$ form an orthonormal basis for a linear space that contains $\mathcal{R}((\Omega_\ell A)^T)$. We compute Q_ℓ by evaluating the QR factorization of $(\Omega_\ell A)^T$. Then we calculate the SVD of AQ_ℓ ,

$$AQ_\ell = \widehat{U} \widehat{\Sigma} \widehat{W}^T,$$

where the matrices $\widehat{U} \in \mathbb{R}^{n \times \ell}$ and $\widehat{W} \in \mathbb{R}^{\ell \times \ell}$ have orthonormal columns, and $\widehat{\Sigma} \in \mathbb{R}^{\ell \times \ell}$ is a diagonal matrix with nonnegative diagonal entries arranged in decreasing order. The expression

$$AQ_\ell Q_\ell^T = \widehat{U} \widehat{\Sigma} \widehat{W}^T Q_\ell^T$$

is an approximation of the SVD of A .

We determine an approximate solution of (3) by solving

$$\min_{\mathbf{x} \in \mathcal{R}(Q_\ell)} \{\|AQ_\ell Q_\ell^T \mathbf{x} - \widehat{U} \widehat{U}^T \mathbf{b}\|^2 + \mu \|L\mathbf{x}\|^2\}, \quad (20)$$

which, with $\mathbf{x} = Q_\ell \widehat{W} \mathbf{y}$, can be written as

$$\min_{\mathbf{y} \in \mathbb{R}^\ell} \{\|\widehat{\Sigma} \mathbf{y} - \widehat{U}^T \mathbf{b}\|^2 + \mu \|LQ_\ell \widehat{W} \mathbf{y}\|^2\}. \quad (21)$$

Finally, let $\widetilde{R} \in \mathbb{R}^{\ell \times \ell}$ be the upper triangular matrix in a QR factorization of $LQ_\ell \widehat{W}$. Since, generally, \widetilde{R} is nonsingular and not very ill-conditioned, we may transform the Tikhonov minimization problem (21) to standard form by the change of variables $\mathbf{z} = \widetilde{R} \mathbf{y}$.

We conclude this section with a discussion on the application of the discrepancy principle, and first consider the situation when $m \geq n$. Then we solve (17) by computing the solution of (19). Assume that the error \mathbf{e} in \mathbf{b} is normally distributed with zero mean and variance ϵ^2 . Then, since $Q_\ell Q_\ell^T$ is an orthogonal projector, the variance of $Q_\ell Q_\ell^T \mathbf{e}$ is $\frac{\ell}{m} \epsilon^2$. Therefore, when determining the regularization parameter μ for the problem (17), we replace ϵ by $\sqrt{\frac{\ell}{m}} \epsilon$ in (6).

When $m < n$, we solve (20) by computing the solution of (21). Since $\widehat{U}\widehat{U}^T$ is an orthogonal projector, the variance of $\widehat{U}\widehat{U}^T \mathbf{e}$ is $\frac{\ell}{n}\epsilon^2$. Therefore, when determining the regularization parameter μ for the minimization problem (21), we replace ϵ by $\sqrt{\frac{\ell}{n}}\epsilon$ in (6).

The value of $\ell = p + k$ affects both the quality of the computed solution and the computing time required. This value has to be large enough so that $AQ_\ell Q_\ell^T$ is a sufficiently accurate approximations of A . The computed examples of the following section illustrate this.

4 Computed examples

The examples reported in this section show the performance of the methods discussed in the previous sections. All computations were carried out on a Windows computer with an i7-8750H @2.2 GHz CPU and 16 GB of memory. The implementations were done in MATLAB R2018b.

The noise level is defined by

$$\delta = \frac{\|\mathbf{e}\|}{\|\mathbf{b}_{\text{exact}}\|}.$$

In all experiments, the regularization parameter was determined with the aid of the discrepancy principle and computed by Newton's method as outlined above with initial iterate $\mu_0 = 0$. More details on the application of Newton's method can be found in [2]. This method also was used to determine the regularization parameter in the randomized SVD (RSVD) method in a similar way.

We consider several examples described in Regularization Tools by Hansen [20] and in IR Tools by Gazzola et al. [13]. Both Regularization Tools and IR Tools are program packages written in MATLAB. Problems in both one and two space-dimensions will be discussed. We compare the performance of the Krylov method of Sect. 2 and the randomized SVD method of Sect. 3. The computed examples show that rapid decay of the singular values to zero with increasing index number is essential for the success of the randomized method. We will refer to the Krylov subspace-based Tikhonov regularization method of Sect. 2 as “K-Tikhonov”, and to the Tikhonov regularization method based on the randomized SVD technique as “R-Tikhonov”. Throughout this section, ℓ denotes the number of bidiagonalization steps carried out by the Golub–Kahan bidiagonalization method, as well as the number of columns of the random matrix Ω_ℓ when $m \geq n$, or the number of rows of the random matrix when $m < n$; see Sect. 3.

Before comparing the two methods, we would like to discuss the computational cost of the K-Tikhonov and R-Tikhonov methods. Let us assume that $\ell \ll n$. Then for K-Tikhonov, the computational cost is dominated by the ℓ matrix–vector product evaluations with A and the ℓ matrix–vector product evaluations with A^T . The R-Tikhonov method requires the evaluation of $A\Omega_\ell$ and $A^T Q_\ell$. The flop count for these evaluations in K-Tikhonov and R-Tikhonov is the same, and of order $\mathcal{O}(mn\ell)$, but

the evaluations in R-Tikhonov can be implemented as matrix–matrix products, while this is not possible in K-Tikhonov, because in Golub–Kahan bidiagonalization the columns of the matrices V_ℓ and $U_{\ell+1}$ are determined sequentially one-by-one. As we will illustrate in the following, when A is stored as a matrix, the matrix–matrix product evaluations in R-Tikhonov are faster than the matrix–vector product evaluations in K-Tikhonov. However, when A is not explicitly stored and it therefore is not possible to evaluate the matrix–matrix products $A\Omega_\ell$ and $A^T Q_\ell$ efficiently, i.e., when we need to compute 2ℓ matrix–vector products (one for each column of Ω_ℓ and Q_ℓ), the computing time required by K-Tikhonov and R-Tikhonov to evaluate the necessary matrix–vector products is almost identical.

Shaw. We first consider the Shaw test problem described in [31]. It is an integral equation of the first kind with a smooth kernel in one space-dimension. MATLAB code that gives a discretization of this integral equation is provided in [20]. This code gives the matrix $A \in \mathbb{R}^{2048 \times 2048}$ and a vector $\mathbf{x}_{\text{exact}} \in \mathbb{R}^{2048}$, from which we compute $\mathbf{b}_{\text{exact}} = A\mathbf{x}_{\text{exact}}$. We add a “noise vector” $\mathbf{e} \in \mathbb{R}^{2048}$ to $\mathbf{b}_{\text{exact}}$ that models white Gaussian noise with noise level $\delta = 0.01$ to obtain the “available” noise-contaminated data vector \mathbf{b} ; cf. (2). The moderate dimension of this problem allows us to compute the solution of the Tikhonov regularized problem also in the full space since it is possible to explicitly compute the SVD of the matrix A . We refer to the latter approach as “standard Tikhonov”. The regularization matrix L used in this example is the discrete Laplacian in one space-dimension.

We apply the K-Tikhonov and R-Tikhonov methods for different dimensions ℓ of the solution subspace and compare the results obtained with those obtained with standard Tikhonov, i.e., the solution of (3). In particular, we are interested in comparing the quality of the computed approximations of $\mathbf{x}_{\text{exact}}$ determined by the different methods, as well as in the timings. Let \mathbf{x} denote an approximate solution computed by one of the methods considered. We define the relative reconstruction error

$$\text{RRE}(\mathbf{x}) = \frac{\|\mathbf{x} - \mathbf{x}_{\text{exact}}\|}{\|\mathbf{x}_{\text{exact}}\|}.$$

Figure 1 reports RREs and timings for the K-Tikhonov and R-Tikhonov methods. The horizontal axes in the subfigures show the dimension, ℓ , of the solution subspaces for the K-Tikhonov and R-Tikhonov method. We can observe that the RRE for R-Tikhonov is slightly smaller than for K-Tikhonov for $\ell > 5$. Moreover, the CPU time required for the computation of the solution with R-Tikhonov is smaller than the CPU time needed for the computation of the solution with K-Tikhonov for solution subspaces of the same dimension. Finally, we note that the RRE obtained with K-Tikhonov rapidly converges to the RRE of the solution determined by standard Tikhonov, while the RRE for the solutions determined by R-Tikhonov typically is smaller. The difference in the quality of the computed solutions is made possibly by the facts that the computed solutions are determined by the discrepancy principle and live in different solution subspaces.

Figure 2 displays the singular values $\sigma_{\ell+1}$ of A and the quantities $\|(I - Q_\ell Q_\ell^T)A\|$, $\|(I - Q_\ell Q_\ell^T)A\|\|\mathbf{x}_{\text{exact}}\|$, and $\|(I - Q_\ell Q_\ell^T)\mathbf{b}\|$ of interest for Proposition 1 as functions of ℓ . We can observe in Fig. 2a that, since the singular values of A decay to zero

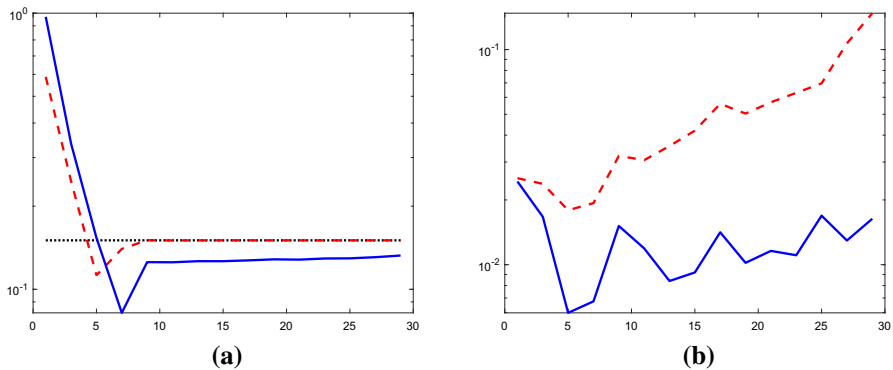


Fig. 1 Shaw test problem, comparison between K-Tikhonov and R-Tikhonov: **a** RRE for solutions determined by K-Tikhonov (red dashed curve), R-Tikhonov (blue solid curve), and standard Tikhonov (black dotted curve). The horizontal axes show the dimension, ℓ , of the solution subspaces for K-Tikhonov and R-Tikhonov. **b** CPU time in seconds for K-Tikhonov (red dashed curve) and R-Tikhonov (blue solid curve) for different values of ℓ (color figure online)

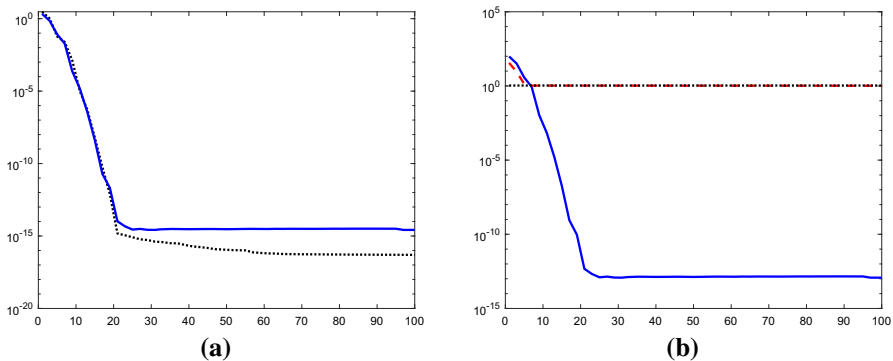


Fig. 2 Shaw test problem, bounds of Proposition 1: **a** Comparison of the approximation error $\|(I - Q_\ell Q_\ell^T)A\|$ (blue solid curve) and the optimal error, i.e., the singular values $\sigma_{\ell+1}$ of the matrix A , (black dotted curve) versus ℓ . **b** Comparison of $\|(I - Q_\ell Q_\ell^T)A\| \|\mathbf{x}_{\text{exact}}\|$ (blue solid curve), $\|(I - Q_\ell Q_\ell^T)\mathbf{b}\|$ (red dashed curve), and $\eta\epsilon$ (black dotted line) versus ℓ (color figure online)

extremely fast with increasing index ℓ , the matrix $Q_\ell Q_\ell^T A$ approximates A well; the approximation error is very close to the optimal one, i.e., to $\sigma_{\ell+1}$. Figure 2b shows that the discrepancy principle already can be satisfied in a subspace of fairly small dimension. We remark that, although the function $\ell \rightarrow \|(I - Q_\ell Q_\ell^T)\mathbf{b}\|$ appears to be constant for ℓ large enough, this function is decreasing very slowly as ℓ increases and vanishes for $\ell = n$.

Heat. We consider the heat test problem in [20]. It is described in [4]. This problem models inverse heat conduction in one space-dimension. We use MATLAB code supplied in [20]. This code requires a parameter κ , which is set to the default value 1. Discretization gives a problem (1) with a matrix $A \in \mathbb{R}^{2048 \times 2048}$ and a vector $\mathbf{x}_{\text{exact}}$, from which we compute the exact data vector $\mathbf{b}_{\text{exact}} = A\mathbf{x}_{\text{exact}}$. We add an error vector \mathbf{e} that models white Gaussian noise and corresponds to a noise level of $\delta = 0.02$ to

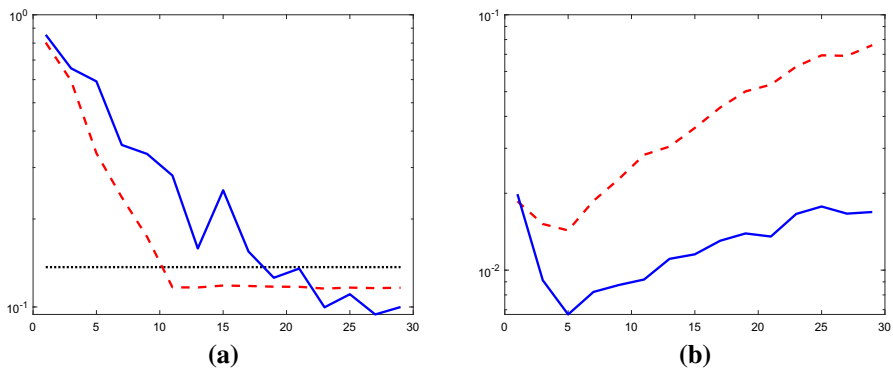


Fig. 3 Heat test problem, comparison between K-Tikhonov and R-Tikhonov: **a** RRE for solutions determined by K-Tikhonov (red dashed curve), R-Tikhonov (blue solid curve), and standard Tikhonov (black dotted curve). The horizontal axes show the dimension, ℓ , of the solution subspaces for K-Tikhonov and R-Tikhonov. **b** CPU time in seconds for K-Tikhonov (red dashed curve) and R-Tikhonov (blue solid curve) for different values of ℓ (color figure online)

$\mathbf{b}_{\text{exact}}$ to obtain the error contaminated vector \mathbf{b} in (1). Similarly as in the previous example, we let L be the discrete Laplacian in one space-dimension, and we display the RRE and CPU times for K-Tikhonov and R-Tikhonov.

Figure 3 shows the RRE and CPU times for several ℓ -values. Similarly as in the Shaw example, the computing times required for the computation of the R-Tikhonov solutions are much smaller than the times required for K-Tikhonov for solution subspaces of the same dimension. For the present example, the RRE-values obtained with R-Tikhonov are slightly larger than those obtained with K-Tikhonov for solution subspaces of the same dimension, at least for $\ell < 23$. However, the RREs are of the same order of magnitude for both methods for solution subspaces of the same dimension. Finally, we observe that, differently from the previous example, the RRE obtained with the K-Tikhonov method is not the same, for ℓ large, as for standard Tikhonov. This is due to the fact that the discrepancy principle does not determine a unique solution; the computed solution depends on the solution subspace used.

Like in the previous example, we report the singular values $\sigma_{\ell+1}$ of A and the quantities $\|(I - Q_\ell Q_\ell^T)A\|$, $\|(I - Q_\ell Q_\ell^T)A\|\|\mathbf{x}_{\text{exact}}\|$, $\|(I - Q_\ell Q_\ell^T)\mathbf{b}\|$ in Fig. 4. In the present example, the singular values of A decrease to zero slower than in the previous example. Although the matrix $Q_\ell Q_\ell^T A$ is still a good approximation of A , the approximation error is visibly larger than the optimal one, given by $\sigma_{\ell+1}$. This is illustrated by Fig. 4b. We observe that in order to satisfy the discrepancy principle, the parameter ℓ has to be larger than in the previous example.

Phillips. Our last example in one space-dimension is the phillips test problem from [20], where MATLAB code is available. This code provides a discretization of a convolution. A background for this problem is given in [28]. We generate the matrix A and the noise-contaminated vector \mathbf{b} in the same manner as in the previous examples. Thus, the noise is white Gaussian and corresponds to the noise level $\delta = 0.05$. The regularization matrix L is the discrete Laplacian in one space-dimension.

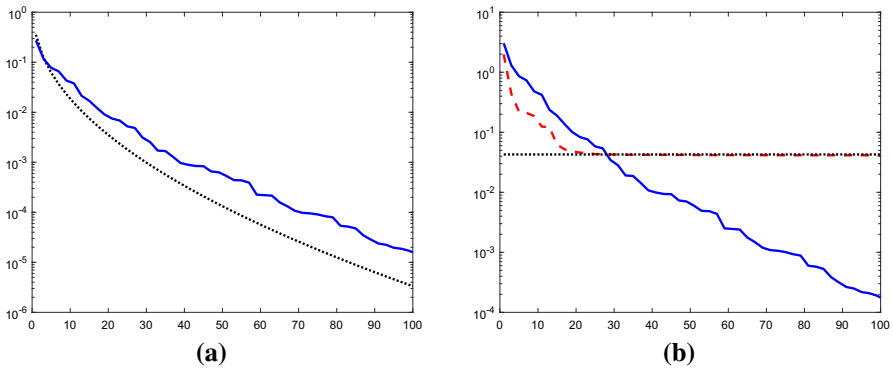


Fig. 4 Heat test problem, bounds of Proposition 1: **a** Comparison of the approximation error $\|(I - Q_\ell Q_\ell^T)A\|$ (blue solid curve) and the optimal error, i.e., the singular values $\sigma_{\ell+1}$ of the matrix A (black dotted curve) as a function of ℓ . **b** Comparison of $\|(I - Q_\ell Q_\ell^T)A\| \|x_{\text{exact}}\|$ (blue solid curve), $\|(I - Q_\ell Q_\ell^T)b\|$ (red dashed curve), and $\eta\epsilon$ (black dotted line) versus ℓ (color figure online)

We report the results obtained with the K-Tikhonov and R-Tikhonov methods in Fig. 5. These results are similar to the ones obtained for the Shaw test problem. Thus, R-Tikhonov outperforms K-Tikhonov in terms of timings and RRE (for ℓ sufficiently large). Similarly to the Shaw test problem, the solutions computed with K-Tikhonov give the same RRE as the solutions computed by standard Tikhonov already for small subspace dimensions ℓ , while the solution computed by R-Tikhonov provide a better approximation of x_{exact} .

As above we display the singular values $\sigma_{\ell+1}$ of A and the norms $\|(I - Q_\ell Q_\ell^T)A\|$, $\|(I - Q_\ell Q_\ell^T)A\| \|x_{\text{exact}}\|$, and $\|(I - Q_\ell Q_\ell^T)b\|$ as functions of ℓ in Fig. 6. This example behaves like the Heat example and we therefore can draw the same conclusions. Nevertheless, let us observe that the decay of the singular values, even though it is slower than in the Shaw example, is still fast enough to yield a fairly accurate approximation of A using the randomized method.

The R-Tikhonov method performs well in all the above examples, even though the matrices A in these examples have different properties. The matrix in the Shaw and Phillips examples are symmetric, while in the Heat example, the matrix is very far from a symmetric matrix. The singular values of the matrix decay to zero quite quickly with increasing index in the Shaw and Heat test problems, while they do not for the problem Phillips.

The above examples are discretizations of problems in one space-dimension. We now turn to problems that are discretizations of ill-posed problems in two space-dimensions. The relative performance of the methods in our comparison will be seen to be different for this kind of problems.

Blur. We determine the matrix A with the MATLAB function `blur(45, 8, 1)` from [20]. This function call generates a symmetric block-Toeplitz-Toeplitz-block (BTTB) matrix $A \in \mathbb{R}^{2025 \times 2025}$, which models a Gaussian point spread function in two space-dimensions. The parameter value 8 is the half-bandwidth of the Toeplitz blocks. Thus, the matrix A is very sparse. It is stored in sparse matrix format. The regularization matrix L is the discrete Laplacian in two space-dimensions.

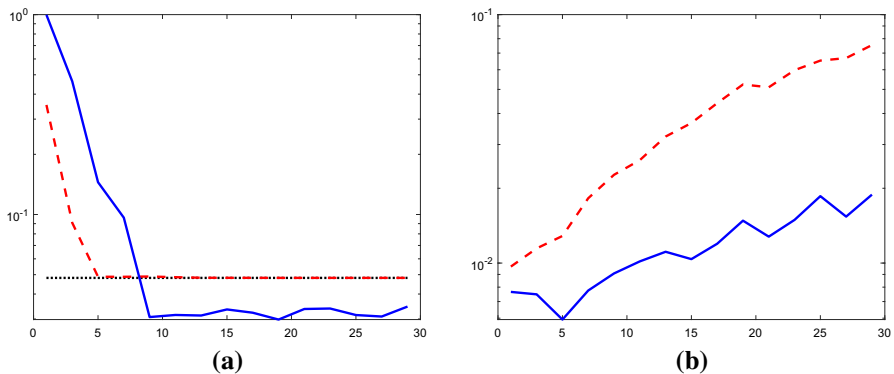


Fig. 5 Phillips test problem, comparison between K-Tikhonov and R-Tikhonov: **a** RRE for solutions determined by K-Tikhonov (red dashed curve), R-Tikhonov (blue solid curve), and standard Tikhonov (black dotted curve). The horizontal axes show the dimension, ℓ , of the solution subspace for K-Tikhonov and R-Tikhonov. **b** CPU time for K-Tikhonov (red dashed curve) and R-Tikhonov (blue solid curve) for different values of ℓ (color figure online)

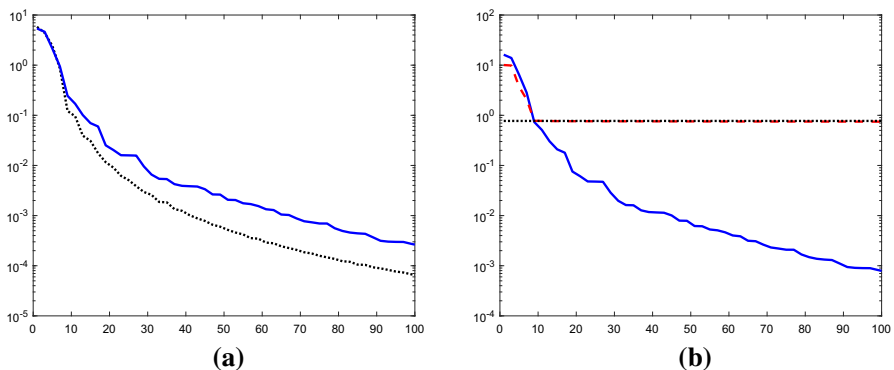


Fig. 6 Phillips test problem, bounds of Proposition 1: **a** Comparison of the approximation error $\|(I - Q_\ell Q_\ell^T)A\|$ (blue solid curve) and the optimal error, i.e., the singular values $\sigma_{\ell+1}$ of the matrix A (black dotted curve), versus ℓ . **b** Comparison of $\|(I - Q_\ell Q_\ell^T)A\| \|\mathbf{x}_{\text{exact}}\|$ (blue solid curve), $\|(I - Q_\ell Q_\ell^T)\mathbf{b}\|$ (red dashed curve), and η_ϵ (black dotted line) versus ℓ (color figure online)

Let the entries of $\mathbf{x}_{\text{exact}}$ be pixel values of an 45×45 -pixel image, with the pixels ordered column-wise. Then $\mathbf{b}_{\text{exact}} = A\mathbf{x}_{\text{exact}}$ represents a blurred image associated with $\mathbf{x}_{\text{exact}}$. Add a vector \mathbf{e} that represents white Gaussian noise with noise level $\delta = 0.03$ to $\mathbf{b}_{\text{exact}}$ to obtain the contaminated data vector \mathbf{b} ; cf. (2).

We compare RREs and CPU times for different values of ℓ similarly as in the previous examples. The results are reported in Fig. 7. The figure shows that the R-Tikhonov method does not perform well in terms of the RRE. In fact, the RRE is very large and does not decrease significantly as the dimension of the solution subspace, ℓ , increases. Moreover, since the matrix A is very sparse, the matrix–vector products required by the K-Tikhonov method are not computational demanding. Therefore, the computational cost of the two methods is about the same.

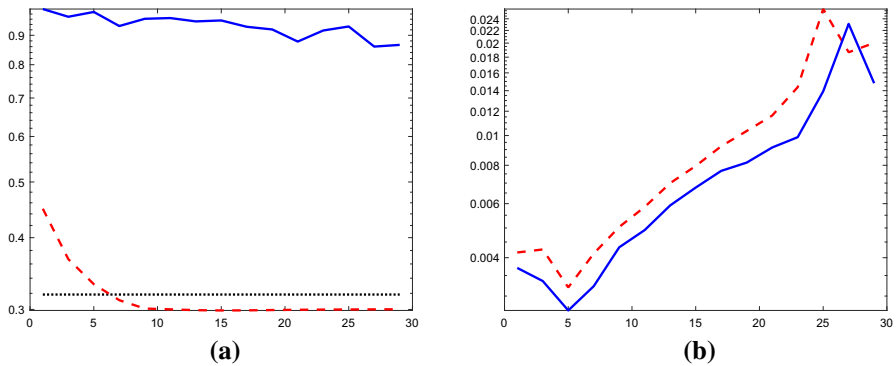


Fig. 7 Blur test problem, comparison between K-Tikhonov and R-Tikhonov: **a** RRE for solutions determined by K-Tikhonov (red dashed curve), R-Tikhonov (blue solid curve), and standard Tikhonov (black dotted curve). The horizontal axes show the dimension, ℓ , of the solution subspaces for K-Tikhonov and R-Tikhonov. **b** CPU time in seconds for K-Tikhonov (red dashed curve) and R-Tikhonov (blue solid curve) for different values of ℓ (color figure online)

Table 1 Blur test problem: RRE and CPU time in seconds for standard Tikhonov, K-Tikhonov, and R-Tikhonov for selected values of k and ℓ

	standard Tikhonov	K-Tikhonov ($\ell = 30$)	R-Tikhonov ($\ell = 1000$)
RRE	$3.18 \cdot 10^{-1}$	$3.00 \cdot 10^{-1}$	$3.16 \cdot 10^{-1}$
CPU time (sec.)	–	$2.00 \cdot 10^{-2}$	$1.21 \cdot 10^0$

The relatively poor performance of the R-Tikhonov method in this example is due to the fact that the singular values of the matrix A decrease fairly slowly to zero with increasing index and are not approximated well by the singular values of the reduced matrix $AQ_{30}Q_{30}^T$ used to compute the R-Tikhonov solution. To see this in more detail, we plot in Fig. 8 the singular values of the matrices $AQ_{30}Q_{30}^T$, \tilde{C}_{30} , and A for all the considered examples. We can observe that in all examples in one space-dimension, the largest singular values of the matrix A are approximated well by the singular values of $AQ_{30}Q_{30}^T$ and \tilde{C}_{30} . However, the singular values of $AQ_{30}Q_{30}^T$ for the present example are very poor approximations of the largest singular values of A . On the other hand, the singular values of \tilde{C}_{30} match very well the largest singular values of A . This suggests that the R-Tikhonov method may not be effective for the solution of linear discrete ill-posed problems with a matrix A whose singular values decay fairly slowly with their index number. To illustrate this, we choose the solution subspace for the R-Tikhonov method to be $\ell = 1000$, and compare the error in the computed solution with the errors in the solution determined by the K-Tikhonov method with $\ell = 30$ and by standard Tikhonov. This comparison is reported in Table 1. We can see that even when $\ell = 1000$, the R-Tikhonov method provides less accurate results than K-Tikhonov with $\ell = 30$, and requires much more execution time (about 45 times as much).

These observations are corroborated by Fig. 9, which shows the singular values $\sigma_{\ell+1}$ of A and compares them to $\|(I - Q_{\ell}Q_{\ell}^T)A\|$ as functions of ℓ . The figure also

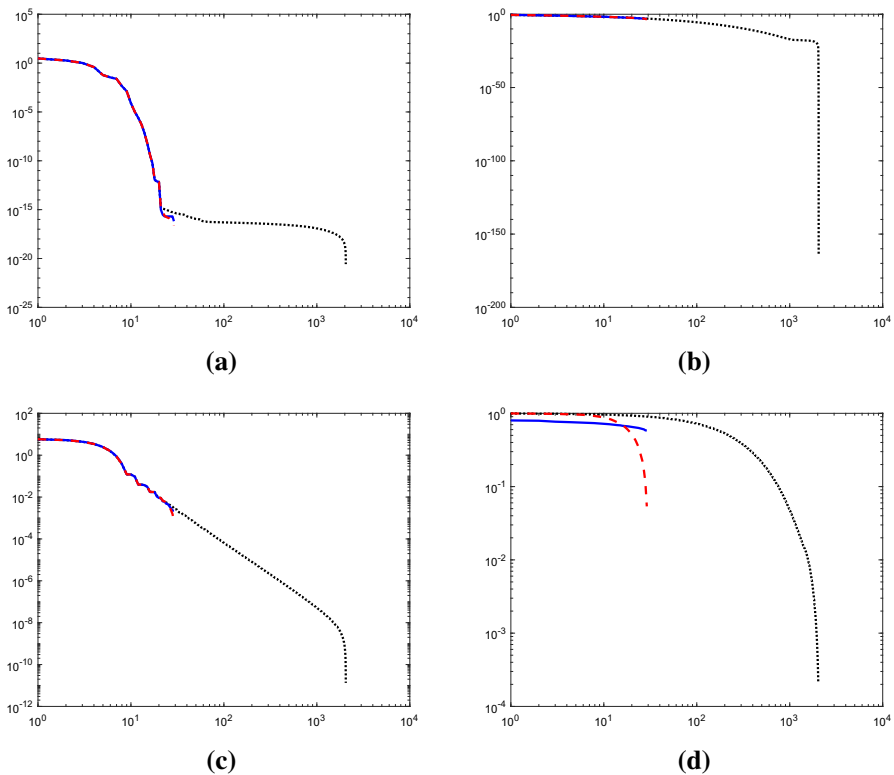


Fig. 8 The largest singular values of the matrices $AQ_{30}Q_{30}^T$ (blue solid curve), \bar{C}_{30} (red dashed curve), and A (black dotted curve) as a function of the index for all the considered examples: **a** Shaw, **b** Heat, **c** Phillips, **d** Blur (color figure online)

displays the norms $\|(I - Q_\ell Q_\ell^T)A\| \|\mathbf{x}_{\text{exact}}\|$ and $\|(I - Q_\ell Q_\ell^T)\mathbf{b}\|$. We can observe in Fig. 9a that since the singular values of A do not decay fast enough to zero with increasing index ℓ , the approximation error $\|(I - Q_\ell Q_\ell^T)A\|$ is extremely large even for $\ell = 100$. Moreover, we can see that the discrepancy principle cannot be satisfied for $\ell \leq 100$. By visual inspection of Fig. 9b, we can deduce that a very large value of ℓ may be required to satisfy the discrepancy principle.

Hubble. We turn to a deblurring problem from [13]. Specifically, we consider the deblurring problem obtained when the available image is blurred by a medium speckle PSF and, in addition, is contaminated by 5% white Gaussian noise. The size of the image is 512×512 pixels; see Fig. 10. We impose periodic boundary conditions. Then the blurring matrix $A \in \mathbb{R}^{n \times n}$, with $n = 512^2$, is block circulant with circulant blocks (BCCB). Thus, A can be diagonalized by the bidimensional Fourier matrix. We can compute the eigenvalues of A in $\mathcal{O}(n \log n)$ flops with the aid of the fast Fourier transform (FFT) algorithm; see, e.g., [21] for a discussion on image deblurring and boundary conditions. By choosing L as the discretization of the bidimensional Laplacian with periodic boundary conditions, we can solve (3) in $\mathcal{O}(n \log n)$ flops for each value of the regularization parameter μ . This allows us to compute the solution

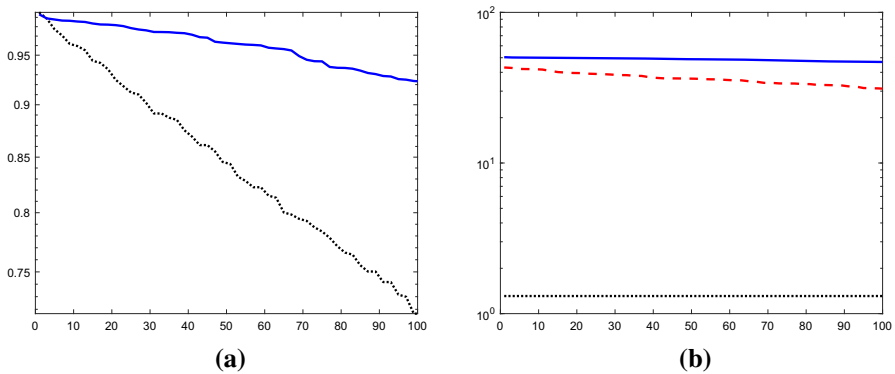


Fig. 9 Blur test problem, bounds of Proposition 1: **a** Comparison of the approximation error $\|(I - Q_\ell Q_\ell^T)A\|$ (blue solid curve) and the optimal error, i.e., the singular values σ_ℓ of the matrix A (black dotted curve), versus ℓ . **b** Comparison of $\|(I - Q_\ell Q_\ell^T)A\| \|\mathbf{x}_{\text{exact}}\|$ (blue solid curve), $\|(I - Q_\ell Q_\ell^T)\mathbf{b}\|$ (red dashed curve), and $\eta\epsilon$ (black dotted line) against ℓ (color figure online)

by standard Tikhonov regularization and to compare results obtained in this manner with those determined by the K-Tikhonov and R-Tikhonov methods.

Similarly as above, we apply both the R-Tikhonov and K-Tikhonov methods for different values of ℓ , and compare results in terms of CPU time and accuracy. The matrix A is not explicitly formed; instead we evaluate matrix–vector products with A and A^T by using the FFT algorithm. Hence, in R-Tikhonov the matrices $A\Omega_\ell$ and $A^T Q_\ell$ are computed by evaluating ℓ matrix–vector products with A and ℓ matrix–vector products with A^T . We therefore expect the computing time for R-Tikhonov and K-Tikhonov to be about the same. This is confirmed by the graphs of Fig. 11b. On the other hand, we can see from Fig. 11a and d that the R-Tikhonov method fails to accurately determine the largest singular values of A , and that the restored image determined by R-Tikhonov is of very poor quality; see Fig. 12b. This is due to the fact that, as we can see in Fig. 11c, the singular values of A do not decrease very fast to zero.

Figure 12 displays the reconstructed images obtained with standard Tikhonov, R-Tikhonov, and K-Tikhonov. Visual inspection shows that K-Tikhonov is able to provide a reconstruction of similar quality as standard Tikhonov, while R-Tikhonov fails to determine an accurate approximation of $\mathbf{x}_{\text{exact}}$.

5 Conclusion

The application of randomized algorithms to the solution of large-scale problems has received considerable attention. This paper compares their performance with a Krylov subspace method when applied to the solution of linear discrete ill-posed problems by Tikhonov regularization. The singular values of linear discrete ill-posed problems “cluster” at the origin, however, their rate of decay towards zero with increasing index is problem dependent. The randomized method is found to be competitive for the solution of linear discrete ill-posed problems in one space-dimension, for which

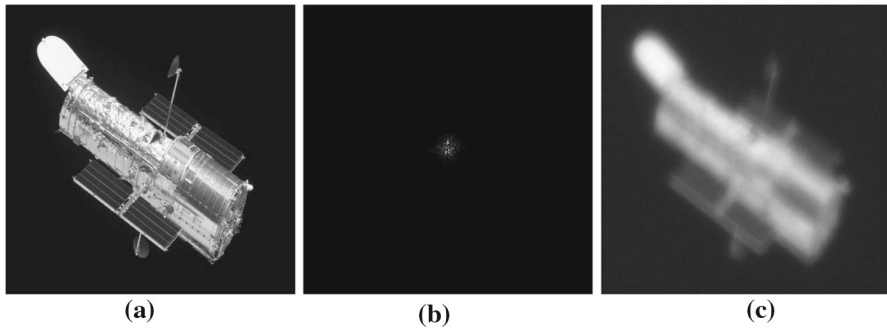


Fig. 10 Hubble test case: **a** true image (512×512 pixels), **b** PSF (512×512 pixels), **c** blurred and noisy image with 5% of white Gaussian noise (512×512 pixels)

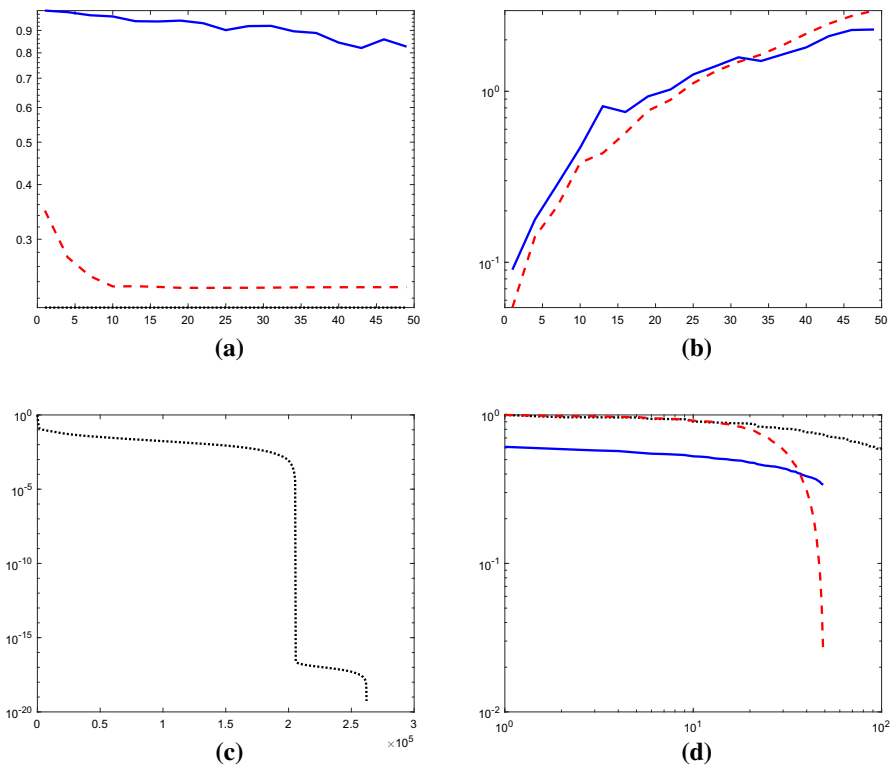


Fig. 11 Hubble test problem: Comparison between K-Tikhonov and R-Tikhonov: **a** RRE for solutions determined by K-Tikhonov (red dashed curve), R-Tikhonov (blue solid curve), and standard Tikhonov (black dotted curve). The horizontal axis shows ℓ for K-Tikhonov and k for R-Tikhonov, **b** CPU times for K-Tikhonov (red dashed curve) and R-Tikhonov (blue solid curve) for different values of ℓ and k , **c** singular values of the blurring matrix A , **d** the largest singular values of the matrices $A Q_{50} Q_{50}^T$ (blue solid curve), \bar{C}_{50} (red dashed curve), and A (black dotted curve) as a function of their index (color figure online)

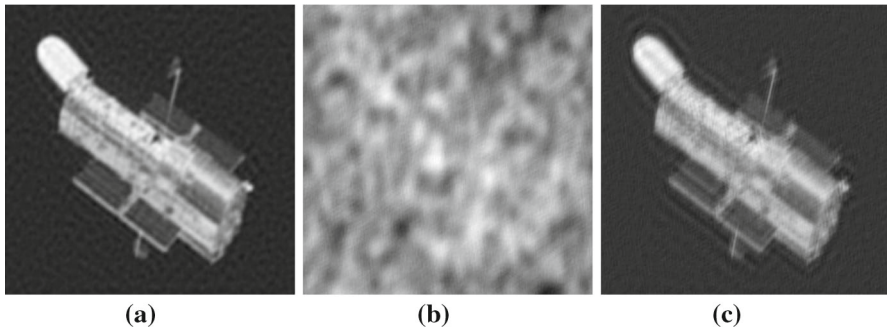


Fig. 12 Hubble test case reconstructions: **a** Tikhonov, **b** R-Tikhonov ($k = 50$), **c** K-Tikhonov ($\ell = 50$)

the singular values decay to zero fast enough with increasing index. However, when the singular values do not decrease quickly enough, the Krylov method considered outperforms the randomized method. This depends on that Krylov methods determine more appropriate solution subspaces of low dimensions for linear discrete ill-posed problems than the randomized method when the singular values do not decay to zero sufficiently rapidly.

We only consider one Krylov subspace method, Golub–Kahan bidiagonalization, in this paper. However, our conclusions carry over to other Krylov subspace solution methods, such as the Arnoldi method, as well, at least when the matrix A is not too far from symmetric. When A is far from symmetric, solution methods for discrete ill-posed problems based on the Arnoldi process are known not to provide satisfactory results; see, e.g., [6, 14] an illustration.

Acknowledgements The authors would like to thank Michiel Hochstenbach and a referee for comments. A.B. is a member of the GNCS-INdAM group that partially supported this work with the Young Researchers Project (Progetto Giovani Ricercatori) “Variational methods for the approximation of sparse data”. Moreover, A.B. research is partially supported by the Regione Autonoma della Sardegna research project “Algorithms and Models for Imaging Science [AMIS]” (RASSR57257, intervento finanziato con risorse FSC 2014–2020—Patto per lo Sviluppo della Regione Sardegna). The work of L.R is supported in part by NSF Grants DMS-1720259 and DMS-1729509.

Funding Open access funding provided by Università degli Studi di Cagliari within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Calvetti, D., Golub, G.H., Reichel, L.: Estimation of the L-curve via Lanczos bidiagonalization. *BIT Numer. Math.* **39**, 603–619 (1999)
2. Calvetti, D., Reichel, L.: Tikhonov regularization of large linear problems. *BIT Numer. Math.* **43**, 263–283 (2003)
3. Calvetti, D., Reichel, L., Shuibi, A.: Invertible smoothing preconditioners for linear discrete ill-posed problems. *Appl. Numer. Math.* **54**, 135–149 (2005)
4. Carasso, A.S.: Determining surface temperatures from interior observations. *SIAM J. Appl. Math.* **42**, 558–574 (1982)
5. Donatelli, M., Reichel, L.: Square smoothing regularization matrices with accurate boundary conditions. *J. Comput. Appl. Math.* **272**, 334–349 (2014)
6. Donatelli, M., Martin, D., Reichel, L.: Arnoldi methods for image deblurring with anti-reflective boundary conditions. *Appl. Math. Comput.* **253**, 135–150 (2015)
7. Dykes, L., Huang, G., Noschese, S., Reichel, L.: Regularization matrices for discrete ill-posed problems in several space-dimensions. *Numer. Linear Algebra Appl.* **25**, e2163 (2018)
8. Eckart, C., Young, G.: The approximation of one matrix by another of lower rank. *Psychometrika* **1**, 211–218 (1936)
9. Eldén, L.: A weighted pseudoinverse, generalized singular values, and constrained least squares problems. *BIT Numer. Math.* **22**, 487–502 (1982)
10. Engl, H.W., Hanke, M., Neubauer, A.: *Regularization of Inverse Problems*. Kluwer, Dordrecht (1996)
11. Fenu, C., Reichel, L., Rodriguez, G., Sadok, H.: GCV for Tikhonov regularization by partial SVD. *BIT Numer. Math.* **57**, 1019–1039 (2017)
12. Gallivan, K., Heath, M., Ng, E., Peyton, B., Plemmons, R., Ortega, J., Romine, C., Sameh, A., Voigt, R.: *Parallel Algorithms for Matrix Computations*. SIAM, Philadelphia (1990)
13. Gazzola, S., Hansen, P.C., Nagy, J.G., Tools, I.R.: A MATLAB package of iterative regularization methods and large-scale test problems. *Numer. Algorithms* **81**, 773–811 (2019)
14. Gazzola, S., Noschese, S., Novati, P., Reichel, L.: Arnoldi decomposition, GMRES, and preconditioning for linear discrete ill-posed problems. *Appl. Numer. Math.* **142**, 102–121 (2019)
15. Gazzola, S., Novati, P., Russo, M.R.: On Krylov projection methods and Tikhonov regularization. *Electron. Trans. Numer. Anal.* **44**, 83–123 (2015)
16. Golub, G.H., Van Loan, C.F.: *Matrix Computations*, 4th edn. Johns Hopkins University Press, Baltimore (2013)
17. Halko, N., Martinsson, P.G., Tropp, J.A.: Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.* **53**, 217–288 (2011)
18. Hanke, M., Hansen, P.C.: Regularization methods for large-scale problems. *Surv. Math. Ind.* **3**, 253–315 (1993)
19. Hansen, P.C.: *Rank-Deficient and Discrete Ill-Posed Problems*. SIAM, Philadelphia (1998)
20. Hansen, P.C.: Regularization Tools version 4.0 for MATLAB 7.3. *Numer. Algorithms* **46**, 189–194 (2007)
21. Hansen, P.C., Nagy, J.G., O’leary, D.P.: *Deblurring Images: Matrices, Spectra, and Filtering*. SIAM, Philadelphia (2006)
22. Hochstenbach, M.E., Reichel, L.: An iterative method for Tikhonov regularization with a general linear regularization operator. *J. Integral Equ. Appl.* **22**, 463–480 (2010)
23. Huang, G., Noschese, S., Reichel, L.: Regularization matrices via matrix nearness problems. *Linear Algebra Appl.* **502**, 366–386 (2016)
24. Huang, G., Reichel, L., Yin, F.: On the choice of solution subspace for nonstationary iterated Tikhonov regularization. *Numer. Algorithms* **72**, 1043–1063 (2016)
25. Kindermann, S.: Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems. *Electron. Trans. Numer. Anal.* **38**, 233–257 (2011)
26. Kindermann, S., Raik, K.: A simplified L-curve method as error estimator. *Electron. Trans. Numer. Anal.* **53**, 217–238 (2020)
27. Lewis, B., Reichel, L.: Arnoldi–Tikhonov regularization methods. *J. Comput. Appl. Math.* **226**, 92–102 (2009)
28. Phillips, D.L.: A technique for the numerical solution of certain integral equations of the first kind. *J. ACM* **9**, 84–97 (1962)

29. Reichel, L., Rodriguez, G.: Old and new parameter choice rules for discrete ill-posed problems. *Numer. Algorithms* **63**, 65–87 (2013)
30. Reichel, L., Shyshkov, A.: A new zero-finder for Tikhonov regularization. *BIT Numer. Math.* **48**, 627–643 (2008)
31. Shaw, C.B., Jr.: Improvements of the resolution of an instrument by numerical solution of an integral equation. *J. Math. Anal. Appl.* **37**, 83–112 (1972)
32. Vatankhah, S., Renaut, R.A., Ardesta, V.E.: A fast algorithm for regularized focused 3D inversion of gravity data using randomized singular-value decomposition. *Geophysics* (2018). <https://doi.org/10.1190/geo2017-0386.1>
33. Xiang, H., Zou, J.: Regularization with randomized SVD for large-scale discrete inverse problems. *Inverse Probl.* **29**, 085008 (2013)
34. Xiang, H., Zou, J.: Randomized algorithms for large-scale inverse problems with general Tikhonov regularization. *Inverse Probl.* **31**, 085008 (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.