Solution of linear discrete ill-posed problems by discretized Chebyshev expansion

Xianglan Bai
Department of Mathematical Sciences
Kent State University
Kent, OH 44242
xbai@kent.edu

Alessandro Buccini
Department of Mathematics and Computer Science
University of Cagliari
Cagliari, Italy
alessandro.buccini@unica.it

Lothar Reichel
Department of Mathematical Sciences
Kent State University
Kent, OH 44242
reichel@math.kent.edu

***Abstract*—Large-scale linear discrete ill-posed problems are generally solved by Krylov subspace iterative methods. However, these methods can be difficult to implement so that they execute efficiently in a multiprocessor environment, because some of the computations have to be carried out sequentially. This is due to the fact that only one new basis vector of the Krylov solution subspace is generated in each iteration. It is therefore interesting to investigate the performance of other solution methods that use a solution subspace basis that can be generated in parallel and, therefore, more efficiently on many computers. This paper proposes solution methods that use a solution subspace basis that is made up of discretized Chebyshev polynomials. It compares their performance to a Krylov subspace method that is based on partial Golub–Kahan bidiagonalization of the system matrix, and to a randomized method. The application of a solution subspace basis made up of discretized Chebyshev polynomial is found to be competitive when solving linear discrete ill-posed problems in one space-dimension and for some problems in higher space-dimensions.**

***Index Terms*—Linear discrete ill-posed problems, Chebyshev expansion method, Tikhonov regularization.**

I. Introduction

We consider the solution of least squares problems

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|A\boldsymbol{x}-\boldsymbol{b}\|_2, \quad A\in\mathbb{R}^{m\times n}, \quad \boldsymbol{x}\in\mathbb{R}^n, \quad \boldsymbol{b}\in\mathbb{R}^m, \quad (1)$$

that stem from the discretization of linear ill-posed problems, such as a Fredholm integral equation of the first kind, say

$$\int_{-1}^{1} \kappa(s,t)x(t)dt = b(s), \qquad -1\le s\le 1, \qquad (2)$$

with a smooth kernel κ. Then the singular values of the matrix A decay fairly rapidly to zero without a significant gap with increasing index number. In particular, the matrix A is severely ill-conditioned and may be rank deficient. The matrix A is allowed to be rectangular, i.e., we consider both the cases $m=n$ and $m\neq n$. The vector $\boldsymbol{b}$ in (1), or equivalently the right-hand side function $b(s)$ in (2), represent measured data that are assumed to be contaminated by a measurement error $\boldsymbol{e}\in\mathbb{R}^m$ in (1) or $e(s)$ in (2).

Most of this paper focuses on the solution of least squares problems (1), which may stem from the discretization of ill-posed problems that are more general than (2), but our derivation of the solution methods is inspired by properties of integral equations of the form (2).

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

$$\boldsymbol{b} = \boldsymbol{b}_{\rm exact} + \boldsymbol{e}. \qquad (3)$$

Assume that the linear system of equations

$$A\boldsymbol{x} = \boldsymbol{b}_{\rm exact} \qquad (4)$$

is consistent. We are interested in determining an accurate approximation of the solution $\boldsymbol{x}_{\rm exact}$ of minimal Euclidean norm of (4) by computing a suitable approximate solution of (1).

Straightforward solution of (1) by using some factorization of A generally does not yield a useful approximation of $\boldsymbol{x}_{\rm exact}$, because the ill-conditioning of A typically causes severe propagation and amplification of the error $\boldsymbol{e}$ in $\boldsymbol{b}$ into the computed solution of (1). A common technique to remedy this difficulty is to replace the minimization problem (1) by a nearby problem, whose solution is less sensitive to the error $\boldsymbol{e}$ in $\boldsymbol{b}$ than the solution of (1). Such a replacement is commonly referred to as regularization. One of the most popular replacements is known as Tikhonov regularization, which replaces the minimization problem (1) by a penalized least squares problem of the form

$$\min_{\boldsymbol{x}\in\mathbb{R}} \left\{ \|A\boldsymbol{x}-\boldsymbol{b}\|_2^2 + \mu\|L\boldsymbol{x}\|_2^2 \right\}, \qquad (5)$$

where $L\in\mathbb{R}^{p\times n}$ is a regularization matrix and $\mu>0$ is a regularization parameter. Throughout this paper $\|\cdot\|_2$ denotes

A.B. is a member of the GNCS-INdAM group and his 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). Research by L.R. is supported in part by NSF grant DMS-1720259.

the Euclidean vector norm or the spectral matrix norm. We will assume that the matrix L is chosen so that

$$\mathcal{N}(A) \cap \mathcal{N}(L) = \{\mathbf{0}\},$$

where $\mathcal{N}(M)$ stands for the null space of the matrix M. Then the Tikhonov minimization problem (5) has a unique solution $\boldsymbol{x}_\mu$ for any $\mu > 0$.

The size of the regularization parameter μ determines how sensitive the solution $\boldsymbol{x}_\mu$ of (5) is to the error $\boldsymbol{e}$ in $\boldsymbol{b}$, and how close $\boldsymbol{x}_\mu$ is to the desired solution $\boldsymbol{x}_{\rm exact}$. We will determine μ by the discrepancy principle, i.e., we choose $\mu > 0$ so that

$$\|\boldsymbol{b} - A\boldsymbol{x}_\mu\|_2 = \eta\epsilon, \tag{6}$$

where ϵ is an upper bound for $\|\boldsymbol{e}\|_2$ and $\eta > 1$ is a user-specified constant that is independent of ϵ. Such a choice of μ is possible for most reasonable error bounds and determines a unique value of μ. It can be shown that $\boldsymbol{x}_\mu \to \boldsymbol{x}_{\rm exact}$ as $\epsilon \searrow 0$; see, e.g., Engl et al. [11] for a proof in Hilbert space setting. We remark that application of the discrepancy principle requires both knowledge of a bound ϵ and consistency of the system (4). There are many so-called "heuristic" techniques for determining a suitable value of the regularization parameter that can be applied in situations when the discrepancy principle cannot; see, e.g., [12], [19]–[22]. These techniques also can be used in conjunction with the solution methods for (5) discussed in the present paper.

Tikhonov minimization problems (5) of small to moderate size can be conveniently solved by first computing the generalized singular value decomposition (GSVD) of the matrix pair $\{A, L\}$ and then determining both a suitable value of the regularization parameter μ and the associated solution $\boldsymbol{x}_\mu$; see, e.g., [5], [8], [17]. We remark that when $p > n$, it may be convenient to replace L by the upper triangular $n \times n$ matrix of its "economical" QR factorization. However, when the matrices A and L are large, the computation of the GSVD of $\{A, L\}$ may be prohibitively expensive. Large-scale Tikhonov minimization problems therefore often are solved by Krylov subspace methods; see, e.g., [4], [6], [13], [17].

Krylov subspace methods generally perform very well. However, they are poorly suited for execution on computers with many processors, because the basis vectors of the solution subspace are generated sequentially one-by-one. This has spurred interest in randomized methods that allow more efficient execution in a multiprocessor environment. Randomized methods for the solution of Tikhonov minimization problems are described in, e.g., [2], [3], [26], [27]. A nice introduction to these methods is presented by Halko, Martinsson, and Tropp [15].

It is the purpose of the present paper to explore the use of discretized Chebyshev polynomials to generate orthonormal solution subspace bases. Solution methods for least squares problems (1) when the matrix A and data vector $\boldsymbol{b}$ are given, as well as solution methods for integral equations (2) when the kernel κ and right-hand side function b are given and have to be discretized, are discussed. The use of bases of discretized Chebyshev polynomials is found to be competitive with randomized methods both with regard to computing time and quality of the computed solution. Both the use of an orthonormal basis of discretized Chebyshev polynomials and randomized methods turn out to be competitive with a commonly used Krylov subspace method, based on partial Golub–Kahan bidiagonalization of the matrix A, for solving (5) when the underlying ill-posed problem is in one space-dimension. These approaches also are competitive for certain ill-posed problems in higher space dimensions.

This paper is organized as follows. Section II describes our methods based on using solution subspaces defined by discretized Chebyshev polynomials. The randomized methods by Xiang and Zou [26], [27] are discussed in Sections III. A Krylov subspace method based on partial Golub–Kahan bidiagonalization is outlined in Section IV, and numerical examples that compare the performance of the methods discussed are presented in Section V. Concluding remarks can be found in Section VI.

II. Solution methods with a solution subspace of discretized Chebyshev polynomials

The solution methods described in the present section and the randomized solution method reviewed in Section III seek to determine a subspace $\mathcal{S}_\ell \subset \mathbb{R}^n$ of fairly small dimension $1 \le \ell \ll \min\{m, n\}$ that allows quite accurate approximation of the matrix A in (1) by a low-rank matrix. We illustrate this with the singular value decomposition (SVD) even though the computation of the SVD typically is too expensive to be attractive for large matrices A. Assume for notational simplicity that $m \ge n$. Then the SVD of A is given by

$$A = W\Sigma V^T, \tag{7}$$

where $W = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m] \in \mathbb{R}^{m\times m}$ and $V = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n] \in \mathbb{R}^{n\times n}$ are orthogonal matrices, and

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

The diagonal entries σ_j are known as the singular values of A. They are nonnegative and ordered according to

$$\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_n.$$

Let $\mathcal{S}_\ell = {\rm span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_\ell\}$ and let $P_{\mathcal{S}_\ell}$ denote the orthogonal projector onto $\mathcal{S}_\ell$. Then

$$\sigma_{\ell+1} = \|A - P_{\mathcal{S}_\ell}A\|_2. \tag{8}$$

Thus, the low-rank matrix $P_{\mathcal{S}_\ell}A$ furnishes an accurate approximation of A if $\sigma_{\ell+1}$ is small. It follows from the Eckart–Young Theorem [9] that the right-hand side in (8) is as small as possible in the sense that if $P_{\mathcal{S}_\ell}$ is an orthogonal projector onto any ℓ-dimensional subspace of $\mathbb{R}^m$, then

$$\sigma_{\ell+1} \le \|A - P_{\mathcal{S}_\ell}A\|_2. \tag{9}$$

A subspace $\mathcal{S}_\ell$ is well suited for the representation of an approximate solution of (5) if it can be computed fairly inexpensively, and the right-hand side of (9) is small. We will

explore the possibility of defining subspaces $\mathcal{S}_\ell$ by discretizing Chebyshev polynomials of the first kind,

$$T_j(t) = \cos(j \arccos(t)), \quad -1 \le t \le 1, \quad j = 0, 1, \ldots, \ell-1,$$

for some $\ell \le m$, at the zeros

$$t_i = \cos\left(\frac{2i-1}{2m}\pi\right), \quad i = 1, 2, \ldots, m, \tag{10}$$

of the Chebyshev polynomial T_m of degree m. For future reference, we define

$$\theta_i = \frac{2i-1}{2m}\pi, \quad i = 1, 2, \ldots, m. \tag{11}$$

Introduce the matrix $U_\ell = [u_{i,j}] \in \mathbb{R}^{m\times\ell}$ by

$$\begin{aligned} u_{i,1} &= \frac{1}{\sqrt{m}}, \quad 1 \le i \le m, \\ u_{i,j} &= \sqrt{\frac{2}{m}}\cos((j-1)\theta_i), \quad 1 \le i \le m,\ 2 \le j \le \ell. \end{aligned} \tag{12}$$

Then U_ℓ has orthonormal columns denoted by $u_1, u_2, \ldots, u_\ell$. We let $\mathcal{S}_\ell = \mathrm{range}(U_\ell)$ and define the orthogonal projector $P_{\mathcal{S}_\ell} = U_\ell U_\ell^T$.

The matrix U_ℓ can be applied in several ways to determine an approximate solution of the Tikhonov minimization problem (5). These different approaches lead to three different numerical methods described below. The third method uses an approach that is analogous to the one used in the randomized method of Section III, but instead of using a Gaussian matrix to determine the matrix U_ℓ, we use (12).

For problems with a Kronecker product structure it can be attractive to use outer products of the columns u_j of U_ℓ to construct a basis. This is illustrated in Section V.

Method II.1. Replace the matrix A in the Tikhonov minimization problem (5) by the matrix $U_\ell U_\ell^T A$ of rank at most $\ell \ll \min\{m, n\}$. Then we obtain the new minimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ \|U_\ell U_\ell^T A\boldsymbol{x} - U_\ell U_\ell^T \boldsymbol{b}\|_2^2 + \mu\|L\boldsymbol{x}\|_2^2 \right\} + \|(I - U_\ell U_\ell^T)\boldsymbol{b}\|_2^2.$$

Requiring an approximate solution of the form $\boldsymbol{x} = U_\ell \boldsymbol{y}$ for some $\boldsymbol{y} \in \mathbb{R}^\ell$, and defining the QR factorization

$$Q_\ell R_\ell = LU_\ell,$$

where the matrix $Q_\ell \in \mathbb{R}^{m\times\ell}$ has orthonormal columns and the matrix $R_\ell \in \mathbb{R}^{\ell\times\ell}$ is upper triangular, yield the low-dimensional minimization problem

$$\min_{\boldsymbol{y}\in\mathbb{R}^\ell} \left\{ \|U_\ell^T A U_\ell \boldsymbol{y} - U_\ell^T \boldsymbol{b}\|_2^2 + \mu\|R_\ell \boldsymbol{y}\|_2^2 \right\}. \tag{13}$$

Theorem 1 below provides a bound for the distance between A and $U_\ell U_\ell^T A$ when A is a suitable discretization of the kernel κ in (2). The result is not concerned with a particular quadrature rule, but just discusses how smoothness of the kernel translates into decay of the magnitude of coefficients of the discrete cosine transform. Method II.2 illustrates how Theorem 1 can be applied when the matrix A stems from a quadrature rule.

Theorem 1. *Assume that the function $\theta \to \kappa(\cos(\theta), \cos(\tau))$ and its partial derivatives through order $\nu - 1$ are absolutely continuous on $[0, \pi]$, and suppose the νth partial derivative is of bounded variation γ, uniformly for $0 \le \tau \le \pi$. Define the matrix $A \in \mathbb{R}^{m\times m}$ by discretizing the kernel of* (2) *according to*

$$A = [a_{i,j}]_{i,j=1}^m, \quad a_{i,j} = \frac{\pi}{m}\kappa(t_i, t_j), \tag{14}$$

where the nodes t_i are given by (10)*. This defines the matrix of the least-squares problem* (1)*. Let the matrix $U_\ell \in \mathbb{R}^{m\times\ell}$ be given by* (12)*. Then for m sufficiently much larger than ℓ, it holds*

$$\|A - U_\ell U_\ell^T A\|_2 \lessapprox \frac{2\gamma}{(\ell-\nu)^{\nu+1}} \tag{15}$$

for $\ell > \nu$.

We note that the discretization of the theorem is associated with the discretization

$$\boldsymbol{b}_{\mathrm{exact}} = [b_i]_{i=1}^m, \quad b_i = b(t_i),$$

of the right-hand side function $b(s)$ in the integral equation (2). The discretization of this function is not used in the theorem.

Proof. Due to the smoothness properties of the kernel, the coefficients

$$\begin{aligned} a_0(\tau) &= \frac{1}{\pi}\int_0^\pi \kappa(\cos(\theta), \cos(\tau))d\theta, \\ a_j(\tau) &= \frac{2}{\pi}\int_0^\pi \kappa(\cos(\theta), \cos(\tau))\cos(j\theta)d\theta,\ j = 1, 2, \ldots, \end{aligned}$$

satisfy

$$|a_j(\tau)| \le \frac{2\gamma}{\pi(j-\nu)^{\nu+1}} \quad \text{for} \quad 0 \le \tau \le \pi, \quad j \ge \nu + 1; \tag{16}$$

see, e.g., [25, Theorem 7.1], where this bound is shown for the coefficients of a Chebyshev expansion of a single function that satisfies the smoothness conditions of the theorem.

Let the matrix $\widetilde{U}_{m-\ell} = [\widetilde{u}_{i,j}] \in \mathbb{R}^{m\times(m-\ell)}$ be defined by

$$\widetilde{u}_{i,j} = \sqrt{\frac{2}{m}}\cos((\ell + j - 1)\theta_i),\ 1 \le i \le m,\ 1 \le j \le m - \ell.$$

Then the concatenated matrix $[U_\ell, \widetilde{U}_{m-\ell}] \in \mathbb{R}^{m\times m}$ is orthogonal and $I - U_\ell U_\ell^T = \widetilde{U}_{m-\ell}\widetilde{U}_{m-\ell}^T$. This yields the bound

$$\|A - U_\ell U_\ell^T A\|_2 \le \|\widetilde{U}_{m-\ell}\|_2 \|\widetilde{U}_{m-\ell}^T A\|_2 = \|\widetilde{U}_{m-\ell}^T A\|_2. \tag{17}$$

We will use (16) to bound the right-hand side of (17). For notational simplicity, we define the function

$$\theta \to f(\cos(\theta)) = \kappa(\cos(\theta), \cos(\theta_k)),$$

for some $1 \le k \le m$. The coefficients in its Fourier expansion

$$f(\cos(\theta)) = \sum_{j=0}^\infty a_j(\theta_k)\cos(j\theta) \tag{18}$$

satisfy the inequality (16). For notational simplicity, we define

$$\alpha_j = a_j(\theta_k), \quad j = 0, 1, 2, \dots \ . \tag{19}$$

The vector $\boldsymbol{f} = [f(\cos(\theta_1)), f(\cos(\theta_2)), \dots, f(\cos(\theta_m))]^T$ is a tabulation of the function f. Define the $(k+1)$st axis vector $\boldsymbol{e}_{k+1} = [0. \dots, 0, 1, 0, \dots, 0]^T$ of suitable dimension. Then the kth Fourier coefficient of $\boldsymbol{f}$ is given by

$$\widehat{\alpha}_k = c_k \boldsymbol{e}_{k+1}^T U_\ell^T \boldsymbol{f}, \quad 0 \le k < \ell.$$

where $c_0 = 1/\sqrt{m}$ and $c_k = \sqrt{2/m}$ for $1 \le k < \ell$. Using (18) and (19), we obtain, for $0 \le k < \ell$,

$$\begin{aligned} \widehat{\alpha}_k &= c_k \sum_{j=0}^{\infty} \alpha_j \boldsymbol{e}_{k+1}^T U_\ell^T \begin{bmatrix} \cos(j\theta_1) \\ \cos(j\theta_2) \\ \vdots \\ \cos(j\theta_m) \end{bmatrix} \\ &= c_k^2 \sum_{j=0}^{\infty} \alpha_j \left(\sum_{p=1}^{m} \cos(k\theta_p) \cos(j\theta_p) \right). \end{aligned} \tag{20}$$

Let

$$s_{k,j} = \sum_{p=1}^{m} \cos(k\theta_p) \cos(j\theta_p), \qquad j, k = 0, 1, \dots \ . \tag{21}$$

When none of $j \pm k$ are multiples of $2m$, we have

$$s_{k,j} = \frac{1}{4} \left[\frac{\sin(\pi(j-k))}{\sin\left(\frac{\pi(j-k)}{2m}\right)} + \frac{\sin(\pi(j+k))}{\sin\left(\frac{\pi(j+k)}{2m}\right)} \right] = 0. \tag{22}$$

For other values of j and k, the sum (21) may be evaluated by taking limits in (22). For instance, when $k = j \ge 1$ and k is not a multiple of m, we obtain in this manner

$$s_{k,k} = \frac{1}{4} \left[\frac{\sin(\pi(2k))}{\sin\left(\frac{\pi(2k)}{2m}\right)} + 2m \right] = \frac{m}{2}. \tag{23}$$

If $j - k$ or $j + k$ are multiples of $2m$, then we can use the trivial bound

$$\left| \sum_{p=1}^{m} \cos(k\theta_p) \cos(j\theta_p) \right| \le m.$$

We conclude from (22) that most of the terms in the sum over j in (20) vanish. In particular, for $k \ge 1$,

$$\widehat{\alpha}_k = \frac{2}{m} (s_{k,k}\alpha_k + s_{k,2m-k}\alpha_{2m-k} + s_{k,2m+k}\alpha_{2m+k} + \dots \ ).$$

When the function f is smooth, the coefficients α_j converge to zero in magnitude quite rapidly; cf. (16). In typical applications $m \gg \ell$. Hence, $m \gg k$, and it suffices to use the approximation

$$\widehat{\alpha}_k \approx \alpha_k \tag{24}$$

when estimating the right-hand side of (17).

The entries of column i of the matrix $\widetilde{U}_{m-\ell}^T A$ are discrete Fourier coefficients of the function

$$\theta \to \frac{\pi}{m} \kappa(\cos(\theta), \cos(\theta_i)) \tag{25}$$

with the coefficient for $\cos(\ell\theta)$ in the first row. The bound (16) suggests that the entries of each row of the matrix $\widetilde{U}_{m-\ell}^T A$ are roughly of the same order of magnitude, and that typically the matrix entries of largest magnitude can be found in the first row. The latter entries provide the major contribution to the norm of $\widetilde{U}_{m-\ell}^T A$ when the functions (25) have the same smoothness for all i.

Denote the entries of the first column of $\widetilde{U}_{m-\ell}^T A$ by $\widetilde{\alpha}_\ell, \widetilde{\alpha}_{\ell+1}, \dots, \widetilde{\alpha}_{m-1}$; they are the discrete Fourier coefficients of the function (25) for $i = 1$. The fact that the entries in each row are of about the same magnitude suggests the approximation

$$\|\widetilde{U}_{m-\ell}^T A\|_2^2 \le \|\widetilde{U}_{m-\ell}^T A^T A\|_F^2 \approx (m-\ell) \sum_{j=\ell}^{m-1} \widetilde{\alpha}_j^2 < m \sum_{j=\ell}^{m-1} \widetilde{\alpha}_j^2.$$

Using the approximation (24) and the scaling $\widetilde{\alpha}_j = \frac{\pi}{m} \widehat{\alpha}_j$ in (25) gives

$$\|\widetilde{U}_{m-\ell}^T A\|_2^2 \lessapprox \frac{\pi^2}{m} \sum_{j=\ell}^{m-1} \alpha_j^2.$$

The bound (16) for the right-hand side then yields

$$\|\widetilde{U}_{m-\ell}^T A\|_2^2 \lessapprox \frac{(m-\ell-1)4\gamma^2}{m(\ell-\nu)^{2\nu+2}} < \frac{4\gamma^2}{(\ell-\nu)^{2\nu+2}} \quad \text{for} \quad \ell > \nu.$$

This gives (15). □

The above theorem shows that the left-hand side of (15) is small when ℓ is large and the kernel κ is smooth, i.e., when ν can be chosen fairly large. If the kernel κ is an analytic function, which is the case in some ill-posed problems, then the bound (15) can be strengthened by using bounds for the Fourier coefficients [25, Theorem 8.1] of analytic functions in a similar way as we applied [25, Theorem 7.1].

We remark that while Theorem 1 requires the discretization (14) of the kernel of (2), the derivation of the reduced problem (13) only demands that the columns of the matrix U_ℓ are orthonormal. This is illustrated in the computed examples of Section V.

The following method describes another discretization of the integral equation (2).

Method II.2. The change of variables $s = \cos(\theta)$ and $t = \cos(\tau)$ in the integral equation (2) gives

$$\int_0^\pi \kappa(\cos(\theta), \cos(\tau)) \sin(\tau) x(\cos(\tau)) d\tau = b(\cos(\theta)),$$

with $0 \le \theta \le \pi$. Application of a Nyström integration method based on a composite midpoint rule with nodes $\theta_j = (2j-1)\pi/(2m)$, $j = 1, 2, \dots, m$, gives a least square problem (1) defined by

$$\begin{aligned} A &= [a_{i,j}]_{i,j=1}^m, & a_{i,j} &= \tfrac{\pi}{m} \kappa(t_i, t_j) \sin(\theta_j), \\ \boldsymbol{b}_{\text{exact}} &= [b_i]_{i=1}^m, & b_i &= b(t_i), \end{aligned}$$

where the t_i are defined by (10). Thus, the matrix in this method differs from the matrix in Theorem 1 by column

scaling. This implies that bounds analogous to (15) can be shown similarly as in the proof of Theorem 1.

The following method describes a third way to apply the solution subspace $\mathrm{range}(U_\ell)$ to the solution of linear discrete ill-posed problems.

Method II.3. Let the matrix $U_\ell \in \mathbb{R}^{n\times\ell}$ be defined similarly as (12) with m replaced by n. Our approach to determine an approximate solution of (1) is analogous to the one applied in the randomized method described in Section III and [26], [27]. First compute the QR factorization

$$Q_\ell R_\ell = AU_\ell,$$

where $Q_\ell \in \mathbb{R}^{m\times\ell}$ has orthonormal columns and $R_\ell \in \mathbb{R}^{\ell\times\ell}$ is upper triangular. Then evaluate the reduced SVD

$$Q_\ell^T A = \widetilde{W}\widetilde{\Sigma}\widetilde{V}^T.$$

Thus, the matrices $\widetilde{W} \in \mathbb{R}^{\ell\times\ell}$ and $\widetilde{V} \in \mathbb{R}^{n\times\ell}$ have orthonormal columns, and $\widetilde{\Sigma} \in \mathbb{R}^{\ell\times\ell}$ is a diagonal matrix with nonnegative nonincreasing diagonal entries. The matrix $Q_\ell Q_\ell^T A$ is a low-rank approximation of A (of rank $\leq \ell$) and we have

$$Q_\ell Q_\ell^T A = Q_\ell\widetilde{W}\widetilde{\Sigma}\widetilde{V}^T, \tag{26}$$

where we consider the right-hand side an approximation of the SVD of A. This approximate SVD of A can be evaluated fairly inexpensively. The computation of the matrices Q_ℓ, $\widetilde{W}$, $\widetilde{\Sigma}$, and $\widetilde{V}$ includes QR factorization that requires $2m\ell^2 - \frac{2}{3}\ell^3$ arithmetic floating point operations (flops) and about $14n\ell^2 + 8\ell^3$ flops for the computation of the SVD of $B = Q_\ell^T A$. We also have to evaluate matrix-matrix products AU_ℓ and B. By taking advantage of level-3 BLAS, these can be evaluated very efficiently.

Replacing A by its low-rank approximation $Q_\ell Q_\ell^T A$ in the Tikhonov minimization problem (5) yields

$$\begin{aligned}\min_{\boldsymbol{x}\in\mathbb{R}^n}\{\|Q_\ell Q_\ell^T A\boldsymbol{x} - Q_\ell Q_\ell^T\boldsymbol{b}\|_2^2 + \mu\|L\boldsymbol{x}\|_2^2\}\\ + \|(I - Q_\ell Q_\ell^T)\boldsymbol{b}\|_2^2,\end{aligned} \tag{27}$$

which is equivalent to the minimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}\{\|\widetilde{W}\widetilde{\Sigma}\widetilde{V}^T\boldsymbol{x} - Q_\ell^T\boldsymbol{b}\|_2^2 + \mu\|L\boldsymbol{x}\|_2^2\}. \tag{28}$$

We would like to determine a solution of as small norm as possible. It is therefore natural to require the solution to be of the form $\boldsymbol{x} = \widetilde{V}\boldsymbol{y}$ for some $\boldsymbol{y} \in \mathbb{R}^\ell$. We then obtain

$$\min_{\boldsymbol{y}\in\mathbb{R}^\ell}\{\|\widetilde{\Sigma}\boldsymbol{y} - \widetilde{W}^T Q_\ell^T\boldsymbol{b}\|_2^2 + \mu\|L\widetilde{V}\boldsymbol{y}\|_2^2\}. \tag{29}$$

Consider the QR factorization $\widetilde{Q}\widetilde{R} = L\widetilde{V}$, where $\widetilde{Q} \in \mathbb{R}^{p\times\ell}$ has orthonormal columns and $\widetilde{R} \in \mathbb{R}^{\ell\times\ell}$ is upper triangular (we assume here that $p \geq \ell$). The matrix $\widetilde{R}$ typically is nonsingular and fairly well-conditioned. Therefore, we can make the change of variables $\boldsymbol{z} = \widetilde{R}\boldsymbol{y}$, which transforms (29) to a Tikhonov minimization problem in standard form. Thus, we obtain

$$\min_{\boldsymbol{y}\in\mathbb{R}^\ell}\{\|\widetilde{\Sigma}\boldsymbol{y} - \widetilde{W}^T Q_\ell^T\boldsymbol{b}\|_2^2 + \mu\|\widetilde{R}\boldsymbol{y}\|_2^2\}. \tag{30}$$

When L is the identity matrix, we have $\|L\widetilde{V}\boldsymbol{y}\|_2 = \|\boldsymbol{y}\|_2$ and no change of variable is necessary.

We turn to the situation when $m < n$. Let $\widehat{U}_\ell^T = U_\ell \in \mathbb{R}^{m\times\ell}$ be defined as in (12) and compute the QR factorization

$$(\widehat{U}_\ell A)^T = Q_\ell R,$$

where $Q_\ell \in \mathbb{R}^{n\times\ell}$ and $R \in \mathbb{R}^{\ell\times\ell}$. Then calculate the SVD of AQ_ℓ,

$$AQ_\ell = \widetilde{U}\widetilde{\Sigma}\widetilde{W}^T.$$

An approximation of the SVD of A is furnished by

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

and an approximate solution of (5) can be determined by solving

$$\min_{\boldsymbol{x}\in\mathcal{R}(Q_\ell)}\{\|AQ_\ell Q_\ell^T\boldsymbol{x} - \widetilde{U}\widetilde{U}^T\boldsymbol{b}\|_2^2 + \mu\|L\boldsymbol{x}\|_2^2\}.$$

The substitution $\boldsymbol{x} = Q_\ell\widetilde{W}\boldsymbol{y}$ gives

$$\min_{\boldsymbol{y}\in\mathbb{R}^\ell}\{\|\widetilde{\Sigma}\boldsymbol{y} - \widetilde{U}^T\boldsymbol{b}\|_2^2 + \mu\|LQ_\ell\widetilde{W}\boldsymbol{y}\|_2^2\}. \tag{31}$$

Using a similar approach as when $m \geq n$, we determine the QR factorization of $LQ_\ell\widetilde{W}$ to obtain an upper triangular matrix $\widetilde{R} \in \mathbb{R}^{\ell\times\ell}$ which, generally, is nonsingular and fairly well conditioned. The minimization problem (31) can be transformed to standard form with the change of variables $\boldsymbol{z} = \widetilde{R}\boldsymbol{y}$.

The computations outlined can be supplemented by iterations with the power method, similarly as described by Bai et al. [3] and Halko et al. [15] in the context of randomized methods. Such an approach determines the singular vectors of the matrix $M = (AA^\top)^q A$ for a small integer $q \geq 1$. Note that the matrix M has the same singular vectors as A, while its singular values satisfy $\sigma_j(M) = \sigma_j(A)^{2q+1}$, $j = 1, 2, \ldots$. It follows that the singular values of M decay faster to zero than the singular values $\sigma_j(A)$ of A. In this approach, the matrix M should not be explicitly formed; see [15] for more details.

We apply the discrepancy principle to determine the regularization parameter as described in [2]. We solve (5) by computing the solution of (30) or (31). With the assumption that the error $\boldsymbol{e}$ in $\boldsymbol{b}$ is normally distributed with zero mean and variance ϵ^2, since $Q_\ell Q_\ell^T$ is an orthogonal projector, the variance of $Q_\ell Q_\ell^T\boldsymbol{e}$ is the fraction $\frac{\ell}{n}\epsilon^2$. We therefore seek to determine the regularization parameter μ such that

$$\|\widetilde{\Sigma}\boldsymbol{y} - \widetilde{W}^T Q_\ell^T\boldsymbol{b}\|_2 = \eta\sqrt{\frac{\ell}{n}}\epsilon.$$

Thus, we replace ϵ by $\sqrt{\frac{\ell}{n}}\epsilon$ in (6).

III. Randomized SVD methods

Randomized SVD methods can be applied to determine a low-rank approximation of the SVD (7); see, e.g., [15]. This approach is applied in truncated SVD methods, see, e.g., [3]. Xiang and Zou [26] describe how randomized SVD methods can be used to compute approximate solutions of Tikhonov

regularization problems in standard form, i.e., when $L = I$ in (5), and in [27] propose several randomized methods for the solution of Tikhonov minimization problems in general form, i.e., when $L \neq I$. We outline these approaches in the following paragraphs and compare their performance in Section V to the approach described in Section II.

First consider the case $m \geq n$. 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 variance one, 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 and assumed to be nonsingular in [26]. 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 of B,

$$B = \widehat{W}\widehat{\Sigma}\widehat{V}^T, \tag{32}$$

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 decreasing diagonal entries. The right-hand side of

$$Q_\ell B = Q_\ell Q_\ell^T A = Q_\ell \widehat{W}\widehat{\Sigma}\widehat{V}^T \tag{33}$$

is an approximation of the SVD (7) of A. The decomposition (33) is much cheaper to compute than (7) when m and n are large and $1 \leq \ell \ll n \leq m$. A bound for the distance between A and $Q_\ell Q_\ell^T A$ easily can be established. The lower bound of the following theorem follows from the Eckart–Young theorem and the upper bound is shown in [15, Corollary 10.9].

Theorem 2. *Suppose that $A \in \mathbb{R}^{m\times n}$ has the singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min\{m,n\}} \geq 0$. 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*

$$\begin{aligned}\sigma_{\ell+1} &\leq \|A - Q_\ell Q_\ell^T A\|_2 \\ &\leq (1 + 6\sqrt{\ell p \log p})\sigma_{k+1} + 3(\ell\Sigma_{j>k}\sigma_j^2)^{1/2}\end{aligned}$$

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

Replacing A by the approximation (33) in the minimization problem (5) gives

$$\begin{aligned}\min_{\boldsymbol{x}\in\mathbb{R}^n} \{&\|Q_\ell Q_\ell^T A\boldsymbol{x} - Q_\ell Q_\ell^T \boldsymbol{b}\|_2^2 + \mu\|L\boldsymbol{x}\|_2^2\} \\ &+ \|(I - Q_\ell Q_\ell^T)\boldsymbol{b}\|_2^2,\end{aligned}$$

which is equivalent to

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \{\|\widehat{W}\widehat{\Sigma}\widehat{V}^T\boldsymbol{x} - Q_\ell^T\boldsymbol{b}\|_2^2 + \mu\|L\boldsymbol{x}\|_2^2\}.$$

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

$$\min_{\boldsymbol{y}\in\mathbb{R}^\ell} \{\|\widehat{\Sigma}\boldsymbol{y} - \widehat{W}^T Q_\ell^T \boldsymbol{b}\|_2^2 + \mu\|L\widehat{V}\boldsymbol{y}\|_2^2\}.$$

When $L = I$, we have $\|\widehat{V}\boldsymbol{y}\|_2 = \|\boldsymbol{y}\|_2$. For a general matrix $L \in \mathbb{R}^{p\times n}$, we compute the QR factorization $\widehat{Q}\widehat{R} = L\widehat{V}$. When the matrix $\widehat{R}$ is of full rank and fairly well-conditioned, we proceed similarly as in Section II, i.e., we carry out the substitution $\boldsymbol{z} = \widehat{R}\boldsymbol{y}$.

We turn to the situation when $m < n$ and $1 \leq \ell \ll m$. We construct a random matrix $\Omega_\ell \in \mathbb{R}^{\ell\times m}$ with random entries with distribution $N(0, 1)$, and compute the economical QR factorization $(\Omega_\ell A)^T = Q_\ell R_\ell$, where $Q_\ell \in \mathbb{R}^{n\times\ell}$ has orthonormal columns and $R \in \mathbb{R}^{\ell\times\ell}$ is upper triangular. Then we proceed similarly as in Section II.

We determine the regularization parameter $\mu > 0$ by discrepancy principle, where we replace ϵ by $\sqrt{\frac{\ell}{n}}\epsilon$ in (6), similarly as in Section II.

IV. A solution method based on Golub–Kahan bidiagonalization

This section reviews the use of Golub–Kahan bidiagonalization to reduce the large matrix A in (1) to a small bidiagonal matrix; see, e.g., [6], [14]. Application of $1 \leq \ell \ll \min\{m, n\}$ steps of Golub–Kahan bidiagonalization to A with initial vector $\boldsymbol{b}$ gives the decompositions

$$AV_\ell = U_{\ell+1}\bar{C}_\ell, \qquad A^T U_\ell = V_\ell C_\ell^T, \tag{34}$$

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

$$U_{\ell+1}\boldsymbol{e}_1 = \boldsymbol{b}/\|\boldsymbol{b}\|_2. \tag{35}$$

Moreover, $\mathrm{range}(V_\ell)$ is the Krylov subspace

$$\mathbb{K}_\ell(A^TA, A^T\boldsymbol{b}) = \mathrm{span}\{A^T\boldsymbol{b}, (A^TA)A^T\boldsymbol{b}, \ldots, (A^TA)^{\ell-1}A^T\boldsymbol{b}\}, \tag{36}$$

and the matrix

$$\bar{C}_\ell = \begin{bmatrix} \rho_1 & & & & 0 \\ \tau_2 & \rho_2 & & & \\ & \ddots & \ddots & & \\ & & \tau_{\ell-1} & \rho_{\ell-1} & \\ & & & \tau_\ell & \rho_\ell \\ 0 & & & & \tau_{\ell+1} \end{bmatrix} \in \mathbb{R}^{(\ell+1)\times\ell}$$

is lower bidiagonal with positive entries τ_k and ρ_k, and $C_\ell \in \mathbb{R}^{\ell\times\ell}$ is obtained by removing the last row of $\bar{C}_\ell$. We assume that ℓ is small enough so that the decompositions (34) with the described properties exist. This is the generic situation.

We compute an approximate solution of (5) by minimizing over the Krylov subspace (36) instead of over $\mathbb{R}^n$,

$$\min_{\boldsymbol{x}\in\mathbb{K}_\ell(A^TA, A^T\boldsymbol{b})} \{\|A\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \mu\|L\boldsymbol{x}\|_2^2\}.$$

Letting $\boldsymbol{x} = V_\ell\boldsymbol{y}$, the relations (34) and (35) give

$$\min_{\boldsymbol{y}\in\mathbb{R}^\ell} \{\|\bar{C}_\ell\boldsymbol{y} - \boldsymbol{e}_1\|\boldsymbol{b}\|_2\|_2^2 + \mu\|LV_\ell\boldsymbol{y}\|_2^2\}. \tag{37}$$

Denote the solution of (37) by $\boldsymbol{y}_{\mu,\ell}$. Then $\boldsymbol{x}_{\mu,\ell} = V_\ell\boldsymbol{y}_{\mu,\ell}$ is an approximate solution of (5).

First consider the situation when $L = I$. Then

$$\|A\boldsymbol{x}_{\mu,\ell} - \boldsymbol{b}\|_2 = \|\bar{C}_\ell \boldsymbol{y}_{\mu,\ell} - \boldsymbol{e}_1\|\boldsymbol{b}\|_2\,\|_2. \tag{38}$$

Therefore, it suffices to choose $\mu > 0$ so that the right-hand side satisfies the discrepancy principle; see [6] for details. It follows that it is quite inexpensive to determine a value of $\mu > 0$ so that the approximate solution $\boldsymbol{x}_{\mu,\ell} = V_\ell \boldsymbol{y}_{\mu,\ell}$ of (5) satisfies (6). A discussion on how this can be done by Newton's method can be found in [6]. Other zero-finders are discussed in [5], [23].

The number of bidiagonalization steps, ℓ, does affect the value of μ, since (38) decrease as ℓ increases. We choose ℓ, as small as possible to satisfy the discrepancy principle for some $0 < \mu < \infty$, i.e., we choose ℓ so that

$$\|\bar{C}_\ell \boldsymbol{y}_{\mu,\ell} - \boldsymbol{e}_1\|\boldsymbol{b}\|_2\,\|_2 < \eta\epsilon \leq \|\bar{C}_{\ell-1} \boldsymbol{y}_{\mu,\ell-1} - \boldsymbol{e}_1\|\boldsymbol{b}\|_2\,\|_2.$$

When $L \neq I$, there are several different approaches; see, e.g., [7], [10], [18]. In the present paper, we will apply the method described in [18] in experiments: Let $L \in \mathbb{R}^{p\times n}$ and assume that ℓ in (34) satisfies $1 \leq \ell \leq \min\{p, n\}$. Compute the QR factorization $Q_\ell R_\ell = LV_\ell$, 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 (37) becomes

$$\min_{\boldsymbol{y}\in\mathbb{R}^\ell} \{\|\bar{C}_\ell \boldsymbol{y} - \boldsymbol{e}_1\|\boldsymbol{b}\|_2\,\|_2^2 + \mu\|R_\ell \boldsymbol{y}\|_2^2\}.$$

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

$$\min_{\boldsymbol{z}\in\mathbb{R}^\ell} \{\|\bar{C}_\ell R_\ell^{-1} \boldsymbol{z} - \boldsymbol{e}_1\|\boldsymbol{b}\|_2\,\|_2^2 + \mu\|\boldsymbol{z}\|_2^2\}.$$

If μ is determined so that the solution $\boldsymbol{z}_{\mu,\ell}$ satisfies $\|\bar{C}_\ell R_\ell^{-1} \boldsymbol{z} - \boldsymbol{e}_1\|\boldsymbol{b}\|_2\,\|_2 = \eta\epsilon$, then the associated approximate solution $\boldsymbol{x}_{\mu,\ell} = V_\ell R_\ell^{-1} \boldsymbol{z}_{\mu,\ell}$ of (5) satisfies the discrepancy principle (6).

V. Computed examples

We consider a few examples from Regularization Tools [16], and compare the three solution methods that use a solution subspace made up of discretized Chebyshev polynomials described in Section II to the randomized and Krylov methods outlined in Sections III and IV, respectively. The methods are compared with respect to efficiency and accuracy. To measure the time, we solve all the problems by all the methods 100 times and report average timings. The accuracy is measured by the Relative Restoration Error (RRE),

$$\mathrm{RRE}(\boldsymbol{x}) = \frac{\|\boldsymbol{x} - \boldsymbol{x}_{\mathrm{exact}}\|_2}{\|\boldsymbol{x}_{\mathrm{exact}}\|_2},$$

where $\boldsymbol{x}_{\mathrm{exact}}$ denotes the desired minimal-norm solution of the linear system of equations (4). We are particularly interested in assessing how the performance of the methods depend on the dimension ℓ of the solution subspace.

The exact data vector is defined by $\boldsymbol{b}_{\mathrm{exact}} = A\boldsymbol{x}_{\mathrm{exact}}$. To simulate error-contaminated data, we add a vector $\boldsymbol{e}$ that models normally distributed Gaussian noise with zero mean and with the variance chosen to achieve a specified noise level

$$\delta = \frac{\|\boldsymbol{e}\|_2}{\|\boldsymbol{b}_{\mathrm{exact}}\|_2};$$

cf. (3).

We will consider two examples in one space-dimension and two examples in two space-dimensions. All examples are discretizations of Fredholm integral equations of the first kind (2). For problems in one space-dimension, we use the regularization matrix

$$L = L_1 = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix},$$

i.e., the Laplacian matrix in one space-dimension, while for examples in two space-dimensions, we use the Laplacian matrix in two space-dimensions. The latter is defined by

$$L = L_1 \otimes I + I \otimes L_1,$$

where I is the identity matrix and $\otimes$ denotes the Kronecker product.

In all the examples, we consider $A \in \mathbb{R}^{4096\times 4096}$. In the tables we denote by

- "Cheb. basis" the method described in Method II.1;
- "Cheb.-Nyström" the method described in Method II.2;
- "Cheb. projection" the method described in Method II.3;
- "Randomized" the method described in Section III;
- "Krylov" the method described in Section IV.

All computations were carried out using MATLAB 2019b with approximately 15 significant decimal digits running on a desktop computer with an Intel Core i5-9400 @2.90GHz CPU and 8 GB of RAM under Windows 10.

a) Baart: We first consider the Baart example. The kernel $k(s,t)$ in (2) is defined by

$$k(s,t) = \mathrm{e}^{s\cos(t)},\ s \in [0, \pi/2],\ t \in [0, \pi], \tag{39}$$

and the right hand-side is given by $b(s) = 2\sinh(s)/s$. Then the solution is given by $x(t) = \sin(t)$. This integral equation is described in [1]. We discretize the integral equation (2) in two ways: For the "Cheb. basis" and "Cheb-Nyström" methods,as described in Methods II.1 and II.2. For the method described in Method II.3 as well as for the randomized and Krylov methods of Sections III and IV, we use the discretization provided by the MATLAB function baart.m in [16]. This function discretizes (2) by a Galerkin method with piece-wise constant test and trial functions. Both discretizations furnish the error-free data vector $\boldsymbol{b}_{\mathrm{exact}}$ to which we add a vector $\boldsymbol{e}$ that represents normally distributed Gaussian noise as described above. This vector is scaled to correspond to the noise level $\delta = 0.01$.

We carry out computations with all methods for subspaces of dimensions $\ell \in \{1, 2, \ldots, 50\}$. Tables I and II report the

RRE and CPU time for all methods for $\ell \in \{10, 25, 50\}$. We observe that the methods described in Methods II.1 and II.2 require about the same CPU time and are faster than the other methods considered. The fact that the CPU time demanded by the "Cheb. basis" and "Cheb-Nyström" methods are about the same depends on that the computations are essentially the same. The "Cheb. projection" method requires about the same CPU time as the "randomized" method, because the computations are quite similar; they only differ in that the starting point for the "Cheb. projection" method is an orthonormal basis of discretized Chebyshev polynomials, while the starting point for the "randomized" method is a Gaussian random matrix. Both the latter methods require more CPU time than the "Cheb. basis" and "Cheb-Nyström" methods. The last method in our comparison is the Krylov subspace Tikhonov regularization method based on partial Golub–Kahan bidiagonalization. This method is referred to as "Krylov". Table II shows this method to require the most CPU time. The behavior of the CPU time as a function of the dimension ℓ of the solution subspace is reported in Figure 3(a).

Turning to the error in the computed approximate solutions, we can observe that the "Cheb-Nyström" method provides the most accurate reconstructions, while the other two Chebyshev methods outperform the randomized and Krylov methods. We report the RRE obtained for solution subspace dimensions $1 \leq \ell \leq 50$ in ℓ in Figure 2(a).

b) Shaw: The integral equation (2) of this example is described by Shaw [24]. It has the kernel

$$k(s,t) = (\cos(s) + \cos(t)) \left(\frac{\sin(u)}{u} \right)^2,$$

where $u = \pi(\sin(s) + \sin(t))$ and $s, t \in [-\pi, \pi]$. The exact solution is $x(t) = \alpha_1 \mathrm{e}^{-\gamma_1 (t-\tau_1)^2} + \alpha_2 \mathrm{e}^{-\gamma_2 (t-\tau_2)^2}$, where the α_j, γ_j, and τ_j are given constants. Discretization gives the matrix A and the vector $\boldsymbol{x}_{\mathrm{exact}}$. The error-free data vector $\boldsymbol{b}_{\mathrm{exact}}$ is obtained by multiplying A and $\boldsymbol{x}_{\mathrm{true}}$. The error-contaminated data vector $\boldsymbol{b}$ is determined by (3). The noise level is $\delta = 0.03$.

Similarly as in the previous example, we report the CPU time required and the RRE for the different methods and for three solution subspace dimensions in Tables I and II. The results for this example are similar to those for the previous one. Also in this example the "Cheb. basis" method outperforms all the other considered methods both in terms of accuracy and computational cost. Figure 2(a) displays all the RREs obtained for different choices of solution subspace dimension ℓ. We can observe that all the methods, except for the "Cheb-Nyström" method, achieve minimal RRE for $\ell = 6$. After this minimum is reached the RRE stabilizes. The "Cheb. basis" method provides the most consistent result. Finally, Figure 3(b) displays the CPU time as a function of the dimension of the solution subspace ℓ.

c) Blur: Our first example in two space-dimensions is an image deblurring problem. We blur the image in Figure 1(a) by the Gaussian PSF shown in Figure 1(b). The PSF is defined by the MATLAB function blur.m from [16] with parameters band $= 30$ and sigma $= 1.5$. The blurred but noise-free image is represented by a vector $\boldsymbol{b}_{\mathrm{exact}}$. We add 2% white Gaussian noise to this vector to obtain the blurred and noisy image shown in Figure 1(c). The latter image is represented by the data vector $\boldsymbol{b}$.

The "Cheb. basis" method of Method II.1 can be applied because the matrix A in (13) can be an arbitrary square matrix; it does not have to be determined by the discretization method of Theorem 1. However, the "Cheb.-Nyström" method of Method II.2 requires the discretization used in the description of the method. Since the MATLAB function blur.m provides a different discretization, the "Cheb.-Nyström" cannot be used. This is a limitation of the "Cheb.-Nyström" method.

The RREs and CPU times for the methods in our comparison are reported in Tables I and II. As in the previous examples, the Krylov method is the slowest among the considered algorithms. However, differently from the examples above, the non-Krylov methods do not provide accurate reconstructions. For the randomized method a similar observation is reported in [2]. The reason for the poor performance of the randomized method is that the singular values of the matrix A do not decay to zero fast enough to make the right-hand side of the inequality of Theorem 2 small. This means that the dimension ℓ of the solution subspace has to be very large to be able to determine an accurate approximation of the desired solution. Expansion in terms of discretized Chebyshev polynomials does not yield high accuracy either, because when the kernel is regarded as a function in one space-dimension it is not smooth. Therefore, its Fourier coefficients will decay slowly with increasing index number. Figures 2(c) and 3(c) report the RREs and CPU times for all methods and values of ℓ considered.

d) Baart2D: Our last example in two space-dimensions is the tensor product of two Baart examples. Let $A_1 \in \mathbb{R}^{64 \times 64}$ be a discretization of (39). Then we define

$$A = A_1 \otimes A_1.$$

The exact solution and the right-hand side are obtained as tensor products of the counterparts in one space-dimension. To simulate realistic data we add 2% white Gaussian noise.

We can observe from the RREs reported in Table I and in Figure 2(d) that the Krylov method, the randomized method, and the "Cheb. projection" method determine accurate reconstructions. The latter two methods are able to determine approximate solutions of high quality due to the fact that the singular values of A decay to zero with increasing index extremely fast; this was already observed in [2]. The "Cheb. basis" and "Cheb-Nyström" methods do not provide satisfactory results. This is due to the fact that we represent a smooth two-dimensional surface in one space-dimension, and when considered as a function in one space-dimension the function is not smooth. We report the CPU times in Table II and Figure 3(d).

The Baart2D problem also can be represented as $A_1 X A_1 \approx B$, where $X, B \in \mathbb{R}^{64 \times 64}$, and B represents the available contaminated data. We represent the solution as a linear

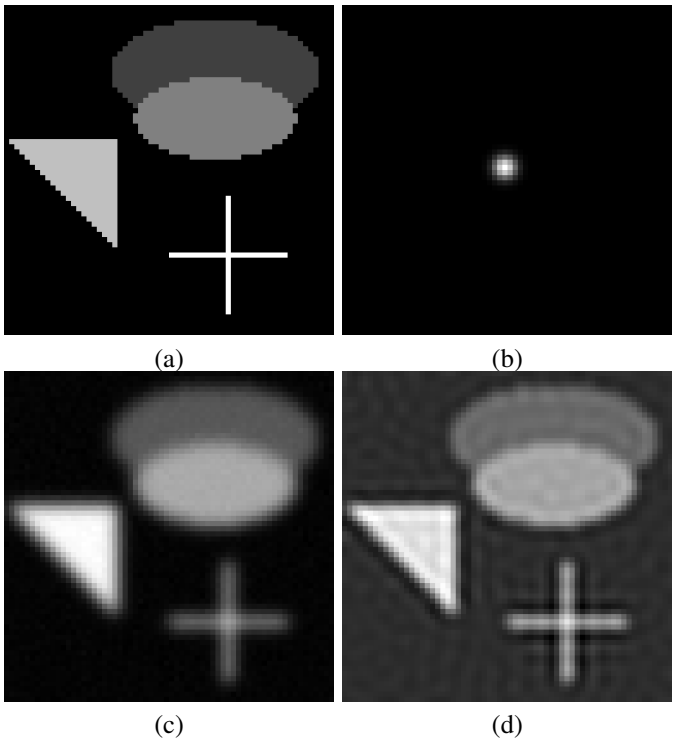


Fig. 1. Blur test case: (a) exact image (64×64 pixels), (b) PSF (64×64 pixels), (c) blurred and noisy image ($\delta = 2\%$), (d) Reconstruction obtained with the considered Krylov method.

combination of outer products of discretized Chebyshev polynomials. Thus, let the u_j, $j = 1, 2, \ldots$, denote orthonormal vectors obtained by discretizing Chebyshev polynomials at 64 nodes. This is analogous to the construction of the vectors u_j in Section II. Then define the matrix basis $\{u_1u_1^T, u_1u_2^T, u_1u_3^T, u_2u_1^T, u_2u_2^T\}$. We seek to determine an approximate solution of the Baart2D problem as a linear combination of these matrices. The noise level in B is chosen to be the same as for the computations reported for Baart2D in Table I. When using the 5-dimensional basis described, we obtain the RRE $1.96 \cdot 10^{-1}$, and when discretizing the integral equation by the trapezoidal rule in two space-dimensions, we achieve RRE $1.98 \cdot 10^{-1}$. The randomized and Krylov methods with $\ell = 5$ basis vectors yield the RREs $3.22 \cdot 10^{-1}$ and $2.23 \cdot 10^{-1}$, respectively. We conclude that discretized Chebyshev expansions can be competitive when the problem to be solved has a Kronecker product structure.

VI. Conclusion

We have presented three methods that use solution subspaces defined by discretized Chebyshev polynomials for the solution of ill-posed inverse problems. These methods can be used to solve large-scale problems and are found to be competitive with methods that are based on randomization and partial Golub–Kahan bidiagonalization for certain ill-posed problems.

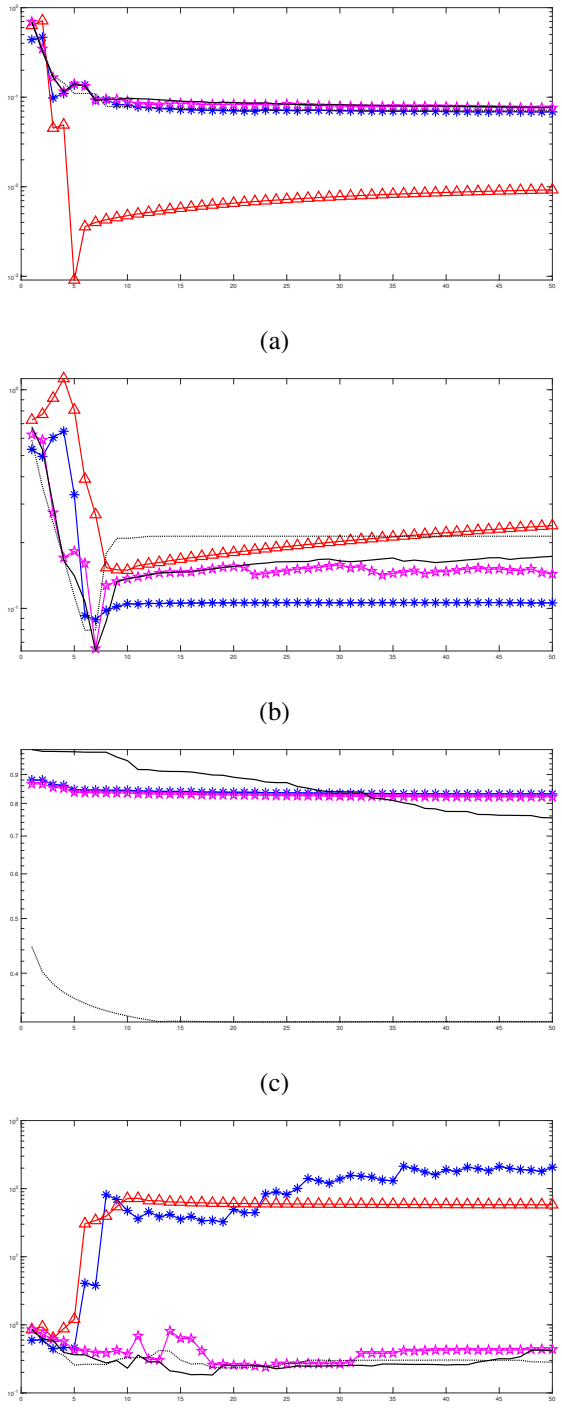
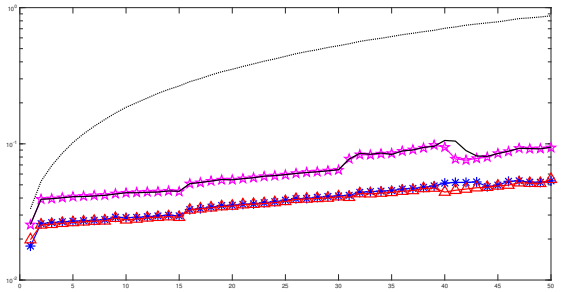
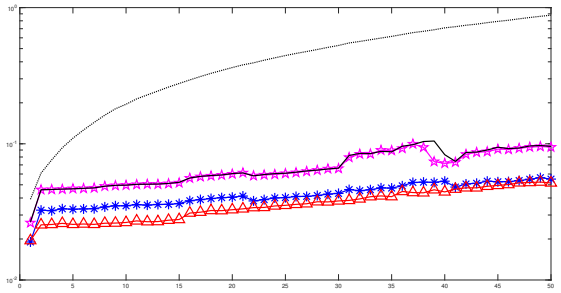


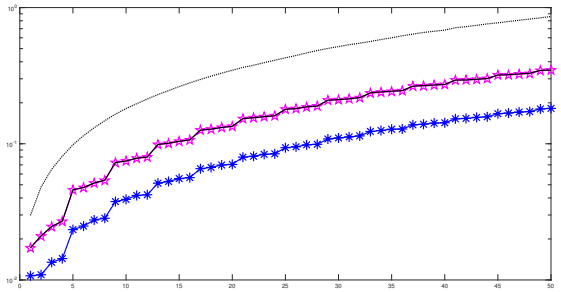
Fig. 2. RRE against ℓ for all considered examples and methods. The blue line with asterisks reports the results obtained using the Chebyshev basis, the red line with triangles reports the results obtained with the Nyström discretization, the magenta line with stars reports the results obtained projecting using the Chebyshev basis, the black solid line reports the results obtained with the randomized method, and the black dotted line reports the results obtained with the considered Krylov method. Panel (a) refers to the Baart example, panel (b) refers to the Shaw example, panel (c) refers to the Blur example, and panel (d) refers to the Baart2D example.



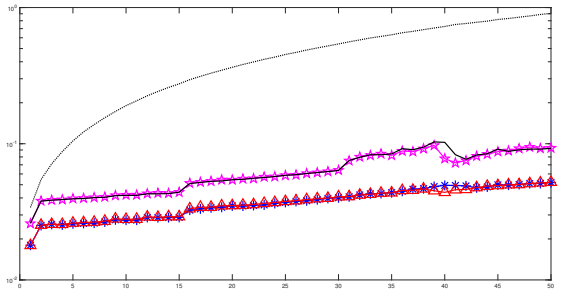
(a)



(b)



(c)



(d)

Fig. 3. CPU time (sec.) against ℓ for all considered examples and methods. The blue line with asterisks reports the results obtained using the Chebyshev basis, the red line with triangles reports the results obtained with the Nystrom discretization, the magenta line with stars reports the results obtained projecting using the Chebyshev basis, the black solid line reports the results obtained with the randomized method, and the black dotted line reports the results obtained with the considered Krylov method. Panel (a) refers to the Baart example, panel (b) refers to the Shaw example, panel (c) refers to the Blur example, and panel (d) refers to the Baart2D example.

TABLE I

RRE OBTAINED WITH THE CONSIDERED EXAMPLES FOR THE DIFFERENT METHODS AND SELECTED VALUES OF ℓ.

Example	Method	$\ell = 10$	$\ell = 25$	$\ell = 50$
Baart	Cheb. basis	$8.26 \cdot 10^{-2}$	$7.11 \cdot 10^{-2}$	$6.86 \cdot 10^{-2}$
	Cheb.-Nyström	$4.75 \cdot 10^{-3}$	$7.22 \cdot 10^{-3}$	$9.24 \cdot 10^{-3}$
	Cheb. projection	$9.05 \cdot 10^{-2}$	$8.36 \cdot 10^{-2}$	$7.61 \cdot 10^{-2}$
	Randomized	$9.77 \cdot 10^{-2}$	$8.38 \cdot 10^{-2}$	$7.68 \cdot 10^{-2}$
	Krylov	$7.85 \cdot 10^{-2}$	$7.19 \cdot 10^{-2}$	$7.14 \cdot 10^{-2}$
Shaw	Cheb. basis	$1.05 \cdot 10^{-1}$	$1.06 \cdot 10^{-1}$	$1.06 \cdot 10^{-1}$
	Cheb.-Nyström	$1.49 \cdot 10^{-1}$	$1.92 \cdot 10^{-1}$	$2.39 \cdot 10^{-1}$
	Cheb. projection	$1.37 \cdot 10^{-1}$	$1.49 \cdot 10^{-1}$	$1.44 \cdot 10^{-1}$
	Randomized	$1.37 \cdot 10^{-1}$	$1.63 \cdot 10^{-1}$	$1.73 \cdot 10^{-1}$
	Krylov	$2.09 \cdot 10^{-1}$	$2.14 \cdot 10^{-1}$	$2.14 \cdot 10^{-1}$
Blur	Cheb. basis	$8.44 \cdot 10^{-1}$	$8.35 \cdot 10^{-1}$	$8.32 \cdot 10^{-1}$
	Cheb. projection	$8.35 \cdot 10^{-1}$	$8.26 \cdot 10^{-1}$	$8.22 \cdot 10^{-1}$
	Randomized	$9.52 \cdot 10^{-1}$	$8.71 \cdot 10^{-1}$	$7.54 \cdot 10^{-1}$
	Krylov	$3.36 \cdot 10^{-1}$	$3.28 \cdot 10^{-1}$	$3.28 \cdot 10^{-1}$
Baart2D	Cheb. basis	$4.76 \cdot 10^{1}$	$8.21 \cdot 10^{1}$	$2.06 \cdot 10^{2}$
	Cheb.-Nyström	$7.19 \cdot 10^{1}$	$5.97 \cdot 10^{1}$	$5.80 \cdot 10^{1}$
	Cheb. projection	$3.69 \cdot 10^{-1}$	$2.69 \cdot 10^{-1}$	$4.37 \cdot 10^{1}$
	Randomized	$2.31 \cdot 10^{-1}$	$2.35 \cdot 10^{-1}$	$4.15 \cdot 10^{-1}$
	Krylov	$3.28 \cdot 10^{-1}$	$2.74 \cdot 10^{-1}$	$2.83 \cdot 10^{-1}$

TABLE II

CPU TIME(SEC.) OBTAINED WITH THE CONSIDERED EXAMPLES FOR THE DIFFERENT METHODS AND SELECTED VALUES OF ℓ.

Example	Method	$\ell = 10$	$\ell = 25$	$\ell = 50$
Baart	Cheb. basis	$2.87 \cdot 10^{-2}$	$3.86 \cdot 10^{-2}$	$5.35 \cdot 10^{-2}$
	Cheb.-Nyström	$2.76 \cdot 10^{-2}$	$3.78 \cdot 10^{-2}$	$5.49 \cdot 10^{-2}$
	Cheb. projection	$4.36 \cdot 10^{-2}$	$5.98 \cdot 10^{-2}$	$9.36 \cdot 10^{-2}$
	Randomized	$4.38 \cdot 10^{-2}$	$5.95 \cdot 10^{-2}$	$9.47 \cdot 10^{-2}$
	Krylov	$1.85 \cdot 10^{-1}$	$4.39 \cdot 10^{-1}$	$8.71 \cdot 10^{-1}$
Shaw	Cheb. basis	$3.51 \cdot 10^{-2}$	$4.03 \cdot 10^{-2}$	$5.51 \cdot 10^{-2}$
	Cheb.-Nyström	$2.64 \cdot 10^{-2}$	$3.53 \cdot 10^{-2}$	$5.16 \cdot 10^{-2}$
	Cheb. projection	$4.96 \cdot 10^{-2}$	$6.05 \cdot 10^{-2}$	$9.45 \cdot 10^{-2}$
	Randomized	$4.97 \cdot 10^{-2}$	$6.08 \cdot 10^{-2}$	$9.57 \cdot 10^{-2}$
	Krylov	$1.95 \cdot 10^{-1}$	$4.45 \cdot 10^{-1}$	$8.77 \cdot 10^{-1}$
Blur	Cheb. basis	$3.91 \cdot 10^{-1}$	$9.36 \cdot 10^{-1}$	$1.82 \cdot 10^{-1}$
	Cheb. projection	$7.45 \cdot 10^{-2}$	$1.79 \cdot 10^{-1}$	$3.47 \cdot 10^{-1}$
	Randomized	$7.48 \cdot 10^{-2}$	$1.79 \cdot 10^{-1}$	$3.48 \cdot 10^{-1}$
	Krylov	$1.81 \cdot 10^{-1}$	$4.28 \cdot 10^{-1}$	$8.60 \cdot 10^{-1}$
Baart2D	Cheb. basis	$2.75 \cdot 10^{-2}$	$3.71 \cdot 10^{-2}$	$5.18 \cdot 10^{-2}$
	Cheb.-Nyström	$2.79 \cdot 10^{-2}$	$3.77 \cdot 10^{-2}$	$5.22 \cdot 10^{-2}$
	Cheb. projection	$4.20 \cdot 10^{-2}$	$5.85 \cdot 10^{-2}$	$9.30 \cdot 10^{-2}$
	Randomized	$4.18 \cdot 10^{-2}$	$5.87 \cdot 10^{-2}$	$9.26 \cdot 10^{-2}$
	Krylov	$1.90 \cdot 10^{-1}$	$4.52 \cdot 10^{-1}$	$9.06 \cdot 10^{-1}$

ACKNOLEDGMENT

The authors would like to thank a referee for comments that lead to a clarified presentation.

REFERENCES

[1] M. Baart, "The use of auto-correlation for pseudo-rank determination in noisy iii-conditioned linear least-squares problems," *IMA Journal of Numerical Analysis*, vol. 2, no. 2, pp. 241–247, 1982.

[2] X. Bai, A. Buccini, and L. Reichel, "Golub–kahan vs. monte carlo: a comparison of bidiagonlization and a randomized svd method for the solution of linear discrete ill-posed problems," *BIT Numerical Mathematics, in press*.

[3] X. Bai, G.-X. Huang, X.-J. Lei, L. Reichel, and F. Yin, "A novel modified trsvd method for large-scale linear discrete ill-posed problems," *Applied Numerical Mathematics*, vol. 164, pp. 72–88, 2021.

[4] A. Buccini, M. Donatelli, and L. Reichel, "Iterated Tikhonov regularization with a general penalty term," *Numerical Linear Algebra with Applications*, vol. 24, no. 4, p. e2089, 2017.

[5] A. Buccini, M. Pasha, and L. Reichel, "Generalized singular value decomposition with iterated Tikhonov regularization," *Journal of Computational and Applied Mathematics*, vol. 373, p. 112276, 2020.

[6] D. Calvetti and L. Reichel, "Tikhonov regularization of large linear problems," *BIT Numerical Mathematics*, vol. 43, no. 2, pp. 263–283, 2003.

[7] D. Calvetti, L. Reichel, and A. Shuibi, "Invertible smoothing preconditioners for linear discrete ill-posed problems," *Applied numerical mathematics*, vol. 54, no. 2, pp. 135–149, 2005.

[8] L. Dykes, S. Noschese, and L. Reichel, "Rescaling the GSVD with application to ill-posed problems," *Numerical Algorithms*, vol. 68, no. 3, pp. 531–545, 2015.

[9] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[10] L. Eldén, "A weighted pseudoinverse, generalized singular values, and constrained least squares problems," *BIT Numerical Mathematics*, vol. 22, no. 4, pp. 487–502, 1982.

[11] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*. Springer Science & Business Media, 1996, vol. 375.

[12] C. Fenu, L. Reichel, G. Rodriguez, and H. Sadok, "GCV for Tikhonov regularization by partial SVD," *BIT Numerical Mathematics*, vol. 57, no. 4, pp. 1019–1039, 2017.

[13] S. Gazzola, P. Novati, and M. R. Russo, "On Krylov projection methods and tikhonov regularization," *Electronic Transactions on Numerical Analysis*, vol. 44, no. 1, pp. 83–123, 2015.

[14] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU press, 2013, vol. 3.

[15] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Review*, vol. 53, no. 2, pp. 217–288, 2011.

[16] P. C. Hansen, "Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems," *Numerical algorithms*, vol. 6, no. 1, pp. 1–35, 1994.

[17] ——, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM, 1998.

[18] M. E. Hochstenbach and L. Reichel, "An iterative method for Tikhonov regularization with a general linear regularization operator," *The Journal of Integral Equations and Applications*, pp. 465–482, 2010.

[19] S. Kindermann, "Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems," *Electronic Transactions on Numerical Analysis*, vol. 38, pp. 233–257, 2011.

[20] S. Kindermann and K. Raik, "A simplified l-curve method as error estimator," *Electronic Transactions on Numerical Analysis*, vol. 53, pp. 217–238, 2020.

[21] Y. Park, L. Reichel, G. Rodriguez, and X. Yu, "Parameter determination for tikhonov regularization problems in general form," *Journal of Computational and Applied Mathematics*, vol. 343, pp. 12–25, 2018.

[22] L. Reichel and G. Rodriguez, "Old and new parameter choice rules for discrete ill-posed problems," *Numerical Algorithms*, vol. 63, no. 1, pp. 65–87, 2013.

[23] L. Reichel and A. Shyshkov, "A new zero-finder for tikhonov regularization," *BIT Numerical Mathematics*, vol. 48, no. 3, pp. 627–643, 2008.

[24] C. Shaw Jr, "Improvement of the resolution of an instrument by numerical solution of an integral equation," *Journal of Mathematical Analysis and Applications*, vol. 37, no. 1, pp. 83–112, 1972.

[25] L. N. Trefethen, *Approximation Theory and Approximation Practice, Extended Edition*. SIAM, 2019.

[26] H. Xiang and J. Zou, "Regularization with randomized SVD for large-scale discrete inverse problems," *Inverse Problems*, vol. 29, no. 8, p. 085008, 2013.

[27] ——, "Randomized algorithms for large-scale inverse problems with general Tikhonov regularizations," *Inverse Problems*, vol. 31, no. 8, p. 085008, 2015.